May 19, 2011

Linear Algebra

Lawrence N. Stout, *Illinois Wesleyan University*

Linear Algebra

Lawrence Neff Stout
Illinois Wesleyan University

May 19, 2011

Contents

Preface

Linear algebra is the study of vector spaces and linear transformations. This means that we study a very large variety of spaces but only one kind of function between them, unlike calculus, in which one space and lots of kinds of functions are studied. The subject provides a nice transition from concrete, problem-oriented mathematics to abstract, structure-oriented mathematics.

At Illinois Wesleyan University, linear algebra is a sophomore level course taken after completion of the calculus sequence by students from a variety of majors. This is the audience I have written for. I have used the links with calculus in the examples to relate linear algebra to material in my students' recent past. One of the reasons that I do not start with solving systems of linear equations is that in thirty years of teaching I have not found a student who entered a linear algebra course wanting to know how to solve large systems of linear equations by hand.

This book is designed to deal with all of the concepts of linear algebra first in $\mathbb{R}^2$, a simple context where algorithmic concerns are minimized and geometric intuition can be brought to bear. Then those same concepts are dealt with in full generality. I find that this conceptual front loading helps in the understanding of the rest of the material. Students can see easily what aspects of the plane are focused on when we think of it as a vector space; linear transformations can be understood as deformations of the plane of a very special type. Nothing more complicated than the solution of quadratic equations or the solution of systems of equations in two unknowns is needed– background which can be quickly resurected from high school algebra.

The limitation to two dimensional space has some drawbacks: the geometry is not as rich as in higher dimensions and some of the subtilties of matrix calculation only become evident in larger matrices. However, the object is to *introduce* the concepts in a context where the technique needed to solve the problems will cause no difficulties. But already in $\mathbb{R}^3$ it is easy to lose sight of the concepts because of the difficulties of solving systems of equations and solving cubics. Staying in $\mathbb{R}^2$ keeps the course from bogging down in chapters 1-4 while introducing all of the concepts (vector spaces, linear transformations, subspaces, kernels, images, bases, matrices, eigenvalues, cannonical forms, inner products, orthogonal projection) that are developed in general in the rest of the course.

This is a book in linear algebra, not matrix algebra, and not numerical linear algebra. I think that the difficulties inherent in doing matrix compu-

tations with floating point arithmetic are best left until it is clear what the meaning of those calculations is; hence my choice to use exact arithmentic throughout. Because so much of linear algebra depends on the interplay between linear transformations and the matrices which make them concrete, I have tried to take an algorithmic approach as well as a conceptual approach to the theory. This is, in essence, a course in how to take the row reduction algorithm and squeeze every bit of information you can out of it. The questions we ask come from the general concepts; the answers often come from careful consideration of the algorithm.

The core of a course in linear algebra lies in Chapters 1-13.1. Sections 3.2, 5.3, 6.3, 7.3, 10.3, and 11.2 can be omitted without loss of continuity (though they do include some nifty stuff!). Once the core material has been covered the instructor has choices to make: one can cover 11.2, 13.2, 13.3, 14.1, and then Chapter 15 if one wants to get a focus on cannonical forms, or one can omit 13.2 and 13.3 and do all of Chapter 14 to get least-squares approximation in function spaces. It is also possible to do Chapter 14 right after Chapter 11 if the instructor is more concerned with getting to the inner product material than with eigenvalues. The core material fits nicely in 13 weeks, the whole book would take a full 16 week semester or two quarters.

When I teach the course I use the supplemental materials in *Mathematica Notebooks for Linear Algebra* by Robin Sue Sanders and myself. Some of these notebooks grew out of notebooks I used to generate illustrations for the text–but they are better than illustrations because the student can play with them and generate further examples or see dynamics animated. I particularly like the notebook on Coding Theory which would accompany section 6.3. Students are surprized when they can actually *see* what is happening in a 7 dimensional space. Several notebooks also provide tools for doing the basic row operations to solve equations, find inverses, find ranks, change basis, and find determinants over any of the fields discussed in the book (plus a few not covered in the book). For large systems it makes more sense to use a computer than to do the row reduction by hand, but for understanding it is important not to make the algorithm into a black box. The notebooks introduce a pivot operation only after use of the elementary row operations has been mastered.

Each section of the book introduces some concepts with a variety of examples. The exercises include straightforward computations, quick corollaries of theorems, proofs of parts of major theorems, and (in many sections) applications or further theoretical developments in the form of Project Problems.

Answers to the odd numbered problems are included in the appendix, and there is a complete solutions manual available to instructors.

Appendix A gives a compilation of the definitions in the text, together with references to where further explanation can be found. Because there are so many concepts in linear algebra and because the book does them first in $\mathbb{R}^2$ and then in general, students may find this appendix useful both as a reminder of the definitions and as a pointer to where to find the two-dimensional case when they encounter the general concept.

This manuscript started out as a joint project with William K. Smith. He withdrew from the project on his retirement from Illinois Wesleyan University. Without his encouragement it would never have been written. Robin Sue Sanders has also helped with numerous conversations, teaching from the manuscript several times, and developing the Mathematica supplement. Thanks are also due to several classes of linear algebra students at Illinois Wesleyan University who used previous versions of the book. Craig Zirbel caught numerous typos and made many valuable editorial comments, many of which were incorporated in the current version. Laura Chik helped with the answer key, both catching my errors and providing some solutions. Justin Rodriguez and Christopher Hatfield both caught many typos.

Chapter 1

$\mathbb{R}^2$ as a Vector Space

Linear algebra is the study of vector spaces and linear transformations. These are abstract notions derived and generalized from useful notions in analytic geometry and physics. Much of the power of mathematics comes from its abstraction, which allows us to recognize similar structures in varied situations and lets us apply techniques and ideas developed using intuition gleaned from one situation to quite different examples. Often a branch of mathematics has developed by abstracting properties in a particular situation, studying those properties in a more general setting, and then noticing much wider application than was apparent in the original situation. The abstraction process is one of focusing in on specific aspects a motivating example, finding varied examples with the same properties, and then making the theory explicit and giving it rigorous form. The power of mathematics comes from all three parts: good motivating examples, a wide variety of interesting examples with extensive application, and the certainty which comes from a rigorous theoretical development.

We will start our study of linear algebra by looking at the structure of the plane $\mathbb{R}^2$ which makes it a vector space. This is a convenient example because the structure which makes it a vector space is sufficiently rich to say interesting things about the plane, but the plane is simple enough that we do not get buried in the technique needed to do the relevant calculations. After we have introduced all of the major notions of linear algebra in the context of the motivating example of the plane, we will give fully general definitions and look at a much wider variety of examples. It is the object of the present chapter to illustrate how notions from mechanics and analytic geometry were abstracted to obtain the fundamental notions of linear algebra.

1.1 Vector arithmetic in $\mathbb{R}^2$

Students of physics learn in their study of mechanics that there are several different physical concepts which require both a magnitude and a direction. Examples are velocity, momentum, acceleration, and force. Beginning our process of abstraction, we see that all of these physical concepts can be given a representation as a directed line segment, the length corresponding to the magnitude and the orientation giving the direction. For ease in drawing diagrams we will use two dimensions; the ideas are the same in three, four, or seventeen dimensions but our geometrical intuition is weaker there. As part of our abstraction process, let us call one of these quantities with both magnitude and direction a **vector**. We will write $\vec{v}$ for a vector and $\|\vec{v}\|$ for its magnitude. When we represent vectors geometrically we will use arrows pointing in the direction of the vector with length given by $\|\vec{v}\|$. A vector with length 0 doesn't really have a direction, so we will treat it as a separate (though important) special case, the zero vector $\vec{0}$.

When we describe a new mathematical object (as we are doing now with vectors) we want to be careful to specify what we mean by equality. For instance, two functions are considered to be equal if they have the same domain, the same codomain, and the same value at each element of the domain, rather than requiring that they be expressed using exactly the same string of symbols. Vectors will be considered to be the same if they have the same direction and the same length, so we can move the starting point without changing the vector. To emphasize our focus on magnitude and direction as the important features of geometric vectors we use those to define equality:

Definition 1.1.1 *Two vectors will be considered equal if they have the same direction and magnitude or if both are the zero vector.*

Engineers frequently calculate the resultant of several forces; physicists often study the relationship between forces, accelerations, and changes of momentum. Addition of forces is done using the "parallelogram law" as shown in Figure 1.1. This law was developed as a result of practical experiments in architecture where it was used to calculate loading in masonry buildings. The two vectors to be added are arranged as adjacent sides of a parallelogram and the sum is the diagonal. The same vector results when the vectors are thought of as giving displacements and the second is traversed starting at

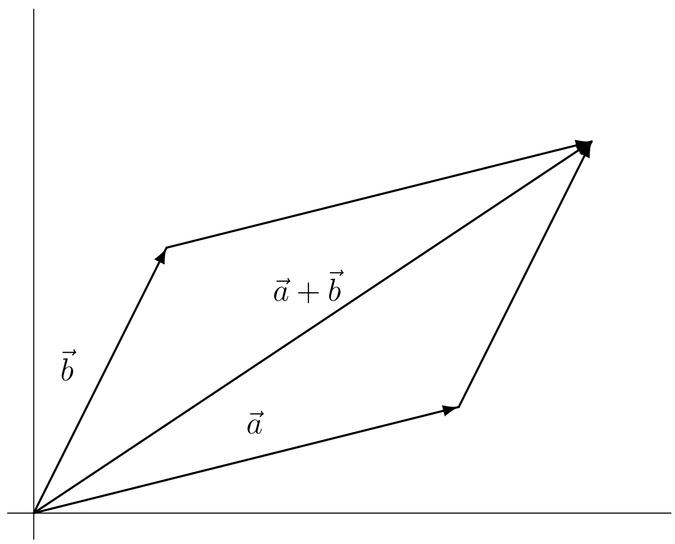


Figure 1.1: Addition Using the Parallelogram Law

the endpoint of the first. Vector addition gives a vector as the sum of two vectors.

There is a second way to obtain new vectors from old which is used in expressing several physical laws. If the magnitude of a force $\vec{F}$ is doubled but the direction is kept the same, it is natural to call the new force $2\vec{F}$. If a train traveling with velocity $\vec{V}$ stops and then goes in reverse at $\frac{1}{3}$ its previous speed, then its new velocity will be $-\frac{1}{3}\vec{V}$. In general, we can multiply a vector by a number (in this context called a **scalar**). If $k = 0$ then $k\vec{v} = \vec{0}$. Otherwise, the magnitude of $k\vec{v}$ is $|k|$ times that of $\vec{v}$. If $k < 0$ then the direction reverses; if $k > 0$ it stays the same. Two vectors will be called **parallel** if one is a non-zero scalar multiple of the other: notice that parallel vectors are represented geometrically by parallel arrows.

The geometric approach to vectors is convenient for drawing pictures and estimating effects, but it becomes very cumbersome when we try to give geometric proofs of all of the properties of the sum and scalar multiplication defined above and when we shift to higher dimensions.

The main contribution of Descartes to modern mathematics was the observation that imposing a coordinate system allows us to look at the geometry using algebra, which is easier to handle and generalize, so we impose a coordinate system. If we agree to draw a vector $\vec{V}$ with its initial point at

the origin, then its terminal point will have coordinates (v_1, v_2). This pair of numbers determines the vector uniquely, and we write it as $\vec{V} = [v_1, v_2]$, using the square brackets to distinguish the vector $[v_1, v_2]$ from the point (v_1, v_2). Imposing coordinates in this way lets us give an algebraic definition of what a vector is:

Definition 1.1.2 *A vector $\vec{a}$ in the plane $\mathbb{R}^2$ is an ordered pair of real numbers $\vec{a} = [a_1, a_2]$. Two vectors $\vec{a} = [a_1, a_2]$ and $\vec{b} = [b_1, b_2]$ are equal if $a_1 = b_1$ and $a_2 = b_2$. The length of the vector $\vec{a}$ is $\|\vec{a}\| = \sqrt{a_1^2 + a_2^2}$.*

It turns out that the two operations defined earlier using geometry have a particularly nice representation in terms of the coordinates, as shown by the next proposition.

Proposition 1.1.1 *The sum $\vec{a} + \vec{b}$ of the vectors $\vec{a} = [a_1, a_2]$ and $\vec{b} = [b_1, b_2]$ is $\vec{a} + \vec{b} = [a_1 + b_1, a_2 + b_2]$, and the product $k\vec{a}$ (where k is a scalar) is $[ka_1, ka_2]$.*

PROOF:

We give a geometric proof using the constructions sketched in Figures 1.2 and 1.3. In Figure 1.2, $\vec{a} + \vec{b}$ is obtained by the parallelogram law. But clearly the two triangles labeled **I** are congruent because they each have sides adjacent to the right angle of length b_1 and b_2. Thus the x-coordinate of $\vec{A} + \vec{B}$ is $a_1 + b_1$. Similarly, the y-coordinate of $\vec{A} + \vec{B}$ is $a_2 + b_2$.

For the second statement refer to Figure 1.3. We notice that the triangles with hypotenuse given by $\vec{A}$ and $k\vec{A}$ are similar. This implies that the ratios of corresponding sides are equal. Since the length of $k\vec{A}$ is k times the length of $\vec{A}$, the x-coordinate of $k\vec{A}$ will be k times as large as the x-coordinate of $\vec{A}$ and similarly for the y-coordinates . ∎

With this result and two more definitions we will be able to prove the algebraic properties of vectors in $\mathbb{R}^2$ from the properties of real numbers.

Definition 1.1.3 *The zero vector $\vec{0} = [0, 0]$. For any vector $\vec{a} = [a_1, a_2]$, the vector $-\vec{a} = (-1)\vec{a} = [-a_1, -a_2]$.*

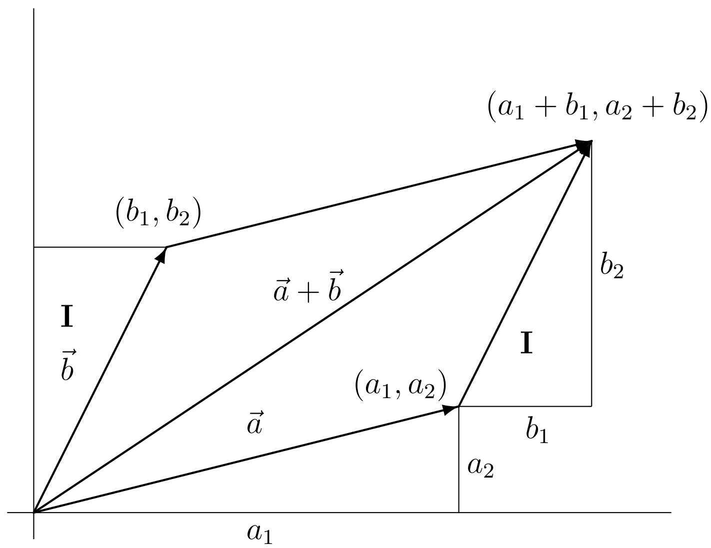


Figure 1.2: Making the sum algebraic

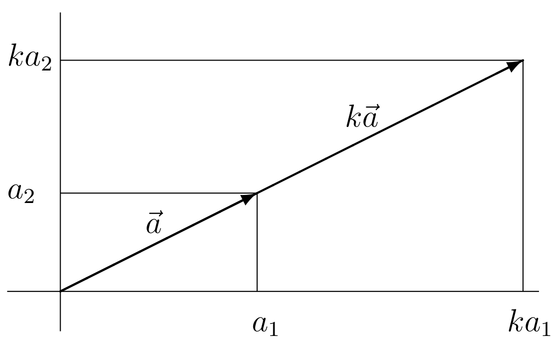


Figure 1.3: Making the product by a scalar algebraic

Before we use the properties of the real numbers we should point out *which* properties are relevant for algebra. We will not be using order properties or completeness (both of which are important in calculus). What we need are the properties which make $\mathbb{R}$ a *field*:

Definition 1.1.4 *The real numbers form a* **field** *because there are operations of addition (written as $a+b$) and multiplication (written ab) such that for all real numbers $a, b,$ and c we have*

Closure:	$a+b$ *is a real number*	ab *is a real number*
Commutative:	$a+b=b+a$	$ab=ba$
Associative:	$a+(b+c)=(a+b)+c$	$a(bc)=(ab)c$
Identity:	*there is a number* 0 *such that* $0+a=a+0=a$	*there is a number* 1 *such that* $1a=a1=a$
Inverses:	*For any a there is a number* $-a$ *with* $a+(-a)=0$	*If $a\neq 0$ then there is a number* $\frac{1}{a}$ *with* $a(\frac{1}{a})=1$
Distributive:	$a(b+c)=(ab)+(ac)$	

The properties listed are abstracted from the properties you use in arithmetic and are precisely the ones needed to be able to manipulate expressions in algebra and solve equations. When we generalize to higher dimensions, we get similar properties for addition of vectors (which takes two vectors and combines them to give another vector) and multiplication by a scalar (which takes a real number and a vector and returns a vector). The structure resulting is that of a *vector space* over $\mathbb{R}$:

Definition 1.1.5 *A vector space $\mathcal{V}$ over the real numbers is a set (whose elements are called vectors) equipped with two operations. Addition takes two vectors and returns a vector as their sum, and multiplication by a scalar takes a real number and a vector and returns a vector. These operations satisfy the following axioms for all vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$, and real numbers k and h:*

Closure:	$\vec{a}+\vec{b}$ *and* $k\vec{a}$ *are vectors*
Commutative law for addition:	$\vec{a}+\vec{b}=\vec{b}+\vec{a}$
Associative law for addition:	$(\vec{a}+\vec{b})+\vec{c}=\vec{a}+(\vec{b}+\vec{c})$
Identity:	*There is a vector* $\vec{0}$ *such that for all* $\vec{a}$, $\vec{a}+\vec{0}=\vec{a}$
Inverses:	*For any vector* $\vec{a}$ *there is a vector* $-\vec{a}$ *with* $\vec{a}+-\vec{a}=\vec{0}$
Absorption:	$k(h\vec{a})=(kh)\vec{a}$
Distributive laws:	$(h+k)\vec{a}=(h\vec{a})+(k\vec{a})$ $h(\vec{a}+\vec{b})=(h\vec{a})+(h\vec{b})$
Identity for scalars:	$1\vec{a}=\vec{a}$

Proposition 1.1.2 *Using vector addition and multiplication by a scalar given componentwise,* $\mathbb{R}^2$ *is a vector space over* $\mathbb{R}$.

PROOF:

Most of these involve little more than converting the vectors to ordered pairs of real numbers and then applying the law of the same form for real numbers twice. For example, to prove commutativity for addition we calculate

$$\vec{a}+\vec{b}=[a_1,a_2]+[b_1,b_2]=[a_1+b_1,a_2+b_2].$$

Now apply the commutative law for addition of real numbers to get $a_1+b_1=b_1+a_1$ and $a_2+b_2=b_2+a_2$. Then apply the definition of equality to get

$$[a_1,a_2]+[b_1,b_2]=[b_1,b_2]+[a_1,a_2].$$

Thus showing that

$$\vec{a}+\vec{b}=\vec{b}+\vec{a}.$$

The absorption law for multiplication by scalars uses two applications of the associative law for multiplication.

$$\begin{aligned}
k(h\vec{a}) &= k(h[a_1,a_2]) \\
&= k[ha_1,ha_2] \\
&= [k(ha_1),k(ha_2] \\
&= [(kh)a_1,(kh)a_2] \\
&= (kh)\vec{a}.
\end{aligned}$$

We use the different name absorption for the vector space property because it actually involves two different kinds of multiplication–multiplication of a vector by a real number and multiplication of two real numbers–rather than two applications of the same kind of multiplication as in the associative property for $\mathbb{R}$.

The other properties are left as exercises.

∎

The proof just given is rather pedantic; very little real work is needed to prove the properties of the operations. This illustrates the power of the algebraic approach. Furthermore, it is almost trivial to generalize this proposition to higher dimensions. If we want to prove commutativity for addition of 17-tuples of real numbers, all we have to do is quote commutativity for real numbers 17 times. This certainly is easier than coming up with a proof in 17-dimensional Euclidean geometry!

Exercises 1.1:

1. Let $\vec{a} = [3, 5]$, $\vec{b} = [-2, 7]$, $\vec{c} = [5, -12]$, and $\vec{d} = [1, -3]$. Find the following:
   (a) $\vec{a} + 5\vec{c}$
   (b) $(\vec{a} + \vec{d}) - 4(\vec{b} + 2\vec{c})$
   (c) $\pi\vec{a} + 3(\vec{b} + \vec{d})$
   (d) $4(\vec{a} + 3\vec{c}) - 6.7(\vec{d} - \vec{c})$
2. Prove the remaining parts of Proposition 1.1.2
3. Prove that a vector has a unique additive inverse; that is, if both
$$\begin{aligned} \vec{v} + -_1\vec{v} &= \vec{0} \\ \vec{v} + -_2\vec{v} &= \vec{0} \end{aligned}$$
then $-_1\vec{v} = -_2\vec{v}$. (Hint: use the associative law to expand $-_1\vec{v} + \vec{v} + -_2\vec{v}$ in two different ways.)

1.2 Bases for $\mathbb{R}^2$

In physics and engineering it is fairly standard practice to write vectors in terms of the standard basis vectors $\vec{\imath} = [1, 0]$ and $\vec{\jmath} = [0, 1]$: for instance

$$[3, 4] = 3\vec{\imath} + 4\vec{\jmath}.$$

The property that makes these *basis vectors* is that any vector $\vec{a} = [a_1, a_2]$ can be represented as $a_1\vec{\imath} + a_2\vec{\jmath}$ in exactly one way.

Definition 1.2.1 *A basis for $\mathbb{R}^2$ is a set of vectors $\{\vec{b}_1, \ldots, \vec{b}_n\}$ such that any vector in $\mathbb{R}^2$ can be written as $k_1\vec{b}_1 + \ldots + k_n\vec{b}_n$ in exactly one way.*

Other bases for $\mathbb{R}^2$ are not difficult to find: any pair of vectors which are not parallel will do.

Example: The set $\{[1, 2], [3, 4]\}$ is a basis.

To show that

$$[a, b] = x[1, 2] + y[3, 4]$$

has a unique solution we solve the pair of equations

$$\begin{aligned} a &= x + 3y \\ b &= 2x + 4y. \end{aligned}$$

We get

$$\begin{aligned} x &= \frac{3b - 4a}{2} \\ y &= \frac{b - 2a}{-2}. \end{aligned}$$

$\diamondsuit$

As long as we stick to two equations in two unknowns this is not too hard. (Efficient techniques for finding solutions to systems of equations will be one of the major tools of linear algebra; we will study them in detail in Chapter 7.)

Example: The parallel vectors $[1, 1]$ and $[-1, -1]$ do not form a basis.

There exist vectors which cannot be written as $x[1, 1]+y[-1, -1]$ for any choice of x and y. For example, if we try to write $[2, -2]$ as $x[1, 1] + y[-1, -1]$ we end up with the system of equations

$$\begin{aligned} x - y &= 2 \\ x - y &= -2 \end{aligned}$$

which clearly has no solutions. $\diamondsuit$

Any set of vectors containing two parallel vectors will run into this same difficulty. A set containing three nonparallel vectors also fails to be a basis for $\mathbb{R}^2$:

Example: The set $\{[1, 0], [0, 1], [1, 1]\}$ is not a basis.

What we lose is the uniqueness of representation. For example

$$[1, 2] = 1[1, 0] + 2[0, 1] + 0[1, 1]$$

but we also have

$$[1, 2] = 0[1, 0] + 1[0, 1] + 1[1, 1].$$

$\diamondsuit$

These examples illustrate the characterization of a basis for $\mathbb{R}^2$ as a pair of nonparallel non zero vectors.

Picking an ordered basis for $\mathbb{R}^2$ is analogous to picking a coordinate system. If we use the standard basis we recover the coordinates of the endpoint of our vector from the coefficients of $\vec{\imath}$ and $\vec{\jmath}$. Using a different ordered basis we will get different "coordinates":

Definition 1.2.2 *If $B = (\vec{b}_1, \vec{b}_2)$ is an ordered basis for $\mathbb{R}^2$, then the B-coordinate representation of $\vec{a} = k_1\vec{b}_1 + k_2\vec{b}_2$ is the column vector $\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$. Because column vectors are awkward within text, we often write this column as $[k_1, k_2]^t$, where the ^{t} indicates transpose, an operation which turns rows into columns.*

Example: Finding B-coordinates

For the ordered basis $B = ([1,2],[3,4])$ the B-coordinate representation of $[1,4]$ is found by solving the system of equations $[1,4] = x[1,2] + y[3,4]$ which is equivalent to

$$\begin{aligned} 1x + 3y &= 1 \\ 2x + 4y &= 4. \end{aligned}$$

This has solution $x = 4$ and $y = -1$, so the B-coordinate representation of $[1,4]$ is $\begin{bmatrix} 4 \\ -1 \end{bmatrix}$. ♢

Exercises 1.2:

1. Let $\vec{v} = [1,1]$ and $\vec{w} = [-1,2]$. For each of the following vectors $\vec{a}$ find scalars x and y so that $\vec{a} = x\vec{v} + y\vec{w}$, then give the B-coordinate representation, where $B = ([1,1],[-1,2])$.
   (a) $\vec{a} = [3,0]$
   (b) $\vec{a} = [-4,3]$
   (c) $\vec{a} = [2,5]$
   (d) $\vec{a} = [0,-4]$
   (e) Show that $\{[1,1],[-1,2]\}$ is a basis for $\mathbb{R}^2$.
2. Repeat exercise 1 using $\vec{v} = [2,-1]$ and $\vec{u} = [-3,0]$.
3. Let $\vec{a} = [-3,2]$ and $\vec{b} = [6,-4]$. Find scalars c and d so that $c\vec{a}+d\vec{b} = \vec{0}$. Is there more than one solution to this problem?
4. Let $\vec{a} = [2,-1]$ and $\vec{b} = [-3,1.5]$. Find scalars c and d so that $c\vec{a}+d\vec{b} = \vec{0}$. Is there more than one solution to this problem?
5. Let $\vec{v}$ and $\vec{w}$ be nonzero vectors in $\mathbb{R}^2$ for which there exist nonzero scalars a and b so that $a\vec{v} + b\vec{w} = \vec{0}$. Prove that $\vec{v}$ and $\vec{w}$ are parallel (that is, show that there is a scalar k such that $\vec{v} = k\vec{w}$).

6. Suppose that the only solution to $x\vec{u}+y\vec{v}=\vec{0}$ is $x=0$ and $y=0$. Show that if $\vec{w}=a_1\vec{u}+b_1\vec{v}$ and $\vec{w}=a_2\vec{u}+b_2\vec{v}$, then $a_1=a_2$ and $b_1=b_2$.

7. Show that if $\vec{u}$ and $\vec{v}$ are non-zero and non-parallel, then there is only one solution to $x\vec{u}+y\vec{v}=\vec{0}$.

8. Show that $\{[1,3],[-2,-6]\}$ is not a basis for $\mathbb{R}^2$.

Chapter 2

Linear Transformations on $\mathbb{R}^2$

If all that we did with vectors were addition and multiplication by scalars they would form a small part of some other course (probably in architecture or physics rather than mathematics) instead of being the start of a subject central to much of modern mathematics. The subject of linear algebra becomes worthy of intensive study when we start thinking about the kind of functions (called linear transformations) between vector spaces which preserve these operations on vectors. Linear transformations are made concrete using matrices. Linear algebra becomes the structural analysis of properties of linear transformations and matrices. In this section we will restrict ourselves to linear transformations from $\mathbb{R}^2$ to itself. They have their origins in questions of rescaling and rotation in drawing graphs.

2.1 Definitions, examples, and operations

In Chapter 1 we have described the structure of $\mathbb{R}^2$ which results from concentrating on operations of vector addition and multiplication by scalars. The next definition tells us that linear transformations are functions which preserve this structure.

Definition 2.1.1 *A linear transformation from $\mathbb{R}^2$ to itself is a function $L : \mathbb{R}^2 \to \mathbb{R}^2$ such that $L(\vec{a} + \vec{b}) = L(\vec{a}) + L(\vec{b})$ and $L(k\vec{a}) = kL(\vec{a})$ for all vectors $\vec{a}$ and $\vec{b}$ and all real numbers k.*

Because we will be using the terms later it might be wise to recall that a function $f : D \to C$ is a rule which assigns to each element of the domain

D a unique element of the codomain C. All elements of D must be assigned values under f but C gives what *could be* hit by f rather than what *is* hit. (The set $\{f(x)|x \in D\}$ is called the image, or range, of f; it is a subset of the codomain.)

First we should note that this is really very restrictive. There aren't a lot of linear transformations. Any linear transformation must preserve the zero vector, since

$$L(\vec{0}+\vec{0}) = L(\vec{0})$$

by the property of the zero vector, and

$$L(\vec{0}+\vec{0}) = L(\vec{0}) + L(\vec{0})$$

by linearity. Thus

$$L(\vec{0}) = L(\vec{0}) + L(\vec{0})$$

and by adding inverses we conclude that

$$L(\vec{0}) = \vec{0}.$$

Furthermore, once we know where a linear transformation sends $\vec{\imath}$ and $\vec{\jmath}$ we will know what it does to all vectors in $\mathbb{R}^2$. To find $L([a,b])$, we write $[a,b] = a\vec{\imath} + b\vec{\jmath}$ so

$$L([a,b]) = L(a\vec{\imath} + b\vec{\jmath}) = aL(\vec{\imath}) + bL(\vec{\jmath}).$$

2.1.1 Examples

With this in mind let us look at some examples. In each case we provide a picture of how the maps transform the unit circle and a square in the plane with an arrow indicating orientation.

Example: A vertical stretch:

Consider the function L with value given by

$$L([a,b]) = [a, 2b]$$

To show that L is linear we need to show that it preserves both addition and scalar multiplication. Both of these are easy computations:

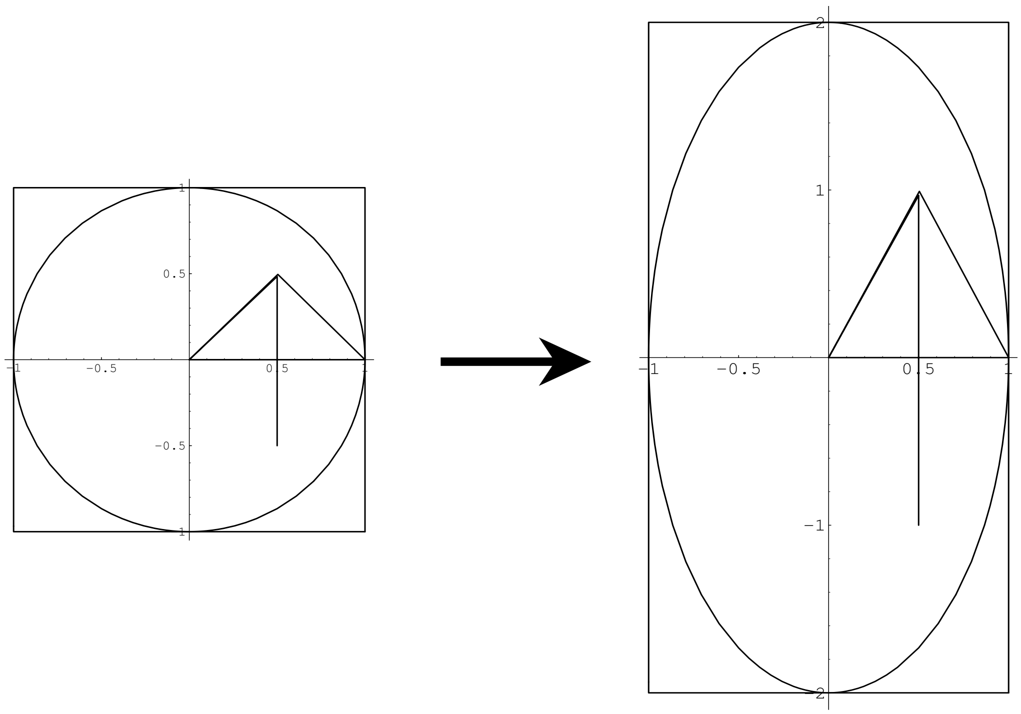


Figure 2.1: A vertical stretch

$$
\begin{aligned}
L([a_1, b_1] + [a_2, b_2]) &= L([a_1 + a_2, b_1 + b_2]) \\
&= [a_1 + a_2, 2b_1 + 2b_2] \\
&= [a_1, 2b_1] + [a_2, 2b_2] \\
&= L([a_1, b_1]) + L([a_2, b_2])
\end{aligned}
$$

and similarly for scalar multiplication:

$$
\begin{aligned}
L(k[a, b]) &= L([ka, kb]) \\
&= [ka, 2kb] \\
&= k[a, 2b] \\
&= kL([a, b])
\end{aligned}
$$

$\diamondsuit$

Example: A horizontal stretch:

Next consider the function with value given by $L_2([a, b]) = [3a, b]$

Again we need to check that this preserves sums and scalar products:

$$\begin{aligned} L_2([a_1, b_1] + [a_2, b_2]) &= L_2([a_1 + a_2, b_1 + b_2] \\ &= [3(a_1 + a_2), b_1 + b_2] \\ &= [3a_1, b_1] + [3a_2, b_2] \\ &= L_2([a_1, b_1]) + L_2([a_2, b_2]) \\ L_2(k[a, b]) &= L_2([ka, kb]) \\ &= [3ka, kb] = k[3a, b] \\ &= kL_2([a, b]) \end{aligned}$$

◊

Example: A rotation through the angle θ

Here we let the value be given by

$$R_\theta([a, b]) = [a\cos(\theta) - b\sin(\theta), a\sin(\theta) + b\cos(\theta)]$$

Here again let's show that we get a linear transformation:

$$L([a_1, b_1] + [a_2, b_2]) = L([a_1 + a_2, b_1 + b_2])$$
$$= [(a_1+a_2)\cos(\theta)-(b_1+b_2)\sin(\theta), (a_1+a_2)\sin(\theta)+(b_1+b_2)\cos(\theta)]$$
$$= [a_1\cos(\theta)+a_2\cos(\theta)-b_1\sin(\theta)-b_2\sin(\theta), a_1\sin(\theta)+a_2\sin(\theta)+b_1\cos(\theta)+b_2\cos(\theta)]$$
$$= [a_1\cos(\theta)-b_1\sin(\theta), a_1\sin(\theta)+b_1\cos(\theta)]+[a_2\cos(\theta)-b_2\sin(\theta), a_2\sin(\theta)+b_2\cos(\theta)]$$
$$= L([a_1, b_1]) + L([a_2, b_2])$$

and similarly for scalar multiplication:

$$\begin{aligned} L(k[a, b]) &= L([ka, kb]) \\ &= [ka\cos(\theta) - kb\sin(\theta), ka\sin(\theta) + kb\cos(\theta)] \\ &= kL([a, b]) \end{aligned}$$

◊

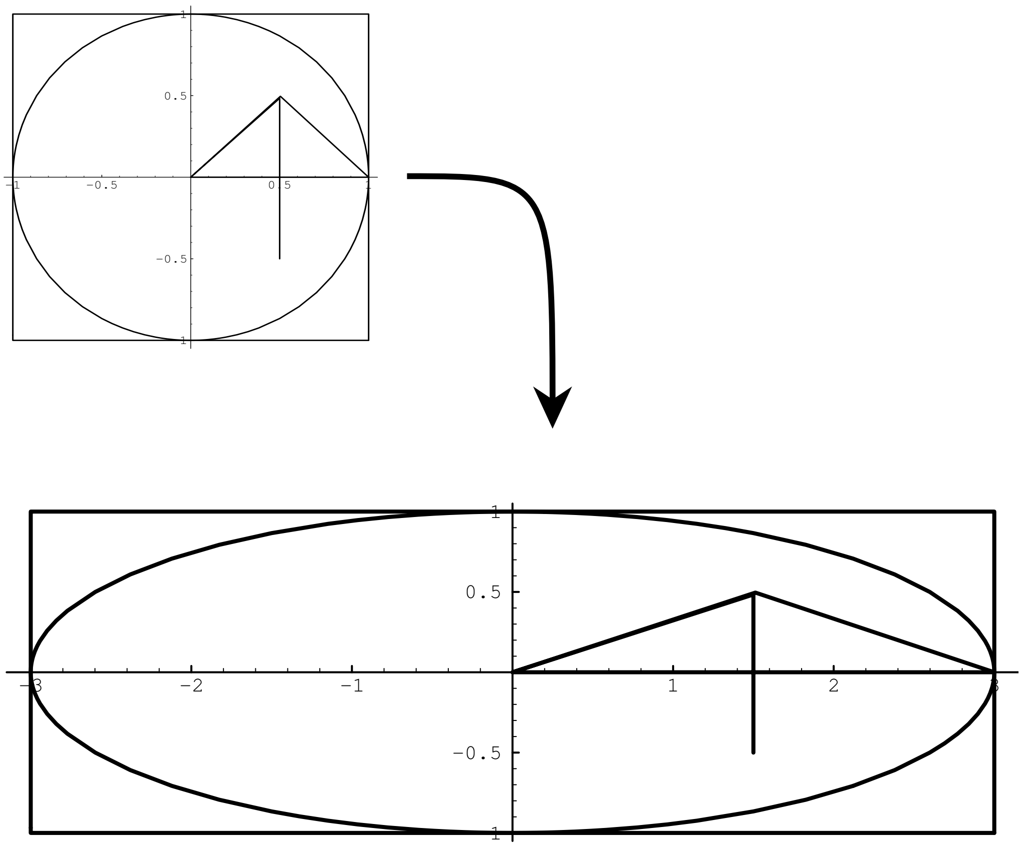


Figure 2.2: A horizontal stretch

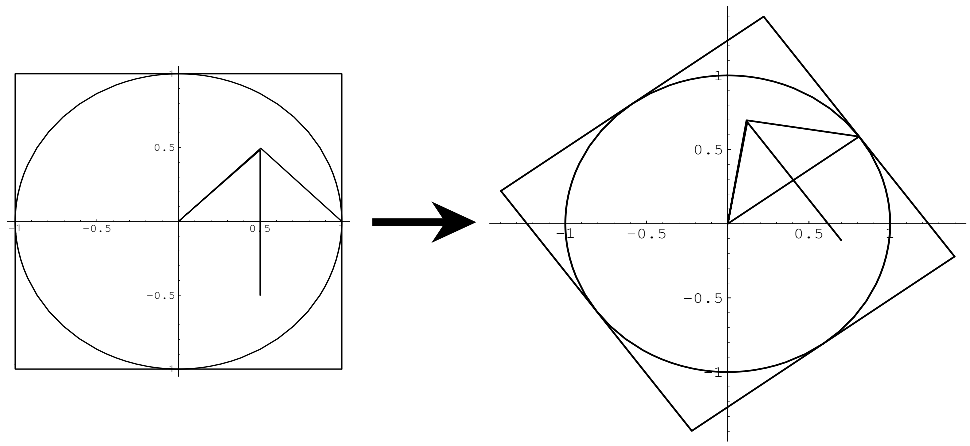


Figure 2.3: Rotation through $\theta = \frac{\pi}{5}$

Example: A reflection:

Another kind of behavior is given by the relfection through the y-axis $R([a, b]) = [-a, b]$

$\diamondsuit$

Example: A general example:

This example combines elements of the previous examples: $L([a, b]) = [2a - 3b, a + 4b]$

Here again let's show that we get a linear transformation:

$$\begin{aligned}
L([a_1, b_1] + [a_2, b_2]) &= L([a_1 + a_2, b_1 + b_2]) \\
&= [2(a_1 + a_2) - 3(b_1 + b_2), (a_1 + a_2) + 4(b_1 + b_2)] \\
&= [2a_1 + 2a_2 - 3b_1 - 3b_2, a_1 + a_2 + 4b_1 + 4b_2] \\
&= [2a_1 - 3b_1, a_1 + 4b_1] + [2a_2 - 3b_2, a_2 + 4b_2] \\
&= L([a_1, b_1]) + L([a_2, b_2])
\end{aligned}$$

and similarly for scalar multiplication:

$$L(k[a, b]) = L([ka, kb])$$

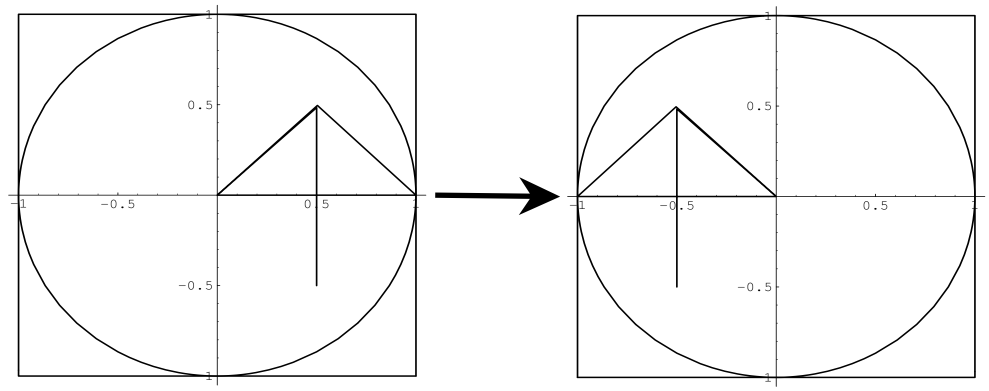


Figure 2.4: Reflection

$$\begin{aligned} &= [2ka - 3kb, ka + 4kb] \\ &= kL([a, b]) \end{aligned}$$

$\diamondsuit$

Example: A nonlinear function:

The function Q from $\mathbb{R}^2$ to $\mathbb{R}^2$ with

$$Q([a, b]) = [a^2, a - ab + 1]$$

is not a linear transformation.

First notice that $Q(\vec{0}) = [0, 1]$ so that this cannot be linear. Also,

$$Q(2[a, b]) = [4a^2, 2a - 4ab + 1] \neq 2[a^2, a - ab + 1]$$

for any choice of a and b.

$\diamondsuit$

Example: Collapsing to a line:

Some linear transformations collapse space in one direction: $C([a, b]) = [a + b, 2a + 2b]$ $\diamondsuit$

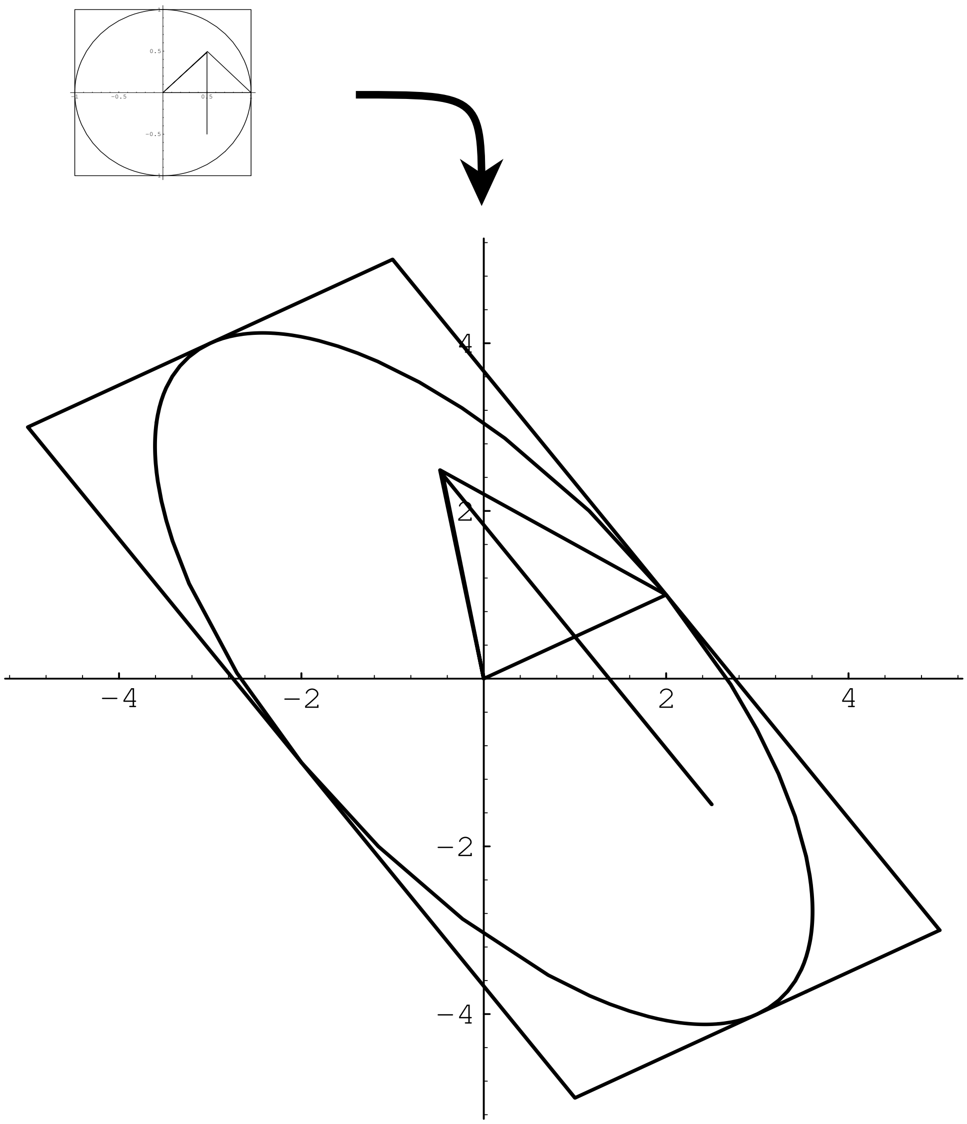


Figure 2.5: A general example

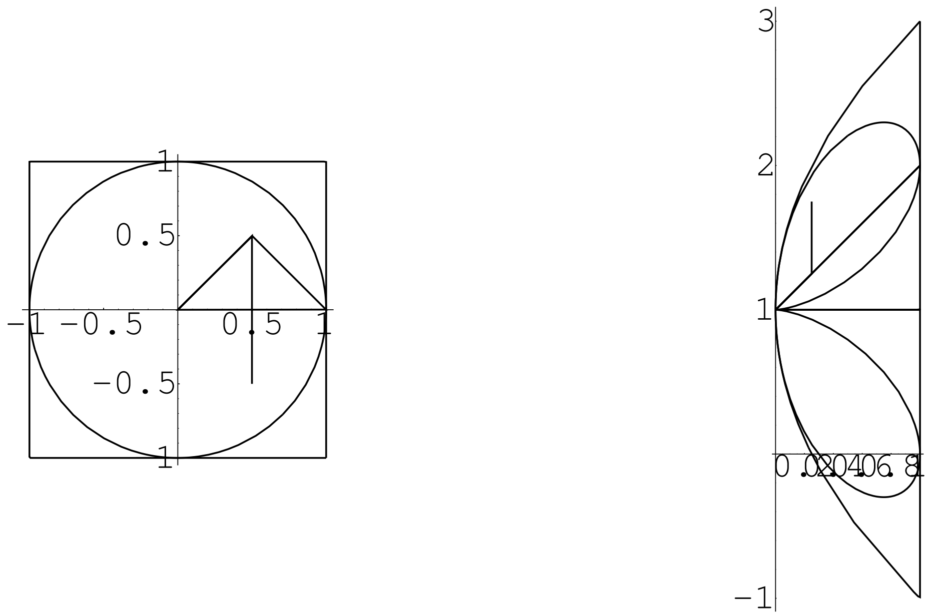


Figure 2.6: A nonlinear function

2.1.2 Operations on linear transformations

Next let us turn our attention to operations on linear transformations which give rise to new linear transformations. The first operation on linear transformations we will consider is composition. Recall that the composition of functions is defined by

$$f \circ g(x) = f(g(x)).$$

Theorem 2.1.1 *If L_1 and L_2 are linear transformations from $\mathbb{R}^2$ to itself, then $L_1 \circ L_2$ is a linear transformation as well.*

PROOF:

The function $L_1 \circ L_2$ from $\mathbb{R}^2$ to itself preserves addition and scalar multiplication because both L_1 and L_2 do:

$$\begin{aligned} L_1 \circ L_2(\vec{a} + \vec{b}) &= L_1(L_2(\vec{a} + \vec{b})) \\ &= L_1(L_2(\vec{a}) + L_2(\vec{b})) \\ &= L_1(L_2(\vec{a})) + L_1(L_2(\vec{b})) \end{aligned}$$

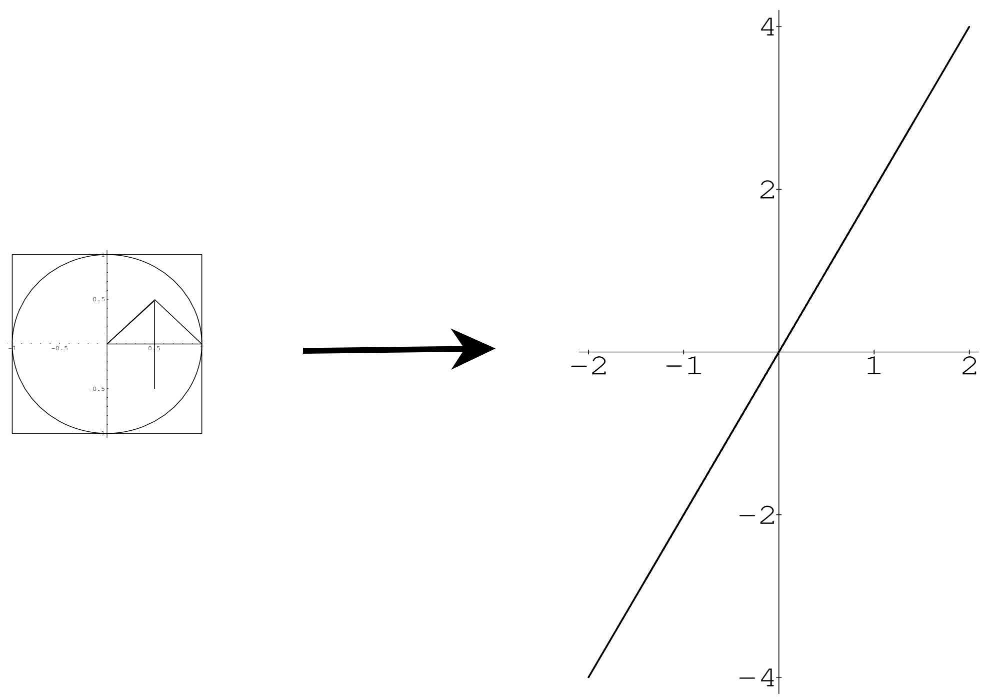


Figure 2.7: Collapsing onto a line

$$\begin{aligned}
&= L_1 \circ L_2(\vec{a}) + L_1 \circ L_2(\vec{b}) \\
L_1 \circ L_2(k\vec{a}) &= L_1(L_2(k\vec{a})) \\
&= L_1(kL_2(\vec{a})) \\
&= kL_1(L_2(\vec{a})) \\
&= kL_1 \circ L_2(\vec{a})
\end{aligned}$$

∎

We can also add linear transformations or multiply by a scalar to get another linear transformation.

Proposition 2.1.2 *If L_1 and L_2 are linear transformations and $k \in \mathbb{R}$, then $L_1 + L_2$, which has value*

$$(L_1 + L_2)(\vec{v}) = L_1(\vec{v}) + L_2(\vec{v}),$$

is linear and so is kL_1, where

$$(kL_1)(\vec{v}) = kL_1(\vec{v}).$$

Proof:

This is a simple calculation: First that the sum of linear transformations preserves sums:

$$\begin{aligned}
(L_1 + L_2)(\vec{a} + \vec{b}) &= L_1(\vec{a}) + L_2(\vec{a}) + L_1(\vec{b}) + L_2(\vec{b}) \\
&= L_1(\vec{a} + \vec{b}) + L_2(\vec{a} + \vec{b})
\end{aligned}$$

then that the sum of linear transformations preserves scalar multiples:

$$\begin{aligned}
(L_1 + L_2)(k\vec{a}) &= L_1(k\vec{a}) + L_2(k\vec{a}) \\
&= kL_1(\vec{a}) + kL_2(\vec{a}) = k(L_1 + L_2)(\vec{a}).
\end{aligned}$$

Similar calculations work for kL_1. ∎

Exercises 2.1:

Which of the maps in problems 1-10 are linear transformations? Either prove the function is a linear transformation or give an example which shows which aspect of linearity fails.

1. $L([x, y]) = [x + 3, y - 2]$
2. $L([x, y]) = [y, x]$
3. $L([x, y]) = [0, x - 2y]$
4. $L([x, y]) = [0, 0]$
5. $L([x, y]) = [x^2 - y^2, x^2 + y^2]$
6. $L([x, y]) = [3x - 4y, x + 7y]$
7. $L([x, y]) = [|x - y|, |y - x|]$
8. $L([x, y]) = [xy, x + 3y]$
9. $L([x, y]) = [ex + \pi y, 2x - y]$
10. $L([x, y]) = [52x + \frac{2}{3}y, x - 3y]$
11. Show that composition of linear transformations is not commutative by showing that if you rotate by $\frac{\pi}{4}$ radians and then stretch by a factor of 2 horizontally you do not get the same result as if you stretch horizontally by a factor of 2 and then rotate by $\frac{\pi}{4}$ radians.
12. Show that translation of axes $f : \mathbb{R}^2 \to \mathbb{R}^2$ with $f([a, b]) = [a+h, b+k]$ is not, in general, a linear transformation.
13. Show that linear transformations from $\mathbb{R}^2$ to itself map straight lines into straight lines or points. Recall that the vector $[x, y]$ is on the line through the point (a, b) in the direction $[d_1, d_2]$ if there is a $t \in \mathbb{R}$ with $[x, y] = [a, b] + t[d_1, d_2]$.
14. Show that the linear transformation $L([x, y]) = [ax + by, cx + dy]$ takes the unit square to a parallelogram with area $|ad-bc|$, this is the absolute value of the determinant of the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$.
15. Show that any linear transformation from $\mathbb{R}^2$ to itself can be written in the form $L([x, y] = [ax+by, cx+dy]$ for suitable choice of $a, b, c,$ and d.

2.2 Matrices and matrix operations

Matrices are rectangular arrays of numbers. Typically they are thought of as having horizontal rows and vertical columns. A matrix is called $m \times n$ if it has m rows and n columns. Each row has n elements in it and each column has m elements. In this section we will be talking about how we can use 2×2 matrices to make linear transformations more concrete.

2.2.1 Using ordered bases to get a matrix for a linear transformation

At the beginning of section 2.1 we noted that a linear transformation from $\mathbb{R}^2$ to itself is completely determined by what it does to the standard basis vectors $\vec{\imath}$ and $\vec{\jmath}$.

Example: Getting a linear transformation from $L(\vec{\imath})$ and $L(\vec{\jmath})$

Suppose that we know that $L(\vec{\imath}) = [1, 2]$ and $L(\vec{\jmath}) = [3, 4]$, then we can find

$$\begin{aligned}
L([a, b]) &= L(a\vec{\imath} + b\vec{\jmath}) \\
&= L(a\vec{\imath}) + L(b\vec{\jmath}) \\
&= aL(\vec{\imath}) + bL(\vec{\jmath}) \\
&= a[1, 2] + b[3, 4] \\
&= [a + 3b, 2a + 4b].
\end{aligned}$$

This tells us that all of the information we need to define L is the value at basis vectors. We can encode this information in a matrix and define multiplication so that it gives the value of the linear transformation. By convention we use column vectors and write the matrix on the left, mirroring the function notation familiar in calculus (though, like any convention, it is a bit artificial and has its drawbacks as well as its advantages). The calculation of $L([a, b])$ is then written as

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a + 3b \\ 2a + 4b \end{bmatrix}.$$

$\diamondsuit$

In general, suppose that we are given an ordered basis $B = (\vec{b}_1, \vec{b}_2)$ for $\mathbb{R}^2$. Then a linear transformation L is completely determined by its action on $\vec{b}_1$ and $\vec{b}_2$. Since any vector can $\vec{v}$ can be written as $k_1\vec{b}_1 + k_2\vec{b}_2$ for a unique pair of numbers (k_1, k_2) we can think of the column vector $\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$ as representing $\vec{v}$ with respect to the ordered basis (the B-coordinate representation of Definition 1.2.2).

If we represent $L(\vec{b}_1)$ with respect to this ordered basis we get a column vector $\begin{bmatrix} p \\ r \end{bmatrix}$ and similarly $L(\vec{b}_2) = \begin{bmatrix} q \\ s \end{bmatrix}$. Combining these column vectors gives a matrix $\begin{bmatrix} p & q \\ r & s \end{bmatrix}$ which contains all the information needed to find $L(\vec{v})$. When $\vec{v} = k_1\vec{b}_1 + k_2\vec{b}_2$ we get

$$L(\vec{v}) = k_1L(\vec{b}_1) + k_2L(\vec{b}_2) = (pk_1 + qk_2)\vec{b}_1 + (rk_1 + sk_2)\vec{b}_2,$$

a vector whose B-coordinates give the column vector $\begin{bmatrix} pk_1 + qk_2 \\ rk_1 + sk_2 \end{bmatrix}$. We define the product of the matrix $\begin{bmatrix} p & q \\ r & s \end{bmatrix}$ and the vector $\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$ so that it will give the value of the linear transformation:

Definition 2.2.1 *The product of a 2×2 matrix and a 2 element column vector is*

$$\begin{bmatrix} p & q \\ r & s \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} pk_1 + qk_2 \\ rk_1 + sk_2 \end{bmatrix}.$$

Note that if the first column of the matrix is the B-coordinate representation of $L(\vec{b}_1)$, the second column of the matrix is the B-coordinate representation of $L(\vec{b}_2)$, and we multiply by the column vector giving the B-coordinate representation of $\vec{v}$, we get the B-coordinate representation of $L(\vec{v})$ precisely because that is how we defined the operation of multiplication of a matrix times a column vector.

Let us now return to our examples of linear transformations and see what their matrices with respect to the standard basis $([1, 0], [0, 1])$ are:

Example: A matrix for $L([a, b]) = [a, 2b]$ using the standard basis

This linear transformation takes the basis vector $\vec{\imath}$ to $1\vec{\imath} + 0\vec{\jmath}$ thus the first column of the matrix for L with respect to the standard basis is

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Since L takes $\vec{\jmath}$ to $0\vec{\imath} + 2\vec{\jmath}$ the second column is

$$\begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

If we take the resulting matrix and multiply on the left by it we get

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \\ 2b \end{bmatrix}$$

$\diamondsuit$

Example:

Our more general example was $L([a, b]) = [2a - 3b, a + 4b]$. This takes $\vec{\imath}$ to $[2, 1] = 2\vec{\imath} + 1\vec{\jmath}$ and $\vec{\jmath}$ to $[-3, 4] = -3\vec{\imath} + 4\vec{\jmath}$, thus the matrix of L with respect to the standard basis is

$$\begin{bmatrix} 2 & -3 \\ 1 & 4 \end{bmatrix}.$$

$\diamondsuit$

Example: Using a non-standard basis

The standard basis is not always the most convenient basis to use in finding a matrix representation. As an example consider the linear transformation G which takes $[x, y]$ to $[x - 2y, x + 4y]$. Using reasoning like that used in the last example we find that its matrix with respect to the standard basis is

$$\begin{bmatrix} 1 & -2 \\ 1 & 4 \end{bmatrix}$$

If, however, we use the (carefully chosen) ordered basis $([1,-1],[-2,1])$ we get a nicer matrix. First we calculate

$$G[1,-1] = [3,-3] = 3[1,-1] + 0[-2,1]$$

and

$$G[-2,1] = [-4,2] = 0[1,-1] + 2[-2,1]$$

so that the matrix of G with respect to the basis $([1,-1],[-2,1])$ is

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}.$$

$\diamondsuit$

Thus two different matrices can represent the same linear transformation with respect to different ordered bases. These matrices are in some sense equivalent. The next definition gives a name to this kind of equivalence.

Definition 2.2.2 *If two matrices* $\mathbf{M}$ *and* $\mathbf{N}$ *represent the same linear transformation with respect to different choices of basis we say that they are* **similar matrices***. We write* $\mathbf{M} \sim \mathbf{N}$*.*

Example 2.2.1 shows that

$$\begin{bmatrix} 1 & -2 \\ 1 & 4 \end{bmatrix} \sim \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

One of the important problems of linear algebra is finding canonical forms for matrices. For example: when can you find a diagonal matrix which is similar to a given matrix? If you can't find a diagonal matrix, how close can you get? Are there other forms which are particularly useful or give significant information about the linear transformations they represent?

2.2.2 Matrix multiplication, addition, and multiplication by a scalar

We have seen that multiplication of a matrix times a column vector can be defined so that it represents evaluation of a linear transformation. Can we make definitions of other operations on matrices so that the other operations on linear transformations are represented?

The first important operation on linear transformations is composition: If L and M are linear transformations from $\mathbb{R}^2$ to itself, then so is $L \circ M$. Suppose that we pick an ordered basis for $\mathbb{R}^2$; how will the matrix for the composition $L \circ M$ be related to the matrices for L and M. To make things both concrete and general, suppose the ordered basis is $(\vec{b}_1, \vec{b}_2)$, the matrix for L is

$$\mathbf{L} = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}$$

and the matrix for M is

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$$

Then we can find the matrix for $L \circ M$ by finding what it does to the basis vectors:

$$\begin{aligned} L \circ M(\vec{b}_1) &= L(M(\vec{b}_1)) \\ &= L(m_{11}\vec{b}_1 + m_{21}\vec{b}_2) \\ &= m_{11}L(\vec{b}_1) + m_{21}L(\vec{b}_2) \\ &= m_{11}(l_{11}\vec{b}_1 + l_{21}\vec{b}_2) + m_{21}(l_{12}\vec{b}_1 + l_{22}\vec{b}_2) \\ &= (l_{11}m_{11} + l_{12}m_{21})\vec{b}_1 + (l_{21}m_{11} + l_{22}m_{21})\vec{b}_2 \end{aligned}$$

and similarly

$$\begin{aligned} L \circ M(\vec{b}_2) &= L(M(\vec{b}_2)) \\ &= L(m_{12}\vec{b}_1 + m_{22}\vec{b}_2) \\ &= m_{12}L(\vec{b}_1) + m_{22}L(\vec{b}_2) \\ &= m_{12}(l_{11}\vec{b}_1 + l_{21}\vec{b}_2) + m_{22}(l_{12}\vec{b}_1 + l_{22}\vec{b}_2) \\ &= (l_{11}m_{12} + l_{12}m_{22})\vec{b}_1 + (l_{21}m_{12} + l_{22}m_{22})\vec{b}_2 \end{aligned}$$

Thus the matrix for $L \circ M$ is

$$\begin{bmatrix} l_{11}m_{11} + l_{12}m_{21} & l_{11}m_{12} + l_{12}m_{22} \\ l_{21}m_{11} + l_{22}m_{21} & l_{21}m_{12} + l_{22}m_{22} \end{bmatrix}.$$

This prompts the definition of matrix multiplication:

Definition 2.2.3 *The* **product** *of the* 2×2 *matrices* $\mathbf{L}$ *and* $\mathbf{M}$ *is*

$$\mathbf{LM} = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} l_{11}m_{11} + l_{12}m_{21} & l_{11}m_{12} + l_{12}m_{22} \\ l_{21}m_{11} + l_{22}m_{21} & l_{21}m_{12} + l_{22}m_{22} \end{bmatrix}.$$

The first column of the product is obtained by taking the first matrix times the first column of the second matrix (as a column vector); the second column of the product is obtained by taking the first matrix times the second column. Without the connection to composition of linear transformations this would be a rather nonobvious way to multiply matrices!

Matrix multiplication has several nice properties, all of which are most easily related to properties of composition of linear transformations. For example, the **identity linear transformation** $Id : \mathbb{R}^2 \to \mathbb{R}^2$ takes $[x, y]$ to itself. It has the property that $L \circ Id = Id \circ L = L$. The matrix corresponding to this, an **identity matrix** has the form

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and has the property that $\mathbf{IM} = \mathbf{M}$ and $\mathbf{MI} = \mathbf{M}$.

Let us look next at the matrix for $L + M$. Again we start by seeing what $L + M$ does to the basis elements:

$$\begin{aligned} (L + M)(\vec{b}_1) &= L(\vec{b}_1) + M(\vec{b}_1) \\ &= l_{11}\vec{b}_1 + l_{21}\vec{b}_2 + m_{11}\vec{b}_1 + m_{21}\vec{b}_2 \\ &= (l_{11} + m_{11})\vec{b}_1 + (l_{21} + m_{21})\vec{b}_2 \end{aligned}$$

and

$$\begin{aligned} (L + M)(\vec{b}_2) &= L(\vec{b}_2) + M(\vec{b}_2) \\ &= l_{12}\vec{b}_1 + l_{22}\vec{b}_2 + m_{12}\vec{b}_1 + m_{22}\vec{b}_2 \\ &= (l_{12} + m_{12})\vec{b}_1 + (l_{22} + m_{22})\vec{b}_2 \end{aligned}$$

so the matrix for the sum is

$$\mathbf{L} + \mathbf{M} = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} + \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} l_{11} + m_{11} & l_{12} + m_{12} \\ l_{21} + m_{21} & l_{22} + m_{22} \end{bmatrix}$$

In a similar fashion we can get multiplication of a matrix by a scalar corresponding to multiplication of a linear transformation by a scalar:

$$k\mathbf{L} = \begin{bmatrix} kl_{11} & kl_{12} \\ kl_{21} & kl_{22} \end{bmatrix}$$

Exercises 2.2:
For problems 1-8

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ -3 & 5 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} -2 & 1 \\ 1 & 7 \end{bmatrix} \text{ and } \mathbf{C} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

find the following:

1. $\mathbf{AC}$
2. $\mathbf{CA}$
3. $\mathbf{BC}$
4. $(\mathbf{BC})\mathbf{A}$
5. $\mathbf{B}(\mathbf{CA})$
6. $\mathbf{A} + \mathbf{B}$
7. $(\mathbf{A} + \mathbf{B})\mathbf{C}$
8. $\mathbf{AC} + \mathbf{BC}$
9. Give the matrices for the linear transformation $K([x, y]) = [3x+y, y-x]$ with respect to the following basis (on both domain and codomain):
   (a) The standard basis
   (b) $([1, -1], [1, 1])$
   (c) $([1, 1], [0, 1])$
   (d) $([2, 3], [-5, 2])$
10. Give the matrices for the linear transformation $M([x, y]) = [-y, x]$ with respect to the following basis (on both domain and codomain):

(a) The standard basis
(b) $([1,-1],[1,1])$
(c) $([1,1],[0,1])$
(d) $([2,3],[-5,2])$

11. Give the matrices for the linear transformation $P([x,y]) = [2x+3y, -5x-8y]$ with respect to the following basis (on both domain and codomain):

    (a) The standard basis
    (b) $([1,-1],[1,1])$
    (c) $([1,1],[0,1])$
    (d) $([2,3],[-5,2])$

12. Give the matrices for the linear transformation $L([x,y]) = [x-2y, 0]$ with respect to the following basis (on both domain and codomain):

    (a) The standard basis
    (b) $([1,-1],[1,1])$
    (c) $([1,1],[0,1])$
    (d) $([2,3],[-5,2])$

13. Show that if the columns of a matrix $\mathbf{M}$ form a basis for $\mathbb{R}^2$ then $\mathbf{M}$ is the matrix of an onto linear transformation. (Recall that a function is onto if its image is all of its codomain.)

14. Show that the matrix
$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
is an identity for the operation matrix multiplication: that is, show $\mathbf{IM} = \mathbf{MI} = \mathbf{M}$ for all matrices $\mathbf{M}$.

15. Give an example which shows that matrix multiplication is not commutative.

16. Prove that matrix multiplication is associative. (Hint: composition of functions is associative.)

17. Show that the set of all 2×2 matrices is a vector space using the operations $+$ and scalar multiplication defined in this section; that is, show that matrices have the same algebraic properties (as in Proposition 1.1.2) that $\mathbb{R}^2$ has. Conclude that the space of linear transformations from $\mathbb{R}^2$ to itself is also a vector space.

2.3 Kernels, Images, and Subspaces

In this section we will consider some vector spaces smaller than $\mathbb{R}^2$ which can be gotten by looking at susbsets which are themselves vector spaces.

Definition 2.3.1 *A subset S of a vector space $\mathcal{V}$ which is itself a vector space using the same operations as in $\mathcal{V}$ is called a* **subspace**. *We will write $\mathcal{S} < \mathcal{V}$ to distinguish subspaces from subsets, which we write $S \subset \mathcal{V}$.*

Example: The x-axis

The set $\{[x, 0] | x \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$. To see this we need to check 10 axioms:

1. Closure under both addition and multiplication by a scalar: If we add two vectors of the form $[x, 0]$ we get a vector of that form and if we multiply such a vector by a real number we get another vector with a second coordinate of 0.
2. Commutativity of $+$:

$$\begin{aligned} [a, 0] + [b, 0] &= [a + b, 0] \\ &= [b + a, 0] \\ &= [b, 0] + a, 0] \end{aligned}$$

3. Associativity of $+$:

$$\begin{aligned} [a, 0] + ([b, 0] + [c, 0]) &= [a + (b + c), 0] \\ &= [(a + b) + c, 0] \\ &= ([a, 0] + [b, 0]) + [c, 0]. \end{aligned}$$

4. Identity: $[0, 0]$ is of the desired form.

5. Inverses: $[-a, 0]$ is of the desired form.
6. Absorption: $k(h[a,0]) = [k(ha), 0] = [(kh)a, 0] = kh[a, 0]$
7. Distribtive laws (both of them): use the distributive laws for $\mathbb{R}$ in the first coordinate, the second coordinate is always 0.
8. Identity for scalars: $1[a, 0] = [a, 0]$.

$\diamondsuit$

In general, to show that a set equipped with two operations is a vector space you need to show that all the properties in Definition 1.1.1 hold. For subspaces, however, it is not necessary to check all of the axioms for a vector space.

Theorem 2.3.1 *A nonempty subset $S \subset \mathcal{V}$ which is closed under addition and multiplication by scalars is a subspace of $\mathcal{V}$.*

PROOF:

The operations of addition and multiplication by a scalar will automatically satisfy all of the axioms of vector spaces given by equations, since those equations hold in the larger vector space $\mathcal{V}$ our subset S is contained in. Thus in addition to closure, we get the commutative, associative, absorption, identity for scalar multiplication, and distributive laws free. Every subspace must contain the zero vector, since that is part of the data for a vector space; we get the zero vector by multiplying the vector we know is in S by 0. Inverses come from scalar multiplication by -1. ∎

If a subspace of $\mathbb{R}^2$ contains a non-zero vector $\vec{a}$ it must also contain the line through the origin given by all the multiples of $\vec{a}$. In $\mathbb{R}^2$ the only subspaces lines through the origin, just the origin, or all of $\mathbb{R}^2$.

Example: The set $\{[x, 2x] | x \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.

To see this we need only show that $[0, 0]$ is in $\{[x, 2x]|x \in \mathbb{R}\}$ and that $\{[x, 2x]|x \in \mathbb{R}\}$ is closed under both addition and scalar multiples. Now

$$[0, 0] = [0, 2(0)]$$

so $[0,0]$ is of the right form. If we add we get another pair of the right form:

$$[x,2x]+[y,2y]=[x+y,2x+2y]=[x+y,2(x+y)]$$

and similarly a constant multiple of a vector of the form $[x,2x]$ is also of that form:

$$k[x,2x]=[kx,k(2x)]=[kx,2(kx)]$$

$\diamondsuit$

In general if a subspace contains vectors $\vec{a}_1\dots\vec{a}_n$ it must contain all vectors $\vec{b}$ which can be obtained from $\vec{a}_1\dots\vec{a}_n$ using sums and multiplication by scalars. These are called **linear combinations** of the given vectors and the set of all linear combinations of the vectors $\vec{a}_1,\dots,\vec{a}_n$ is called the **subspace spanned** by $\{\vec{a}_1,\dots,\vec{a}_n\}$, which we will denote $\mathrm{Span}(\{\vec{a}_1,\dots,\vec{a}_n\})$.

Definition 2.3.2 *A linear combination of the vectors $\vec{a}_1\dots\vec{a}_n$ is any vector of the form*

$$\vec{b}=\sum_{i=1}^{n}k_i\vec{a}_i$$

. For any finite set of vectors $S=\{\vec{a}_1\dots\vec{a}_n\}$, the set of linear combinations of vectors in S is Span(S)

Proposition 2.3.2 *For any finite set of vectors $S=\{\vec{a}_1\dots\vec{a}_n\}$, the set Span$(S)$ of linear combinations of vectors in S is a subspace.*

PROOF:

We need to show that $\mathrm{Span}(S)$ is non-empty and that it is closed under both operations. Notice that $\vec{a}_1\in\mathrm{Span}(S)$ (by letting $k_1=1$ and all other $k_i=0$), giving non-emptiness. Closure under addition comes from

$$\sum_{i=1}^{n}k_i\vec{a}_i+\sum_{i=1}^{n}h_i\vec{a}_i=\sum_{i=1}^{n}(k_i+h_i)\vec{a}_i$$

and closure under multiplication by scalars comes from

$$r\sum_{i=1}^{n}k_i\vec{a}_i=\sum_{i=1}^{n}(rk_i)\vec{a}_i.$$

∎

Later on we will extend this concept to the span of arbitrary sets of vectors. Since the smallest possible vector space consists of just the zero vector, we define $\text{Span}(\emptyset) = \{\vec{0}\}$.

Example: What is the subspace of $\mathbb{R}^2$ spanned by the set $\{[1,2],[2,4]\}$?

A vector $[a,b]$ will be in $\text{Span}(\{[1,2],[2,4]\})$ if and only if it can be written as $[a,b] = x[1,2] + y[2,4]$. This is the same as asking for a and b for which solutions exist for the system of equations

$$\begin{aligned} x + 2y &= a \\ 2x + 4y &= b \end{aligned}$$

It is clear that this is hopeless unless $b = 2a$. Thus $\text{Span}(\{[1,2],[2,4]\}) = \{[a,b] | b = 2a\}$. $\diamondsuit$

Example: What is the subspace of $\mathbb{R}^2$ spanned by the set $\{[1,2],[2,-4]\}$?

A vector $[a,b]$ will be in $\text{Span}(\{[1,2],[2,-4]\})$ if and only if it can be written as $[a,b] = x[1,2] + y[2,-4]$. This is the same as asking for a and b for which solutions exist for the system of equations

$$\begin{aligned} x + 2y &= a \\ 2x - 4y &= b \end{aligned}$$

This has solutions

$$\begin{aligned} x &= \frac{2a+b}{4} \\ y &= \frac{2a-b}{4} \end{aligned}$$

so $\text{Span}(\{[1,2],[2,-4]\}) = \mathbb{R}^2$. $\diamondsuit$

If $\text{Span}(S) = \mathcal{V}$ then we say that S is a **spanning set** for $\mathcal{V}$. Notice that the requirement on $(\vec{b}_1, \vec{b}_2)$ that every element $\vec{v}$ of $\mathbb{R}^2$ have a representation of the form $k_1\vec{b}_1 + k_2\vec{b}_2$ tells us that a basis must be a spanning set.

2.3.1 Images of linear transformations

In our study a linear transformations L from $\mathbb{R}^2$ (the domain) to $\mathbb{R}^2$ (the codomain) we sometimes get every element of the codomain as the image of an element in the domain (in which case L is onto) and sometimes we do not. For example, the linear transformation in Example 2.1.1, which collapsed $\mathbb{R}^2$ onto a line, is not onto; its image is smaller than its codomain. This suggests that the image of a linear transformation might be an interesting set to look at.

Definition 2.3.3 *The image of a linear transformation $L : \mathcal{V} \to \mathcal{W}$ is the subset $Im(L) = \{L(\vec{v}) | \vec{v} \in \mathcal{V}\} \subset \mathcal{W}$.*

Proposition 2.3.3 *The image of a linear transformation is a subspace of its codomain.*

Proof:

If $\vec{v}_1 \in \text{Im}(L)$ and $\vec{v}_2 \in \text{Im}(L)$, then there are vectors $\vec{u}_1$ and $\vec{u}_2$ with $L(\vec{u}_1) = \vec{v}_1$ and $L(\vec{u}_2) = \vec{v}_2$. Then $r\vec{v}_1 = L(r\vec{v}_1)$ and $\vec{v}_1 + \vec{v}_2 = L(\vec{u}_1 + \vec{u}_2)$. This tells us that both $r\vec{v}_1$ and $\vec{v}_1 + \vec{v}_2$ are in $\text{Im}(L)$. We know as well that $L(\vec{0}) = \vec{0}$, so the image also contains the zero vector. Since $\text{Im}(L)$ is nonempty and closed under both sum and multiplication by a scalar, it is a subspace of $\mathcal{W}$. ∎

Proposition 2.3.4 *If $\{\vec{b}_1, \vec{b}_2\}$ spans $\mathbb{R}^2$, then $\{L(\vec{b}_1), L(\vec{b}_2)\}$ spans $Im(L)$.*

Proof:

A vector $\vec{w}$ is in $\text{Im}(L)$ if it is of the form $L(\vec{v})$ for some $\vec{v} \in \mathbb{R}^2$. Now since $\{\vec{b}_1, \vec{b}_2\}$ spans $\mathbb{R}^2$, $\vec{v} = k_1\vec{b}_1 + k_2\vec{b}_2$ for some k_1, k_2. Then

$$\begin{aligned}
\vec{w} &= L(\vec{v}) \\
&= L(k_1\vec{b}_1 + k_2\vec{b}_2) \\
&= k_1 L(\vec{b}_1) + k_2 L(\vec{b}_2).
\end{aligned}$$

∎

In the last section we saw that once we have an ordered basis for $\mathbb{R}^2$ in mind, talking about linear transformations and talking about matrices is really the same thing. With this in mind recall that the columns of the matrix for a linear transformation were obtained by looking at the image of the basis vectors. Now the image of a basis gives a spanning set for the image of a linear transformation, so if we look at the subspace of $\mathbb{R}^2$ spanned by the columns of a matrix for L we will obtain the image of L. This subspace of $\mathbb{R}^2$ is called the **column space** of the matrix. Summarizing, we get the following definition:

Definition 2.3.4 *The* **column space** *of a matrix is the subspace spanned by its column vectors.*

If we represent a linear transformation with respect to a choice of ordered bases, then the column space of the matrix will give the image.

2.3.2 Kernels of linear transformations

The other question we ask about linear transformations is whether or not they are one-to-one. This is equivalent to asking whether you can conclude that $\vec{a} = \vec{b}$ from knowing that $L(\vec{a}) = L(\vec{b})$ or equivalently that

$$L(\vec{a}) - L(\vec{b}) = L(\vec{a} - \vec{b}) = 0.$$

This suggests that it might be useful to look at the set of vectors a linear transformation sends to $\vec{0}$.

Definition 2.3.5 *We call the set of vectors $\vec{v}$ such that $L(\vec{v}) = \vec{0}$ the kernel of L, written $Ker(L)$.*

Proposition 2.3.5 *The kernel of a linear transformation L is a subspace of the domain of L.*

PROOF:

It is clear that $\vec{0} \in \mathrm{Ker}(L)$, so $\mathrm{Ker}(L) \neq \emptyset$. If $\vec{a} \in \mathrm{Ker}(L)$ and $\vec{b} \in \mathrm{Ker}(L)$ then

$$L(\vec{a} + \vec{b}) = L(\vec{a}) + L(\vec{b}) = \vec{0} + \vec{0} = \vec{0}$$

Thus $\vec{a}+\vec{b} \in \mathrm{Ker}(L)$. Since $L(k\vec{a}) = kL(\vec{a}) = k\vec{0} = \vec{0}$, $k\vec{a} \in \mathrm{Ker}(L)$ as well, giving closure under both operations. Thus $\mathrm{Ker}(L)$ is a subspace of the domain of L. ∎

In terms of the matrix for L with respect to a basis, the kernel will be exactly the set of all solutions to the pair of equations given by the matrix equation

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

A system of equations in which all of the constants are 0 is called homogeneous. Thus we conclude that the set of solutions to a homogeneous system of linear equations is a subspace. We also observe that if we want to get our hands on the kernel of a linear transformation what we really need to know how to do is describe the solutions to such systems. For systems of two equations in two unknowns most of us can muddle through without much difficulty. Clearly for larger systems we will need some systematic approach.

2.3.3 Solution of systems of equations in two unknowns

For two by two matrices we can find the solution to the system

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} m \\ n \end{bmatrix}.$$

by solving the first equation for x

$$x = \frac{m - by}{a}$$

and then substituting in the second equation:

$$\begin{aligned} c\frac{m - by}{a} + dy &= n \\ (ad - cb)y &= an - cm \\ y &= \frac{an - cm}{ad - bc} \end{aligned}$$

Then we can substitute back into the expression for x to get

$$x = \frac{md - bn}{ad - bc}.$$

This tells us that there will be a unique solution if the number $ad - bc \neq 0$.

If $ad - bc = 0$ then $\frac{a}{c} = \frac{b}{d}$. In this case our system has solutions only if $n = \frac{a}{c}m$ in which case the second equation gives no new information, so there

are an infinite number of solutions. If $n \neq \frac{a}{c}m$ then the system describes two distinct parallel lines and there are no solutions.

We have encountered the expression $ad - bc$ before (in the exercises on linear transformations, Exercises 2.1#6, where it was related to areas in the plane). Thus we will give it a name:

Definition 2.3.6 *The* **determinant** *of a* 2×2 *matrix is given by the formula*

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

We note that a system of two equations in two unknowns has a unique solution if and only if the determinant of the matrix of coefficients is non-zero.

We can use $\text{Ker}(L)$ to find out if L is one-to-one:

Proposition 2.3.6 *The linear transformation L is one-to-one if and only if Ker*$(L) = \{\vec{0}\}$*.*

PROOF:

Certainly if $\text{Ker}(L)$ contains two distinct vectors, then L cannot be one-to-one, since two vectors go to $\vec{0}$. On the other hand, if $\text{Ker}(L) = \{\vec{0}\}$ and $L(\vec{a}) = L(\vec{b})$ then

$$L(\vec{a} - \vec{b}) = L(\vec{a}) - L(\vec{b}) = \vec{0}$$

so $\vec{a} - \vec{b} \in \text{Ker}(L)$, so it must be $\vec{0}$. This makes $\vec{a} = \vec{b}$, so L is one-to-one. ∎

Corollary 2.3.7 *A linear transformation is one to one if and only if the determinant of any matrix representing it is non-zero.*

PROOF:

By Proposition 2.3.6 a linear transformation is one-to-one if and only if its kernel consists only of the zero vector. If the matrix for our linear transformation with respect to some ordered basis $(\vec{b}_1, \vec{b}_2)$ is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then a vector $x\vec{b}_1 + y\vec{b}_2$ is in the kernel if and only if x and y give a solution to

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This system has a unique solution if and only if the determinant $ad - bc \neq 0$. ∎

Example: Finding image and kernel of a linear transformation

Let us apply these ideas to find the image and kernel of the linear transformation L which takes $[x, y]$ to $[x - y, 2x - 2y]$. The matrix for this linear transformation with respect to the standard bases is

$$\begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix}.$$

This matrix has determinant $(1 \cdot -2) - (-1 \cdot 2) = 0$, so this linear transformation is not one-to-one.

The image is spanned by the set of vectors

$$\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ -2 \end{bmatrix} \right\}.$$

Now, the second vector is just -1 times the first, so we can identify this as the line through the origin given by all multiples of the first vector. A more familiar form for this line is as the set of points (x, y) satisfying the equation $y = 2x$.

To find the kernel we need to solve the system of equations

$$\begin{aligned} x - y &= 0 \\ 2x - 2y &= 0. \end{aligned}$$

This is easy. The set of solutions is spanned by $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ and is recognizable as the line with equation $y = x$. ♢

Exercises 2.3:

For problems 1 to 8, describe the subspaces spanned by the following sets

of vectors: (Since subspaces are either $\{[0,0]\}$, lines through the origin, or all of $\mathbb{R}^2$, these should be the kinds of answers you give.)

1. $\{[0,0]\}$
2. $\{[1,0]\}$
3. $\{[1,2]\}$
4. $\{[1,3],[-2,-6]\}$
5. $\{[1,3],[1,2]\}$
6. $\{[1,2],[2,4],[-3,-6]\}$
7. $\{[1,2],[-2,4],[3,-6]\}$
8. $\{[2,4],[-1,3],[3,4],[2,5]\}$
9. Find the kernel of the following linear transformations:
   (a) $L_1([x,y]) = [x+y, 2x+2y]$
   (b) $L_2([x,y]) = [x+2y, 2x+y]$
   (c) $L_3([x,y]) = [0, 2x+y]$
   (d) $L_1([x,y]) = [0,0]$
10. Find the image of the following linear transformations:
    (a) $L_1([x,y]) = [x+y, 2x+2y]$
    (b) $L_2([x,y]) = [x+2y, 2x+y]$
    (c) $L_3([x,y]) = [0, 2x+y]$
    (d) $L_1([x,y]) = [0,0]$
11. Give examples of linear transformations with the following subspaces as kernel:
    (a) $\{[0,0]\}$
    (b) $\mathbb{R}^2$
    (c) $\{[x,3x] | x \in \mathbb{R}\}$

12. Give examples of linear transformations with the following subspaces as image:

    (a) $\{[0, 0]\}$
    (b) $\mathbb{R}^2$
    (c) $\{[x, 3x] | x \in \mathbb{R}\}$

    For problems 11-20 find the determinant and identify which represent linear transformations which are one-to-one.

13. $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$

14. $\begin{bmatrix} 2 & 1 \\ 0 & .1 \end{bmatrix}$

15. $\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$

16. $\begin{bmatrix} .2 & .3 \\ -.3 & .2 \end{bmatrix}$

17. $\begin{bmatrix} .2 & 1 \\ 0 & .2 \end{bmatrix}$

18. $\begin{bmatrix} .2 & 1 \\ 0 & 0 \end{bmatrix}$

19. $\begin{bmatrix} 2 & 3 \\ -1 & 1 \end{bmatrix}$

20. $\begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}$

21. $\begin{bmatrix} 4 & -2 \\ 1 & 3 \end{bmatrix}$

22. $\begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$

23. (Project Problem) In our later treatment of determinants we will ask for certain properties and then show that there is only one function satisfying those properties. For each of the following properties, give a numerical example and then a proof using 2 by 2 matrices with variables as entries:

    (a) The determinant of the product of two matrices is the product of their determinants.

    (b) The determinant of the transpose of a matrix is the same as the determinant of the matrix. The transpose switches rows and columns:
    $$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^t = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

    (c) If you switch two rows of a matrix you multiply the determinant by -1.

    (d) Multiplying a row of a matrix by a constant multiplies the determinant by the same constant.

    (e) Any matrix with two identical rows has determinant 0.

2.4 Finding inverses

In the last section we saw that matrix multiplication represents composition of linear transformations. The inverse of a matrix will correspond to the inverse under composition of the linear transformation it represents.

Definition 2.4.1 *The inverse of the matrix* $\mathbf{M}$, *if it exists, is the matrix* $\mathbf{M}^{-1}$ *such that* $\mathbf{M}\mathbf{M}^{-1} = \mathbf{M}^{-1}\mathbf{M} = \mathbf{I}$. *The inverse of the linear transformation* L, *if it exists, is a linear transformation* L^{-1} *such that* $L^{-1} \circ L = Id = L \circ L^{-1}$.

Finding the inverse of a matrix is equivalent to finding the inverse of a linear transformation and, in general, it takes some work. Inverses do not always exist so it is an interesting problem to discover how to tell if the inverse exists or not. Determinants provide an answer, almost as if by magic. However, for large matrices it takes almost as much work to find the determinant as it does to find the inverse of a matrix.

A function has an inverse (under composition) if it is one to one and onto. We looked at conditions on the matrix which guarantee that the linear transformation is one to one (having $Ker(L) = \{\vec{0}\}$ is equivalent to the system of equations $\mathbf{L}\vec{x} = \vec{0}$ having only the trivial solution). We also looked at how to recognize onto linear transformations by noticing that the column space of a matrix is the image of the linear transformation it represents.

Finding the inverse (under matrix multiplication) of a matrix will give us an expression for the inverse of the linear transformation it represents. Finding such an inverse actually boils down to finding the solution to two systems of two equations, with constant terms given by the columns of the identity matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

This gives us the formula

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}.$$

Thus we can find the inverse if the determinant is not 0. If the determinant is 0, then we will not be able to find an inverse. While for the 2×2 case this isn't bad, in general this is a lousy way to find the inverse of a matrix because for larger systems it has a very high operation count compared with other methods.

In this section we will see how to find the inverse of a matrix using *elementary row operations.*

There are three kinds of elementary row operations: interchanging two rows, multiplying a row by a constant, and adding a constant multiple of one row to another. For example interchanging the first and second rows of

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

gives

$$\begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix};$$

multiplying the second row by 3 then gives

$$\begin{bmatrix} 3 & 4 \\ 3 & 6 \end{bmatrix},$$

and then adding -1 times the first row to the second gives

$$\begin{bmatrix} 3 & 4 \\ 0 & 2 \end{bmatrix}.$$

We would write this whole sequence as

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \overset{R_1 \leftrightarrow R_2}{\rightsquigarrow} \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} \overset{3R_2}{\rightsquigarrow} \begin{bmatrix} 3 & 4 \\ 3 & 6 \end{bmatrix} \overset{R_2 - R_1}{\rightsquigarrow} \begin{bmatrix} 3 & 4 \\ 0 & 2 \end{bmatrix}$$

Here we have labeled the row operations using a reasonably common convention. Clearly we can use elementary row operations to change the form of a matrix; systematic use of row operations can reduce matrices to particularly nice forms.

Notice that each of the row operations can be accomplished by multiplying on the left by a matrix:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 3 & 6 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 0 & 2 \end{bmatrix}$$

The matrix which accomplishes the row operation is obtained by performing the row operation on an identity matrix.

Each row operation can be undone: to undo $R_1 \leftrightarrow R_2$ do it again; to undo $3R_2$ do $\frac{1}{3}R_2$; to undo $R_2 - R_1$ do $R_2 + R_1$. Thus each of the matrices which accomplishes a row operation must have an inverse; the matrix for the row operation which undoes it.

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

That means any sequence of elementary row operations can be accomplished by multiplying on the left by a matrix (which has an inverse) found by applying the sequence of row operations to the identity matrix. Our algorithm for finding inverses of matrices will use elementary row operations to make our matrix into an identity (if possible). We keep track of what we've done by performing the same row operations on an identity matrix.

Example: Find the inverse of $\begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}$

First augment the matrix by adjoining an identity matrix to keep track of our work and then do the following row operations:

$$\begin{aligned}
\begin{bmatrix} 2 & 4 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{bmatrix} \quad & \underset{\rightsquigarrow}{R_1 \leftrightarrow R_2} \quad \begin{bmatrix} 1 & 3 & 0 & 1 \\ 2 & 4 & 1 & 0 \end{bmatrix} \\
& \underset{\rightsquigarrow}{R_2 - 2R_1} \quad \begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & -2 & 1 & -2 \end{bmatrix} \\
& \underset{\rightsquigarrow}{-\tfrac{1}{2}R_2} \quad \begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & 1 & -\frac{1}{2} & 1 \end{bmatrix} \\
& \underset{\rightsquigarrow}{R_1 - 3R_2} \quad \begin{bmatrix} 1 & 0 & \frac{3}{2} & -2 \\ 0 & 1 & -\frac{1}{2} & 1 \end{bmatrix}
\end{aligned}$$

We conclude that the inverse of

$$\begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \text{ is } \begin{bmatrix} \frac{3}{2} & -2 \\ -\frac{1}{2} & 1 \end{bmatrix}.$$

$\diamondsuit$

Example: Find the inverse of the linear transformation $L([a, b]) = [2a + 4b, a + 3b]$

The matrix for this linear transformation with respect to the standard ordered basis is

$$\begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix}$$

which we just showed has inverse

$$\begin{bmatrix} \frac{3}{2} & -2 \\ -\frac{1}{2} & 1 \end{bmatrix}.$$

Thus the inverse of L is given by $L^{-1}([a,b]) = [\frac{3}{2}a - 2b, -\frac{1}{2}a + b]$, obtained by multiplying

$$\begin{bmatrix} \frac{3}{2} & -2 \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \frac{3}{2}a - 2b \\ -\frac{1}{2}a + b \end{bmatrix}.$$

$\diamondsuit$

This algorithm may be summarized in terms of its strategy: first use interchange or division of a row to get a 1 on the main diagonal (the one from upper left to lower right) then use that row to get 0 in the rest of its column. The tactics are the use of elementary row operations. We decide what to do by looking at the left hand side of the augmented matrix. The right hand side keeps track of the row operations we have performed.

If you reach an impasse in attempting to use this algorithm, then the matrix you started with does not have an inverse. The impasse shows up as a zero on the diagonal which you cannot get rid of without messing up your previous work.

Example: The matrix $\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$ does not have an inverse.

Again we try our algorithm:

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix} \quad \overset{R_2 - R_1}{\rightsquigarrow} \quad \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

at which point we are stuck. There is no way we can get the 1 on the lower right corner. We conclude that there is no inverse. $\diamondsuit$

Exercises 2.4:
For problems 1-5 find the inverse, if it exists, of the matrix

1. $\begin{bmatrix} 2 & 3 \\ 4 & 8 \end{bmatrix}$

2. $\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$

3. $\begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix}$

4. $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$

5. $\begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix}$

For problems 6-10 find the inverse, if it exists, of the linear transformation

6. $L([x, y]) = [2x + 4y, 3x + 8y]$

7. $L([x, y]) = [2x + 4y, 3x + 6y]$

8. $L([x, y]) = [x + 2y, -x + 2y]$

9. $L([x, y]) = [x - y, -x + y]$

10. $L([x, y]) = [2x + 4y, 2x + 4y]$

11. Show that if a matrix has an inverse, then the inverse is unique: if $\mathbf{MN}_1 = \mathbf{I}$ and $\mathbf{N}_2\mathbf{M} = \mathbf{I}$ then $\mathbf{N}_1 = \mathbf{N}_2$.

12. Show that if $\mathbf{M}$ and $\mathbf{N}$ have inverses, then so does $\mathbf{MN}$.

13. Show that if a linear transformation has an inverse as a function, then that inverse is a linear transformation.

14. What happens if you multiply on the right by an elementary matrix instead of multiplying on the left?

15. One way of deciding which of several algorithms is best for solving a kind of problems is to calculate worst case computational complexity. One way to do this in linear algebra involves counting the number of multiplications (or divisions) needed to carry out an algorithm (addition and some other manipulations involved in matrix algorithms are much less costly in terms of computation time unless the matrices involved are so large that memory management considerations dominate the problem). Count how many multiplication operations are used in the algorithm in this section for finding the inverse of a 2 by 2 matrix. Compare with the 6 multiplications needed if you use the formula from the determinant.

Chapter 3

Eigenvalues in $\mathbb{R}^2$

3.1 Dynamics of iteration: introduction to eigenvalues

One of the main uses of linear algebra comes in the description of linear systems, either using linear differential equations or linear difference equations. Often this is the first step in understanding the behavior of nonlinear systems as well. In this section we will consider the dynamics of iteration for linear transformations in the plane, asking what happens when we start with a nonzero vector and then repeatedly apply L. We are asking what happens to $L^n(\vec{a})$ as n gets large.

The easiest kind of vectors to follow are the **eigenvectors** of L because iteration only changes their length:

Definition 3.1.1 *A number λ is called an* **eigenvalue** *for L if there is a nonzero vector $\vec{a}$ (called an* **eigenvector** *) with $L(\vec{a}) = \lambda\vec{a}$.*

Now suppose we are interested in the behavior of iterates of an eigenvector with eigenvalue λ. Since

$$\begin{aligned}
L(\vec{a}) &= \lambda\vec{a} \\
L^2(\vec{a}) &= L(L(\vec{a})) \\
&= L(\lambda\vec{a}) \\
&= \lambda^2\vec{a} \\
L^n(\vec{a}) &= \lambda^n\vec{a}.
\end{aligned}$$

If $\lambda > 1$ then the length grows with each iteration: the values are headed straight out to infinity. If $-1 < \lambda < 1$ then the length is shrinking with each iteration and the iterates are approaching zero. A negative eigenvalue leads to a change of direction on each iteration, hence to oscillations. Large negative eigenvalues lead to divergent oscillations. We will return to examples illustrating this behavior after we find out how to find eigenvalues and their corresponding eigenvectors.

The first thing we do is shift our attention from finding eigenvalues of a linear transformation L to finding eigenvalues of the matrix $\mathbf{L}$ which represents that transformation with respect to some choice of ordered basis B. A vector $\vec{a}$ which has B-coordinate representation $\vec{x}^t$ will have B-coordinates given by $\mathbf{L}\vec{x}^t$. If applying the linear transformation to a particular vector multiplies that vector by λ, then multiplying the corresponding B-coordinate column vector on the left by the matrix corresponding to the linear transformation will also have the effect of multiplying by λ. This changes the eigenvalue problem to that of finding a λ such that there is a non-trivial solution to the system of equations

$$\mathbf{L}\vec{x}^t = \lambda\vec{x}^t$$

or equivalently

$$(\mathbf{L} - \lambda\mathbf{I})\vec{x}^t = \vec{0}^t.$$

This will happen only when the matrix $\mathbf{L} - \lambda\mathbf{I}$ does not have an inverse, since if it does have an inverse the only answer for $\vec{x}$ is $\vec{0}$. When we discussed inverses of 2×2 matrices in Section 2.4 we noted that there is an inverse if and only if the determinant is nonzero. Thus we need

$$\det(\mathbf{L} - \lambda\mathbf{I}) = 0.$$

This gives a quadratic equation in λ and thus poses no particular difficulty (for other linear transformations the problem of finding eigenvalues is more difficult).

Definition 3.1.2 *The characteristic polynomial for a matrix* $\mathbf{L}$ *is* $\det((L) - \lambda\mathbf{I})$. *The characteristic equation is* $\det((L) - \lambda\mathbf{I}) = 0$.

Example:

The diagonal matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

has eigenvalues 1 and 2. We can find these by finding where

$$\det \begin{bmatrix} 1-\lambda & 0 \\ 0 & 2-\lambda \end{bmatrix} = (1-\lambda)(2-\lambda) = 0.$$

Obviously finding eigenvalues for diagonal matrices isn't much of a challenge; the eigenvalues are just the entries on the diagonal. $\diamondsuit$

Example:

In the Section 2.2.1, fourth example, we noted that the matrix

$$\begin{bmatrix} 1 & -2 \\ 1 & 4 \end{bmatrix}$$

could be diagonalized by choosing an appropriate basis. We can now see how that basis was found using eigenvalues. We start by finding the solution to

$$\det \begin{bmatrix} 1-\lambda & -2 \\ 1 & 4-\lambda \end{bmatrix} = (1-\lambda)(4-\lambda) + 2 = 0$$

This gives the characteristic equation

$$\lambda^2 - 5\lambda + 6 = 0,$$

so $\lambda = 3$ and $\lambda = 2$ are the eigenvalues.

Next let us look for an eigenvector for the eigenvalue 3. We want x and y so that

$$\begin{bmatrix} 1 & -2 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 3 \begin{bmatrix} x \\ y \end{bmatrix}$$

This gives two equations in two unknowns

$$\begin{aligned} x - 2y &= 3x \\ x + 4y &= 3y \end{aligned}$$

which becomes

$$\begin{aligned} -2x - 2y &= 0 \\ x + y &= 0 \end{aligned}$$

when we put all the variables on the left. This is not enough to specify both variables, since the second equation gives no new information. Any pair with $x = -y$ will satisfy both equations. In our diagonalization we used $[1, -1]$ for the eigenvector for 3.

Similarly for the eigenvalue $\lambda = 2$ we get the system

$$\begin{aligned} x - 2y &= 2x \\ x + 4y &= 2y \end{aligned}$$

which simplifies to the single equation $x + 2y = 0$. The vector $[2, -1]$ satisfies this equation and hence is an eigenvector for the eigenvalue 2.

Since $[1, -1]$ and $[2, -1]$ are not collinear, they form a basis for $\mathbb{R}^2$. As noted in Example 2.2.4, the matrix for L with respect to the ordered basis $([1, -1], [2, -1])$ is

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}.$$

$\diamondsuit$

A matrix with real entries need not have real eigenvalues. It is possible for the characteristic equation $\det(\mathbf{M} - \lambda\mathbf{I}) = 0$ to have complex roots. The rotation example we looked at in Example 2.1.3 is typical:

Example:

A rotation of the plane through $\frac{\pi}{4}$ radians has the matrix with respect to the standard ordered basis

$$\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

The characteristic equation is

$$\det \begin{bmatrix} \frac{\sqrt{2}}{2} - \lambda & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} - \lambda \end{bmatrix} = 0$$

or $1-\sqrt{2}\lambda+\lambda^2=0$. This equation has no real roots, however, it does have the complex roots $\lambda = \frac{\sqrt{2}}{2} \pm \frac{\sqrt{2}}{2}i$. Later on we will see how to find vectors in $\mathbb{C}^2$ which are eigenvectors corresponding to these complex eigenvalues. ♢

Let us now look at some examples that show how the behavior of an iterative dynamical system can vary depending on the kind and size of eigenvalues:

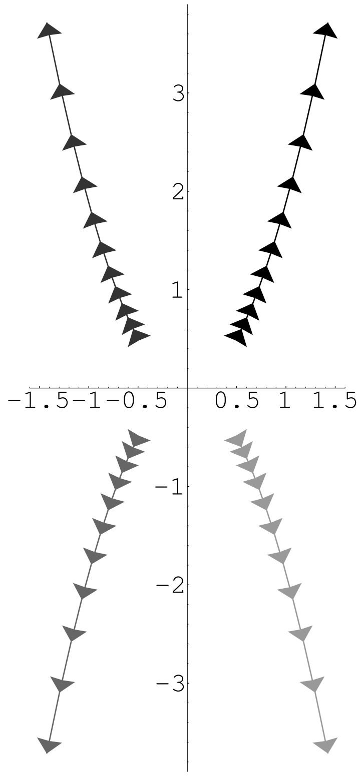
3.1.1 Behavior when the eigenvalues are both real

Example: Both λ_1 and $\lambda_2 > 1$

An example here is the linear transformation with matrix

$$\begin{bmatrix} 1.1 & 0 \\ 0 & 1.2 \end{bmatrix}$$

When we iterate this map points run away from the origin. Here is a graph showing the paths taken by some selected points.



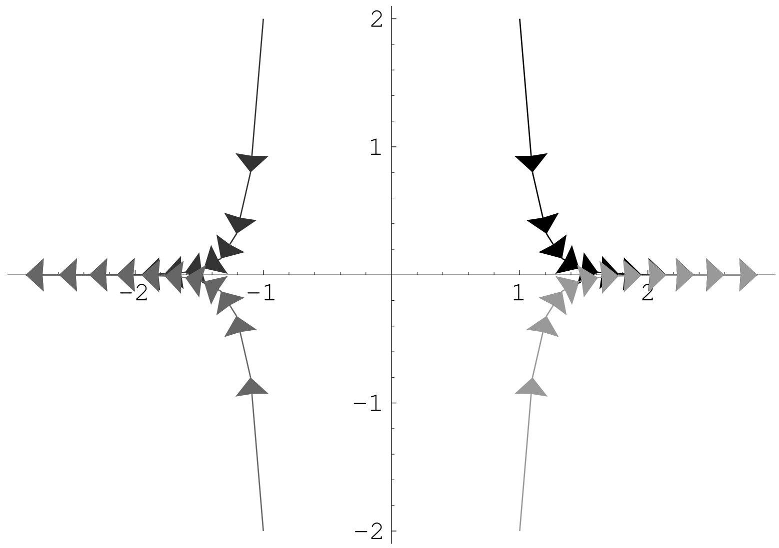
$\diamondsuit$

Example: $\lambda_1 > 1$ **and** $0 < \lambda_2 < 1$

Let's use the matrix

$$\begin{bmatrix} 1.1 & 0 \\ 0 & .4 \end{bmatrix}$$

for this example. We expect that points will shrink fairly rapidly in the y direction while growing in the x direction as shown in this graph:



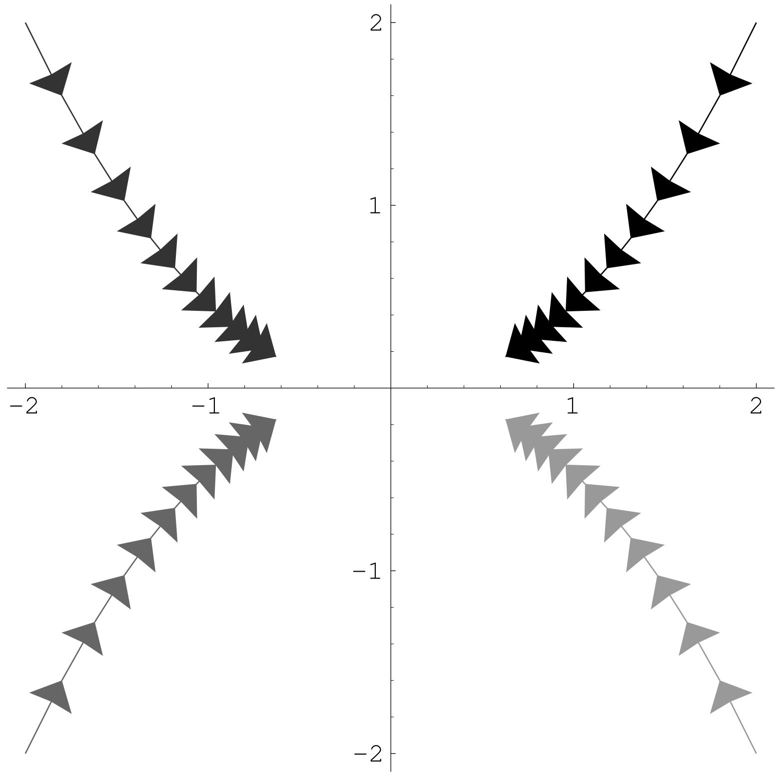
◊

Example: Both λ_1 and λ_2 strictly between 0 and 1

An example here has matrix

$$\begin{bmatrix} .9 & 0 \\ 0 & .8 \end{bmatrix}$$

In this case all points approach the origin as shown in this graph:



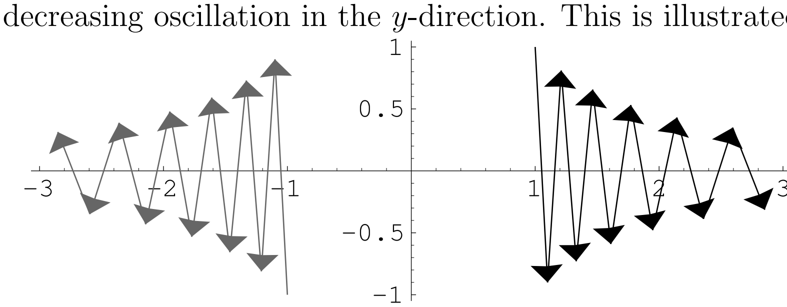
◊

Example: One positive and one negative eigenvalue

Here let us look at the dynamics of the linear transformation with matrix

$$\begin{bmatrix} 1.1 & 0 \\ 0 & -.9 \end{bmatrix}$$

This system shows growth in the direction of the x-axis and a decreasing oscillation in the y-direction. This is illustrated in



◊

3.1.2 Dynamics of systems with complex eigenvalues

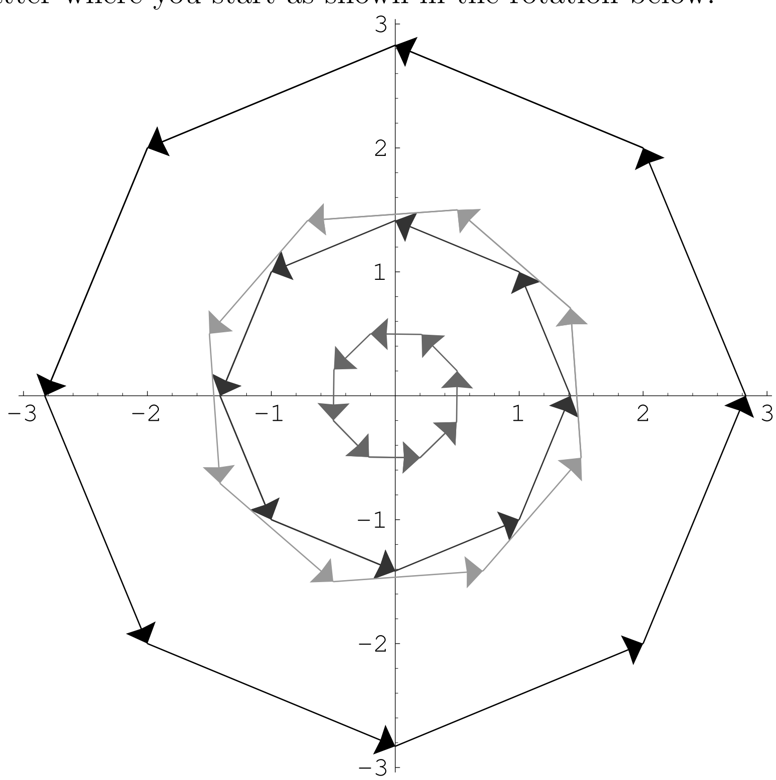
So far our examples have had the characteristic equation having real roots. A quadratic equation can also have a pair of complex roots. We can represent a complex number as $a+bi$ where a and b are real numbers and $i^2=-1$. The conjugate of $a+bi$ is $a-bi$. Both $a+bi$ and $a-bi$ lie on a circle centered at the origin of radius $\sqrt{a^2+b^2}$, which gives a measure of the size of a complex number, called the modulus and written $\|a+bi\|$.

Example: Complex eigenvalues of modulus 1

If the eigenvalues both have modulus (or absolute value) 1 we get rotation. The matrix

$$\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$$

which we considered above gives dynamics with period eight no matter where you start as shown in the rotation below:



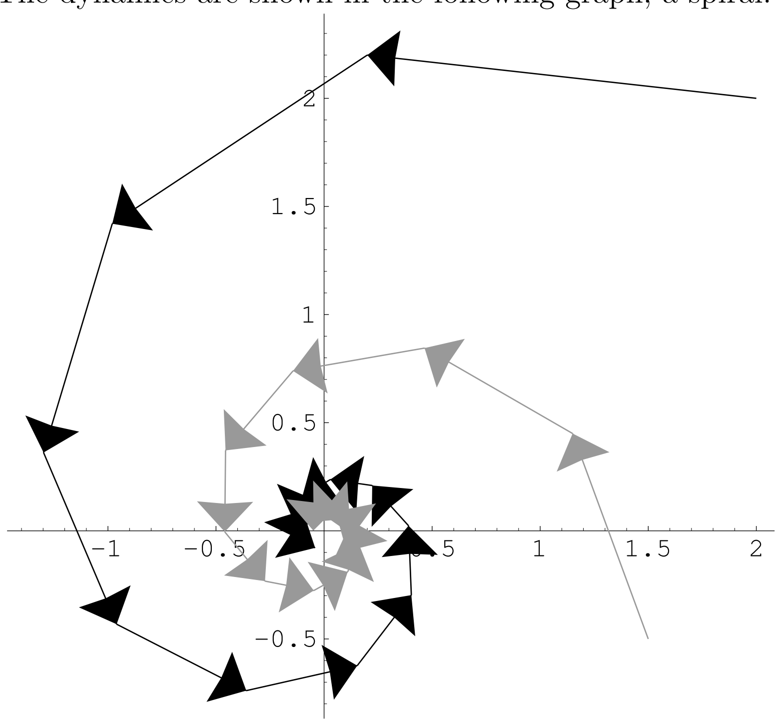
$\diamondsuit$

Example: Complex eigenvalues with modulus less than 1

For an example of complex eigenvalues appearing in a conjugate pair, we can use the matrix

$$\begin{bmatrix} 0.6 & -0.5 \\ 0.5 & 0.6 \end{bmatrix}$$

which has eigenvalues $0.6 \pm 0.5i$. The dynamics give a spiral moving toward the origin. We can tell that the motion is toward the origin and not away from it because $\|0.6 \pm 0.5i\| = \sqrt{0.6^2 + 0.5^2} < 1$. The dynamics are shown in the following graph, a spiral:



$\diamondsuit$

Exercises 3.1:

For numbers 1 through 8 find the characteristic equation, the eigenvalues and an eigenvector for each eigenvalue, and then use your work to describe

the dynamics for the linear transformations with the following matrices with respect to the standard basis:

1. $\begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$

2. $\begin{bmatrix} 2 & 1 \\ 0 & .1 \end{bmatrix}$

3. $\begin{bmatrix} 2 & 3 \\ 4 & 1 \end{bmatrix}$

4. $\begin{bmatrix} .2 & .3 \\ -.3 & .2 \end{bmatrix}$

5. $\begin{bmatrix} .2 & 1 \\ 0 & .2 \end{bmatrix}$

6. $\begin{bmatrix} .2 & 1 \\ 0 & .3 \end{bmatrix}$

7. $\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$

8. $\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$

9–16. Since we know how to multiply matrices, add matrices, and multiply matrices by a scalar, we can evaluate polynomials at a matrix. For example if

$$p(x) = x^2 - 2x + 1$$

and

$$\mathbf{M} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$

then

$$\begin{aligned} p(\mathbf{M}) &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - 2\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + 1\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 7 & 10 \\ 15 & 22 \end{bmatrix} - \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 6 \\ 9 & 15 \end{bmatrix} \end{aligned}$$

For each of the matrices in problems 1-8 find $p(\mathbf{M})$ where $p(x)$ is the characteristic polynomial of $\mathbf{M}$. (What you are seeing at work here is the Cayley-Hamilton Theorem.)

For the next five problems the dynamics of iteration of the linear transformation $L : \mathbb{R}^2 \to \mathbb{R}^2$ are described below. What can you say about the eigenvalues of L?

17. Successive iterations spiral clockwise inward
18. Successive iterations spiral counterclockwise inward
19. Successive iterations spiral outward
20. Iterations grow rapidly approaching the line $y = 3x$ with successive iterates on opposite sides of the line
21. Iterations approach the origin with no apparent oscillation
22. (Project Problem) A *stochastic* matrix has all entries non-negative and has the sum of each column equal to 1.

    (a) Find the eigenvalues of the stochastic matrix

    $$\begin{bmatrix} .5 & .7 \\ .5 & .3 \end{bmatrix}.$$

    (b) Find the eigenvalues of the stochastic matrix

    $$\begin{bmatrix} .9 & .8 \\ .1 & .2 \end{bmatrix}.$$

    (c) Show that the general case

    $$\begin{bmatrix} p & q \\ 1-p & 1-q \end{bmatrix}$$

    has eigenvalues $\lambda = 1$ and $\lambda = p - q$. Since long term behavior is determined by the largest eigenvalue, this tells us that dynamics for a stochastic matrix tend to a fixed point (an eigenvector for the eigenvalue 1).

(d) An epidemic with particularly simple dynamics has 10% of those who are not sick becoming ill each week. Fortunately, 80% of those who were sick recover in the next week. Show that the disease becomes endemic (there is always a certain percentage of the population which is sick) and find the long term incidence level for the disease.

3.2 Canonical forms for 2×2 matrices

Given the importance of eigenvalues it would be nice if we could find forms for matrices in which the eigenvalues can be determined by inspection. Since similar matrices represent the same linear transformation with respect to different ordered bases and eigenvalues are properties of the linear transformation, similar matrices will have the same eigenvalues. Finding a matrix similar to a given matrix but in a nicer form is the same as finding an convenient ordered basis for the linear transformation the matrix represents. In this section we will discuss some of the nice forms that matrices can be put into by judicious choice of basis.

We start with a proposition telling us that eigenvectors for different eigenvalues must point in independent directions:

Proposition 3.2.1 *If T is a linear transformation with eigenvalues $\lambda_1 \neq \lambda_2$, $\vec{a}$ is an eigenvector for the eigenvalue λ_1, and $\vec{b}$ is an eigenvector for λ_2, then the only solution to $x\vec{a} + y\vec{b} = \vec{0}$ is $x = 0$ and $y = 0$.*

PROOF:

Suppose we have $x\vec{a} + y\vec{b} = \vec{0}$. Then we also have

$$\begin{aligned} T(x\vec{a} + y\vec{b}) &= T(\vec{0}) \\ x\,T(\vec{a}) + y\,T(\vec{b}) &= \vec{0} \\ x\lambda_1\vec{a} + y\lambda_2\vec{b} &= \vec{0}. \end{aligned}$$

Multiplying $x\vec{a} + y\vec{b} = \vec{0}$ by λ_2 and subtracting it from $x\lambda_1\vec{a} + y\lambda_2\vec{b} = \vec{0}$ gives $x(\lambda_1 - \lambda_2)\vec{a} = \vec{0}$. This tells us that $x = 0$, since $\lambda_1 \neq \lambda_2$. Multiplying by λ_1 and subtracting will show that $y = 0$ as well. ∎

This tells us that if T has distinct eigenvalues, then there is a basis of eigenvectors. With respect to that basis the matrix of T will be the matrix

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Actually we can get a diagonal matrix whenever we have a basis of eigenvectors, not just when the eigenvalues are distinct.

If the characteristic equation has the form $(\lambda_1 - \lambda)^2 = 0$, we may find that we get the eigenvalue λ_1 twice, but that there is no basis for $\mathbb{R}^2$ consisting of eigenvectors. In this case we look for vectors $\vec{a}$ and $\vec{b}$ such that $T(\vec{a}) = \lambda_1 \vec{a}$ and $T(\vec{b}) = \vec{a} + \lambda_1 \vec{b}$. Using those vectors for our basis we get the form

$$\begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{bmatrix}.$$

Example: Repeated eigenvalue

The linear transformation $L([x, y]) = [5x + y, -x + 3y]$ has matrix

$$\begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix}$$

with respect to the standard bases. This has characteristic equation

$$\lambda^2 - 8\lambda + 16 = 0,$$

so its only eigenvalue is 4. The system

$$\begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 4 \begin{bmatrix} x \\ y \end{bmatrix}$$

becomes

$$\begin{aligned} x + y &= 0 \\ -x - y &= 0 \end{aligned}$$

from which we conclude that $y = -x$. We have only one free choice in finding eigenvectors, not two. Thus there is not a basis of $\mathbb{R}^2$ consisting of eigenvectors.

If we use the eigenvector $\vec{a} = [1, -1]$ as our first basis vector then we can find another using

$$\begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 4 \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

This yields

$$\begin{aligned} x + y &= 1 \\ -x - y &= -1. \end{aligned}$$

This system also has one degree of freedom in its solutions. A second basis vector is $\vec{b} = [.5, .5]$.

To find the matrix for the linear transformation with respect to the new basis we calculate

$$\begin{aligned} L([1,-1]) = [4,-4] &= 4[1,-1] + 0[.5,.5] \\ L([.5,.5] = [3,1] &= 1[1,-1] + 4[.5,.5], \end{aligned}$$

so the matrix with respect to this new basis is

$$\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}.$$

♢

We noted earlier that matrices with real entries could have a pair of conjugate complex numbers for eigenvalues. In such a case the eigenvectors will usually involve complex numbers. But we are working with $\mathbb{R}^2$, not $\mathbb{C}^2$, so we need to see how this situation can be handled using real numbers. The general result is that if the eigenvalues for a linear transformation L are $a \pm bi$ then we can find a basis such that the matrix for L is

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

If the matrix $\mathbf{L}$ has eigenvalues $a \pm bi$, this will give a matrix similar to $\mathbf{L}$ which has a canonical form from which we can read off the eigenvalues. We will illustrate this process with an example:

Example: Complex conjugate eigenvalues

Suppose we want to find a basis such that the matrix for the linear transformation $L[x, y] = [x - 2y, 5x + 3y]$, which has eigenvalues $2 \pm 3i$, is

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

Eigenvectors are found by solving the systems $L([x, y]) = 2 + 3i[x, y]$ and $L([x, y] = 2 - 3i[x, y]$. The system for $2 + 3i$ gives

$$\begin{array}{rcl} x - 2y & = & (2+3i)x \\ 5x + 3y & = & (2+3i)y \\ & \text{so} & \\ (-1-3i)x - 2y & = & 0 \\ 5x + (1-3i)y & = & 0 \\ & \text{giving} & \\ y & = & \dfrac{-1-3i}{2}x \end{array}$$

So $\vec{e_1} = [1, \frac{-1-3i}{2}]$ is an eigenvector for $\lambda_1 = 2 + 3i$ and similarly $\vec{e_2} = [1, \frac{-1+3i}{2}]$ is an eigenvector for $\lambda_2 = 2 - 3i$. We can get real vectors with the desired properties by taking $\vec{b_1} = \vec{e_1} + \vec{e_2} = [2, -1]$ and $\vec{b_2} = i(\vec{e_1} - \vec{e_2}) = [0, -3]$. Now $L(\vec{b_1}) = [4, 7] = 2[2, -1] - 3[0, -3]$ and $L(\vec{b_2}) = [6, -9] = 3[2, -1] + 2[0, -3]$. Thus we get the desired matrix. $\diamondsuit$

To summarize, given a linear transformation $L : \mathbb{R}^2 \to \mathbb{R}^2$ we can find a basis for $\mathbb{R}^2$ which gives a matrix in one of the following canonical forms, where λ_1 and λ_2 have eigenvectors $\vec{v}_1$ and $\vec{v}_2$, respectively:

eigenvalues	basis	canonical form
λ_1, λ_2 real, distinct	$(\vec{v}_1, \vec{v}_2)$	$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$
$\lambda_1 = \lambda_2 = \lambda$ real	$(\vec{v}_1, \vec{v}_2)$ non-collinear	$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$
$\lambda_1 = \lambda_2 = \lambda$ real	no basis of eigenvectors Use $(\vec{v}_1, \vec{v}_3)$ with $L(\vec{v}_3) = \lambda\vec{v}_3 + \vec{v}_1$	$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$
$\lambda = a \pm b\, i$	$\{\vec{v}_1 + \vec{v}_2, i(\vec{v}_1 - \vec{v}_2)\}$	$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$

Exercises 3.2:

For problems 1-12, find the canonical form described in this section and a basis which gives it for the following matrices:

1. $\begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$

2. $\begin{bmatrix} 2 & -3 \\ 2 & 1 \end{bmatrix}$

3. $\begin{bmatrix} -2 & 3 \\ 1 & 1 \end{bmatrix}$

4. $\begin{bmatrix} -1 & -1 \\ 5 & 3 \end{bmatrix}$

5. $\begin{bmatrix} 2 & -3 \\ 0 & 1 \end{bmatrix}$

6. $\begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix}$

7. $\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$

8. $\begin{bmatrix} -1 & 3 \\ -3 & 5 \end{bmatrix}$

9. $\begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix}$

10. $\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$

11. $\begin{bmatrix} 1 & 4 \\ -1 & 1 \end{bmatrix}$

12. $\begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$

13. Show that if $\mathbf{M}$ is in one of the canonical forms of this section then the eigenvalues of $k\mathbf{M}$ are k times the eigenvalues of $\mathbf{M}$. The form of the matrices involved makes it easy to see what the eigenvalues are. (Since any 2×2 matrix is similar to a matrix in canonical form, and since eigenvalues are properties of the linear transformation the matrix represents and not just of the matrix, this shows that multiplying a matrix by a constant multiplies the eigenvalues by the same constant for **all** matrices, not just those in our canonical form.)

14. Find an example to show that the eigenvalues of $\mathbf{M} + \mathbf{N}$ need not be the sum of eigenvalues of $\mathbf{M}$ and $\mathbf{N}$.

15. Find examples to show that the eigenvalues of a product $\mathbf{MN}$ need not be related to those of $\mathbf{M}$ and $\mathbf{N}$. In particular, find $\mathbf{M}$ and $\mathbf{N}$ and eigenvalues λ_i such that

    (a) λ is an eigenvalue for $\mathbf{M}$ but not for $\mathbf{MN}$.

    (b) λ_1 is an eigenvalue for $\mathbf{M}$ and λ_2 is an eigenvalue for $\mathbf{N}$ but $\lambda_1\lambda_2$ is not an eigenvalue for $\mathbf{MN}$.

16. Show that if 0 is an eigenvalue for $\mathbf{N}$ then it is also an eigenvalue for $\mathbf{MN}$.

17. Show that 0 is an eigenvalue for $\mathbf{N}$ if and only if $\mathbf{N}$ does not have an inverse.

18. Using canonical forms show that the product of the eigenvalues of a matrix is its determinant.

19. The sum of the diagonal entries of a matrix is called its **trace**. For canonical forms it is clear that the sum of the eigenvalues is the trace. Show that in general the trace of a matrix is the sum of its eigenvalues. This shows that similar matrices will have the same trace.

Chapter 4

Properties of the dot product

An observant reader may have noticed something peculiar about what we have done so far. We started by requiring that a vector have both magnitude and direction and we have not mentioned either the magnitude or the direction of a vector since. This is because we have concentrated on the properties that vectors have because they can be added and multiplied by scalars. The notion of a vector space captures those aspects of $\mathbb{R}^2$ but not the notions of angle between vectors and length. These need an additional kind of structure, an inner product, exemplified in $\mathbb{R}^2$ by the dot product.

4.1 Dot Products in $\mathbb{R}^2$

Let us return temporarily to the geometric approach to vectors and the way that they are used in physics. In one dimension, if we move an object a distance d against a force F, then the work done is given by the number Fd. But both force and displacement are actually vector quantities, so in two dimensions (or more) we need to take into account the direction. We reduce to the one dimensional case by taking the component of the force in the direction of the displacement, rather than the whole force in our calculation.

Given two vectors $\vec{a}$ and $\vec{b}$, we often write $\vec{a}$ as a sum of a vector in the same direction as $\vec{b}$ and one perpendicular to $\vec{b}$,

$$\vec{a} = k\vec{b} + \vec{c} \text{ with } \vec{c} \perp \vec{b}.$$

The piece parallel to $\vec{b}$ is called the **component** of $\vec{a}$ in the direction of $\vec{b}$. It gives the vector which is a multiple of $\vec{b}$ which is closest to $\vec{a}$. The situation is illustrated in Figure 4.1.

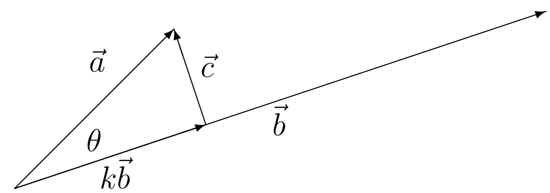


Figure 4.1: Geometry of the dot product

We can find the length of $k\vec{b}$ by applying some simple trigonometry to the right triangle formed by $\vec{a}$ as hypotenuse with sides $k\vec{b}$ and $\vec{c}$. Since $\cos(\theta)$ is the ratio of the length of the adjacent side to the length of the hypotenuse,

$$\cos\theta = \frac{k\|\vec{b}\|}{\|\vec{a}\|}.$$

Solving for k gives

$$k = \frac{\|\vec{a}\| \ \|\vec{b}\| \cos(\theta)}{\|\vec{b}\|^2}.$$

This concept forms the basis for the geometric definition of the dot product between two vectors:

Definition 4.1.1 *The dot product of two vectors is defined geometrically as*

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \ \|\vec{b}\| \cos(\theta).$$

We may notice several immediate consequences of this definition.

1. The dot product gives a real number (a scalar), not a vector.

2. The right side is symmetric in $\vec{a}$ and $\vec{b}$ so $\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}$.

3. If $\vec{a}$ and $\vec{b}$ are perpendicular then $\vec{a} \cdot \vec{b} = 0$.

4. The dot product of a vector with itself is the square of its magnitude.

We also notice that the definition is hard to use if we do not know θ, so that an algebraic approach to the definition, if possible, would be preferable. Fortunately, such a definition is readily available, as the next proposition shows.

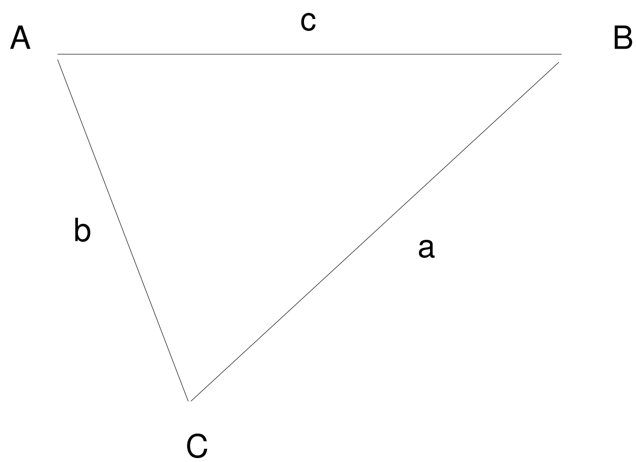


Figure 4.2: Law of Cosines

Proposition 4.1.1 *Let $\vec{a} = [a_1, a_2]$ and $\vec{b} = [b_1, b_2]$ and let θ be the angle between them ($0 \leq \theta \leq \pi$). Then $\vec{a} \cdot \vec{b} = \|\vec{a}\|\|\vec{b}\| \cos(\theta) = a_1b_1 + a_2b_2$.*

PROOF:

We will use the law of cosines from trigonometry, which in a common formulation states that in a triangle with angles A, B, and C, opposite the sides of length a, b, and c, respectively,

$$c^2 = a^2 + b^2 - 2ab\cos(C).$$

(See Figure 4.2)

We will express this same result in vector terminology:

$$\|\vec{b} - \vec{a}\|^2 = \|\vec{a}\|^2 + \|\vec{b}\|^2 - 2\|\vec{a}\|\|\vec{b}\| \cos(\theta).$$

When we solve this equation for the last term on the right and then express the norms of the vectors in terms of their components we find

$$\begin{aligned} 2\|\vec{a}\|\|\vec{b}\| \cos\theta &= \|\vec{a}\|^2 + \|\vec{b}\|^2 - \|\vec{b} - \vec{a}\|^2 \\ &= a_1^2 + a_2^2 + b_1^2 + b_2^2 - ((b_1 - a_1)^2 + (b_2 - a_2)^2) \end{aligned}$$

$$
\begin{aligned}
&= a_1^2 + a_2^2 + b_1^2 + b_2^2 - (b_1^2 - 2a_1b_1 + a_1^2 + b_2^2 - 2a_2b_2 + a_2^2) \\
&= a_1^2 - a_1^2 + a_2^2 - a_2^2 + b_1^2 - b_1^2 + b_2^2 - b_2^2 + 2a_1b_1 + 2a_2b_2 \\
&= 2(a_1b_1 + a_2b_2).
\end{aligned}
$$

Thus

$$\|\vec{a}\|\|\vec{b}\| \cos\theta = a_1b_1 + a_2b_2.$$

∎

Because the algebraic form given by this proposition is easier to work with than our geometric definition we will use the following definition for the dot product of algebraic vectors for our further generalization:

Definition 4.1.2 *For vectors $\vec{a} = [a_1, a_2]$ and $\vec{b} = [b_1, b_2]$, the dot product is given by*

$$\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2.$$

Example: Finding the angle between vectors

We use the dot product to determine the angle between $\vec{a} = [1, 2]$ and $\vec{b} = [3, -4]$. To find θ we calculate,

$$
\begin{aligned}
\cos(\theta) &= \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|\|\vec{b}\|} \\
&= \frac{(1 \times 3) + (2 \times -4)}{\|\vec{a}\|\|\vec{b}\|} \\
&= \frac{-5}{5\sqrt{5}} \\
&\approx -.4472135955,
\end{aligned}
$$

and thus

$$
\begin{aligned}
\theta &\approx \arccos(-.4472135955) \\
&\approx 2.03443936 \text{ radians.}
\end{aligned}
$$

◇

Example: Direction Cosines

The direction of a vector is often given by specifying the angle formed by the vector and the standard basis vectors. Since the dot product gives an easy way to recover the cosine of that angle, what are usually given are the **direction cosines**. For example, the vector $[3,-4]$ forms an angle θ_x with the x-axis and θ_y with the y-axis, where

$$\begin{aligned}\cos(\theta_x) &= \frac{[3,-4]\cdot[1,0]}{\|[3,-4]\|\ \|[1,0]\|}\\ &= \frac{3}{5}\\ \cos(\theta_y) &= \frac{[3,-4]\cdot[0,1]}{\|[3,-4]\|\ \|[0,1]\|}\\ &= \frac{-4}{5}.\end{aligned}$$

$\diamondsuit$

The dot product is easily calculated from Definition 4.1.2; at the same time Proposition 4.1.1 justifies our interpreting the dot product in the geometric sense originally used. Moreover, the essential properties of the dot product are easily proven from its algebraic description.

Proposition 4.1.2 *Let $\vec{a},\vec{b},\vec{c}$ be arbitrary vectors and let $k\in\mathbb{R}$, then*

1. *Dot product is commutative:* $\vec{a}\cdot\vec{b}=\vec{b}\cdot\vec{a}$
2. *Dot products preserve multiplication by a scalar:*$k(\vec{a}\cdot\vec{b})=(k\vec{a})\cdot\vec{b}=\vec{a}\cdot(k\vec{b})$
3. *Dot products distribute over sums:*$\vec{a}\cdot(\vec{b}+\vec{c})=\vec{a}\cdot\vec{b}+\vec{a}\cdot\vec{c}$
4. *Lengths can be recovered from dot products:* $\vec{a}\cdot\vec{a}=\|\vec{a}\|^2$
5. $\vec{a}\cdot\vec{a}\geq 0$ *and* $\vec{a}\cdot\vec{a}=0$ *if and only if* $\vec{a}=\vec{0}$.

PROOF:

Commutativity follows immediately from commutativity of multiplication in $\mathbb{R}$.

For the preservation of scalar multiples we calculate

$$\begin{aligned}
k(\vec{a}\cdot\vec{b}) &= k(a_1b_1 + a_2b_2) \\
&= ka_1b_1 + ka_2b_2 \\
(k\vec{a})\cdot\vec{b} &= [ka_2, ka_2]\cdot[b_1, b_2] \\
&= ka_1b_1 + ka_2b_2 \\
\vec{a}\cdot(k\vec{b}) &= [a_1, a_2]\cdot[kb_1, kb_2] \\
&= ka_1b_1 + ka_2b_2.
\end{aligned}$$

Since these all give the same result, they must have been equal.

For the distributivity over sums, we have

$$\begin{aligned}
\vec{a}\cdot(\vec{b}+\vec{c}) &= [a_1, a_2]\cdot([b_1, b_2] + [c_1, c_2]) \\
&= [a_1, a_2]\cdot[b_1 + c_1, b_2 + c_2] \\
&= a_1(b_1 + c_1) + a_2(b_2 + c_2) \\
&= a_1b_1 + a_1c_1 + a_2b_2 + a_2c_2 \\
&= a_1b_1 + a_2b_2 + a_1c_1 + a_2c_2 \\
&= \vec{a}\cdot\vec{b} + \vec{a}\cdot\vec{c}.
\end{aligned}$$

We leave the proofs of 4 and 5 as exercises. ∎

The first three parts of this proposition say that the dot product behaves very much like multiplication of numbers. However, in one way the dot product departs markedly from multiplication of numbers. If the product of two *numbers* is 0, then one of the numbers must have been 0. For dot product, however, this is not true. If $\vec{a}$ and $\vec{b}$ are perpendicular then since $\theta = \frac{\pi}{2}$, $\cos(\theta) = 0$, and $\vec{a}\cdot\vec{b} = 0$.

The converse is almost true by Proposition 4.1.1. If $\vec{a}\cdot\vec{b} = 0$, then it must follow that $\|\vec{a}\| = 0$ or $\|\vec{b}\| = 0$, or $\cos(\theta) = 0$. Thus if $\vec{a}\cdot\vec{b} = 0$, either the vectors are perpendicular or at least one of them is the zero vector. If we agree that the zero vector is perpendicular to every vector, then we can conclude that $\vec{a}\cdot\vec{b} = 0$ if and only if $\vec{a}$ and $\vec{b}$ are perpendicular (a numerical description of a geometric property).

We conclude this section with an important result which follows from the geometric description of the dot product.

Proposition 4.1.3 *(Cauchy-Schwarz Inequality). For any two vectors $\vec{a}$ and $\vec{b}$ in $\mathbb{R}^2$*

$$(\vec{a} \cdot \vec{b})^2 \leq \|a\|^2 \|b\|^2$$

Equality holds if and only if $\vec{b} = k\vec{a}$ for some $k \in \mathbb{R}$ (which says that $\vec{a}$ and $\vec{b}$ are parallel).

Proof:

A good starting point is the geometric description of the dot product:

$$(\vec{a} \cdot \vec{b}) = \|\vec{a}\| \|\vec{b}\| \cos(\theta).$$

If we square both sides we get

$$(\vec{a} \cdot \vec{b})^2 = \|\vec{a}\|^2 \|\vec{b}\|^2 \cos^2(\theta),$$

or, since $\cos^2(\theta) \leq 1$,

$$(\vec{a} \cdot \vec{b})^2 \leq \|\vec{a}\|^2 \|\vec{b}\|^2.$$

Now suppose $(\vec{a} \cdot \vec{b})^2 = \|\vec{a}\|^2 \|\vec{b}\|^2$. This is true, trivially, if either $\vec{a} = \vec{0}$ or $\vec{b} = \vec{0}$. If neither vector is zero, then equality holds if and only if $\cos\theta = \pm 1$. Now $\cos\theta = 1$ if and only if $\theta = 0$ and $\cos\theta = -1$ if and only if $\theta = \pi$ (recall that we chose $0 \leq \theta \leq \pi$). In either case there is a $k \in \mathbb{R}$ with $\vec{b} = k\vec{a}$, where $k \geq 0$ if and only if $\theta = 0$ and $k < 0$ if and only if $\theta = \pi$. ∎

Exercises 4.1:
For 1-6, find $\cos\theta$, where θ is the angle between $\vec{a}$ and $\vec{b}$:

1. $\vec{a} = [1, 1]$ and $\vec{b} = [2, 1]$
2. $\vec{a} = [-1, 1]$ and $\vec{b} = [6, 2]$
3. $\vec{a} = [0, 3]$ and $\vec{b} = [-2, 1]$
4. $\vec{a} = [5, 1]$ and $\vec{b} = [2, 1]$
5. $\vec{a} = [1, 3]$ and $\vec{b} = [2, 6]$

6. $\vec{a} = [1, -2]$ and $\vec{b} = [2, 1]$
7. Find the direction cosines for $[3, 4]$.
8. Find the direction cosines for $[12, -5]$.
9. Find the direction cosines for $[1, 1]$.
10. Find the direction cosines for $[-3, 4]$.
11. Find the cosines of the angles of the triangle with vertices $A = (-3, -2)$, $B = (5, 1)$ and $C = (2, 4)$.
12. Show that the diagonals of a rhombus are perpendicular by showing that if $\|\vec{a}\| = \|\vec{b}\|$, then $\vec{a} + \vec{b} \perp \vec{a} - \vec{b}$.
13. Find c so that $[3, -4]$ and $[c, 1]$ are perpendicular.
14. Find b so that $[b, 2] \perp [3, 4]$.
15. Find k so that $[1, 3] - k[2, 4]$ is perpendicular to $[2, 4]$.
16. Find k so that $[1, -2] - k[3, 1]$ is perpendicular to $[3, 1]$.
17. Prove part 4 of Proposition 4.1.2.
18. Prove part 5 of Proposition 4.1.2.
19. Use the Cauchy-Schwarz inequality to prove that
$$\|\vec{a} + \vec{b}\|^2 \leq (\|\vec{a}\| + \|\vec{b}\|)^2.$$
Hint:$\|\vec{v}\|^2 = \vec{v} \cdot \vec{v}$

4.2 Orthogonal Projection

The dot product was introduced to allow us to find the component of a vector $\vec{a}$ in the direction of another vector $\vec{b} \neq \vec{0}$. In this section we will see how the relationship with the dot product can arise from consideration of this problem using trigonometry, calculus, or geometry and linear algebra. Our problem is

Find k so that the length of $\vec{a} - k\vec{b}$ is minimized.

Solution 1—using calculus: The length of $\vec{a} - k\vec{b}$ is a function of k. Because it involves a radical it is easier to minimize the square of the length, so we'll do that.

$$\begin{aligned} f(k) &= \|\vec{a} - k\vec{b}\|^2 \\ &= (a_1 - kb_1)^2 + (a_2 - kb_2)^2 \\ &= a_1^2 - 2k\ a_1b_1 + k^2\ b_1^2 + a_2^2 - 2k\ a_2b_2 + k^2\ b_2^2 \\ &= a_1^2 + a_2^2 - 2(a_1b_1 + a_2b_2)\ k + (b_1^2 + b_2^2)\ k^2 \\ &= \|\vec{a}\|^2 - 2\vec{a}\cdot\vec{b}\ k + \|\vec{b}\|^2\ k^2 \end{aligned}$$

To minimize this we find k such that $f'(k) = 0$. Now

$$f'(k) = -2\vec{a}\cdot\vec{b} + 2\|\vec{b}\|^2 k$$

so we conclude that

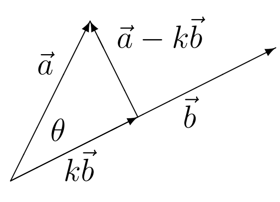
$$k = \frac{\vec{a}\cdot\vec{b}}{\|\vec{b}\|^2} = \frac{\vec{a}\cdot\vec{b}}{\vec{b}\cdot\vec{b}}$$

is the only critical point. Now

$$f''(k) = 2\|\vec{b}\|^2 > 0$$

so this gives a minimum as desired.

Solution 2—using trigonometry: Here what we can find is the length of the projection of $\vec{a}$ onto $\vec{b}$. A picture will help:



Using the definition of $\cos(\theta)$, it becomes clear from the picture that we want $\|k\ \vec{b}\|$ to be $\|\vec{a}\|\cos(\theta)$. Now from the geometric description of the dot product

$$\cos(\theta) = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\ \|\vec{b}\|}.$$

Now $\|k\ \vec{b}\| = |k|\ \|\vec{b}\|$ so this means we want

$$\begin{aligned} k &= \frac{\|\vec{a}\| \cos(\theta)}{\|\vec{b}\|} \\ &= \|\vec{a}\| \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|\ \|\vec{b}\|^2} \\ &= \frac{\vec{a} \cdot \vec{b}}{\vec{b} \cdot \vec{b}} \end{aligned}$$

Solution 3—using geometry and linear algebra: We know from geometry that the shortest distance from a point (the endpoint of $\vec{a}$) to a line (the one determined by the origin and $\vec{b}$) is found by dropping a perpendicular to the line from the point and measuring its length. So we want to write

$$\vec{a} = k\vec{b} + h\vec{b}^{\perp}$$

where $\vec{b}^{\perp}$ is some vector perpendicular to $\vec{b}$. Thus $\vec{b} \cdot \vec{b}^{\perp} = 0$. We can use properties of the dot product to find k:

$$\begin{aligned} \vec{b} \cdot \vec{a} &= \vec{b} \cdot (k\vec{b} + h\vec{b}^{\perp}) \\ &= k\vec{b} \cdot \vec{b} + h\vec{b} \cdot \vec{b}^{\perp} \\ &= k\vec{b} \cdot \vec{b} + 0 \end{aligned}$$

Thus

$$k = \frac{\vec{a} \cdot \vec{b}}{\vec{b} \cdot \vec{b}}$$

as in the other derivations.

These calculations lead us to the following definition:

Definition 4.2.1 *The projection of $\vec{a}$ onto $\vec{b}$ is the vector*

$$\overrightarrow{\mathrm{proj}}_{\vec{b}}\ (\vec{a}) = \frac{\vec{a} \cdot \vec{b}}{\vec{b} \cdot \vec{b}}\ \vec{b}.$$

Because the ideas involved in projections are useful in settings where the word "perpendicular" might have other connotations the term "orthogonal" is used in inner product spaces.

Definition 4.2.2 *Vectors $\vec{a}$ and $\vec{b}$ are said to be* **orthogonal** *if $\vec{a} \cdot \vec{b} = 0$.*

Notice that with this definition the zero vector $\vec{0}$ is orthogonal to any vector.

Example: Finding a projection

Find the vector in the direction of $[1, 2]$ closest to $[-3, 5]$.

Using the ideas in this section it is clear we want the projection of $[-3, 5]$ onto $[1, 2]$. This is

$$\overrightarrow{\text{proj}}_{[1,2]}([-3,5]) = \frac{[-3,5] \cdot [1,2]}{[1,2] \cdot [1,2]}[1,2] = \frac{7}{5}[1,2] = [\frac{7}{5}, \frac{14}{5}].$$

$\diamondsuit$

Exercises 4.2:

For problems 1-6, find the vector $\overrightarrow{\text{proj}}_{\vec{u}}\,(\vec{v})$ in the direction of $\vec{u}$ closest to $\vec{v}$.

1. $\vec{u} = [1, 2]$ and $\vec{v} = [3, 4]$
2. $\vec{u} = [-1, 4]$ and $\vec{v} = [3, -2]$
3. $\vec{u} = [5, -2]$ and $\vec{v} = [1, 1]$
4. $\vec{u} = [1, 4]$ and $\vec{v} = [-1, 2]$
5. $\vec{u} = [-2, -1]$ and $\vec{v} = [2, 3]$
6. $\vec{u} = [-3, 2]$ and $\vec{v} = [1, 1]$

7–12 For each of the pairs of vectors in exercises 1–6 write $\vec{v}$ as the sum of a vector parallel to $\vec{u}$ and a vector perpendicular to $\vec{u}$.

13. Let $\vec{v}$ and $\vec{w}$ be two fixed, non-zero, nonparallel vectors in $\mathbb{R}^2$. Let $\vec{a}$ be an arbitrary vector.
    (a) Find scalars h and k so that $\vec{a} = h\vec{v} + k\vec{w}$.
    (b) Show that if $\vec{v}$ and $\vec{w}$ are perpendicular there is a nice form for h and k in terms of dot products.

4.3 Summary and Preview

In this part of the book we have discussed the properties of vectors in the plane, concentrating first on those properties which do not involve angles and then introducing dot products to recover angles and length. We defined linear transformations and showed how to get the matrix for a linear transformation with respect to a choice of basis. We saw how particular choices of basis might give nicer forms for the matrix and further information about the linear transformation. We saw how the dynamics of iteration is related to the eigenvalues of the linear transformation. Because we were working in $\mathbb{R}^2$ none of the technique was particularly difficult or involved. Most calculations needed little more than solving a system of two equations in two unknowns or solving a quadratic equation.

In the chapters which follow we will generalize these ideas to vector spaces over the reals or over other fields. (For much of the theory it is more useful to work over the complex numbers because polynomials always have enough roots in the complex numbers; for applications we sometimes want to work over finite fields; we can avoid some problems more properly considered in a numerical analysis course if we work in the rational numbers when possible.) We will start by looking at lots of examples. Because these examples are a bit more involved than $\mathbb{R}^2$ we will need more systematic methods for solving the systems of equations which result. These systematic methods give rise to an algorithm for reducing a matrix to what is called echelon form which we will use for just about everything.

A first course on linear algebra focuses on vector spaces and linear transformations, systems of linear equations, matrices, and some of the things you can do with them. The methods have wide applicability and the ideas lead to varied generalizations making them a key to modern mathematics. No subject in the undergraduate curriculum is more central to the further development of mathematics, both pure and applied, algebraic, analytic, geometric, or combinatorial.

Chapter 5

Vector Spaces and Linear Transformations

5.1 Vector Spaces: Definitions and Examples

5.1.1 Definitions of fields and vector spaces

The example of vectors in the plane which we looked at in Chapter 1 is only one example of a large collection of similarly behaved systems. In elementary functions courses we study addition and scalar multiplication of functions from the reals to the reals and find the same properties that we enumerated in Proposition 1.4. In calculus we used these operations to break the problem of finding derivatives and indefinite integrals into manageable pieces, noting that differentiable functions and continuous functions were closed under addition and multiplication by constants. Whenever mathematicians find several examples of similar behavior in different settings, they look for a more abstract formulation which has the examples as special cases. If the abstraction gives a notion which is easy to work with and which gives useful information about the examples it becomes a mathematical object in its own right. Linear algebra is the study of such an abstraction.

Let us first concentrate on the properties of the scalars. In Chapter 1 we summarized the algebraic properties of the real numbers that we would be using. The relevant properties are those that make the real numbers a *field*. They include the fact that there are two operations on the reals, addition and multiplication, which satisfy eleven axioms (listed below) which will be familiar because of their heavy use in high school algebra. In most of this

book we will be working with vector spaces over the reals, though in some cases it makes more sense to work over other fields. (We often want to solve systems of equations over the rationals; eigenvalue problems are most easily solved over the complex numbers, and coding theory uses vector spaces over finite fields. We will consider these examples in section 2.3.)

Definition 5.1.1 *A* **field** *is a set F equipped with two binary operations $+ : F \times F \to F$ and $\times : F \times F \to F$ satisfying the following axioms for all a, b, and $c \in F$:*

Closure:	$a + b \in F$	$a \times b \in F$
Associativity:	$(a + b) + c = a + (b + c)$	$(a \times b) \times c = a \times (b \times c)$
Commutativity:	$a + b = b + a$	$a \times b = b \times a$
Identity:	$\exists_{0\in F}\forall_a(a + 0 = a)$	$\exists_{1\in F}\forall_a(a \times 1 = a)$
Inverses:	$\forall_{a\in F}\exists_{-a}(-a + a = 0)$	$\forall_{a\neq 0}\exists_{\frac{1}{a}\in F}(a \times \frac{1}{a} = 1)$
Distributive:	$a \times (b + c) = (a \times b) + (a \times c)$	

In this definition we have used some useful shorthand notation from symbolic logic: $\forall$ reads as "for all" and $\exists$ reads as "there exists". Thus the expression for inverses, $\forall_{a\in F}\exists_{-a}(-a + a = 0)$, reads "for every $a \in F$ there is an element $-a$ such that $-a + a = 0$". Similarly, $\forall_{a\neq 0}\exists_{\frac{1}{a}\in F}(a \times \frac{1}{a} = 1)$ translates as "for every $a \in F$ which is not 0, there is an element $\frac{1}{a}$ in F such that $a \times \frac{1}{a} = 1$."

The examples of fields we will use in this course are the set of real numbers with the usual addition and multiplication, the set of complex numbers with the usual addition and multiplication, the rationals with the usual operations, and the integers with addition and multiplication mod 2. The field axioms summarize the properties of numbers used in high school algebra; when working in a field nothing untoward happens.

There are some immediate consequences of these axioms which we will want to use later which we have not made part of the definition. They deal with the uniqueness of the identities and inverses. We will prove these results in rather general form so that they will apply later in other situations:

Proposition 5.1.1 *If a commutative operation has an identity then that identity is unique.*

PROOF:

Suppose that the operation $*$ has two identity elements, e_1 and e_2. Then consider $e_1 * e_2$. By the commutative law

$$e_1 * e_2 = e_2 * e_1.$$

Since e_2 is an identity

$$e_1 * e_2 = e_1$$

and since e_1 is an identity

$$e_2 * e_1 = e_2.$$

thus $e_1 = e_2$. ∎

If e is a (two sided) identity element for $*$ then whenever $l * r = e$ we say that l is the left inverse of r and r is the right inverse of l. The designation of left and right only matters if the operation is not commutative (and we will see some later in the course which are not) so in the case of field operations the following proposition tells us that inverses are unique.

Proposition 5.1.2 *If $*$ is an associative operation with identity e and a has both a right inverse r and a left inverse l, then $r = l$.*

PROOF:

We are given that $a * r = e$ and $l * a = e$. Let us evaluate $l * (a * r)$ in two different ways:

$$l = l * e = l * (a * r) = (l * a) * r = e * r = r.$$

This shows that $l = r$. ∎

In the $\mathbb{R}^2$ there is a second set (the vectors) which has two operations. We can add vectors and we can multiply a scalar and a vector. For the time being we will ignore the additional structure that the dot product gives, so we will not be talking about length, angles, or projections at this point. In order to talk about abstract vector spaces we need to say what properties we want the operations to have. The properties of $\mathbb{R}^2$ summarized in Definition 1.1.5 give us the axioms we want.

Definition 5.1.2 *A* **vector space** *over a field F (whose elements are called scalars) is a set $\mathcal{V}$ (whose elements are called vectors) which has two operations: $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and scalar multiplication $F \times \mathcal{V} \to \mathcal{V}$ (usually indicated by juxtaposition) which are required to satisfy the following axioms for all vectors $\vec{a}, \vec{b}, \vec{c}$, and scalars h and k:*

Closure:	*both $\vec{a} + \vec{b}$ and $k\vec{a}$ are vectors*
Commutativity of $+$:	$\vec{a} + \vec{b} = \vec{b} + \vec{a}$
Associativity of $+$:	$(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$
Identity for $+$:	*There is a unique vector $\vec{0}$ with $\vec{0} + \vec{a} = \vec{a}$ for all $\vec{a}$.*
Inverses for $+$:	*For each $\vec{a}$ there is a unique $-\vec{a}$ so that $\vec{a} + -\vec{a} = \vec{0}$*
Absorption:	$h(k\vec{a}) = (hk)\vec{a}$
Distributivity:	$(h + k)\vec{a} = (h\vec{a}) + (k\vec{a})$
	$h(\vec{a} + \vec{b}) = (h\vec{a}) + (h\vec{b})$
Identity for scalars:	$1\vec{a} = \vec{a}$

As soon as you have a new definition two things cry out to be done at once: find several examples, and determine some of the consequences. We will start by proving some of the properties which could have been included as axioms but weren't so as to keep the list manageable.

Proposition 5.1.3 *In any vector space $0\vec{a} = \vec{0}$.*

PROOF:

> We know that $0+1 = 1$, so $(0+1)\vec{a} = 1\vec{a}$. By the distributive law $(0 + 1)\vec{a} = (0\vec{a}) + (1\vec{a})$. Identity for scalars gives $1\vec{a} = \vec{a}$ in both places so $(0\vec{a}) + \vec{a} = \vec{a}$. Now add $-\vec{a}$ to both sides and use the associative law and the properties of inverse and identities to get $(0\vec{a}) = \vec{0}$. ∎

Proposition 5.1.4 *In any vector space $-1\vec{a} = -\vec{a}$.*

PROOF:

> Recall that the inverse axiom says that $-\vec{a}$ is the unique vector with $\vec{a} + -\vec{a} = \vec{0}$. All we have to show is that $\vec{a} + (-1\vec{a}) = \vec{0}$. Now $\vec{a} = 1\vec{a}$ by the identity axiom for scalars, so we can reduce the problem to showing that $(1\vec{a}) + (-1\vec{a}) = \vec{0}$. But by distributivity $(1\vec{a}) + (-1\vec{a}) = (1 + (-1))\vec{a}$. This in turn is equal to $0\vec{a}$, which by Proposition 5.1.3 is $\vec{0}$. ∎

Next we will explore a wide variety of examples of vector spaces over the reals. Our object is to show the scope of the definition and give a hint as to its utility.

5.1.2 $\mathbb{R}^n$

Example: $\mathbb{R}^3$

The set $\mathbb{R}^3$ is the set of ordered triples of real numbers. The sum is defined by $[a_1, a_2, a_3] + [b_1, b_2, b_3] = [a_1 + b_1, a_2 + b_2, a_3 + b_3]$ and the scalar product is given by $k[a_1, a_2, a_3] = [ka_1, ka_2, ka_3]$. Commutativity of $+$ follows from three applications of commutativity of addition for real numbers. Similarly associativity of $+$ follows from three applications of associativity of addition for the reals. The identity is the vector $[0, 0, 0]$. Its properties follow from the fact that 0 is the identity for addition of real numbers. The inverse of $[a_1, a_2, a_3]$ is $[-a_1, -a_2, -a_3]$. Its properties follow from those of the inverse for addition in the reals. Absorption becomes three applications of the associative law for multiplication. Write it out so you can see it. The two distributive laws both follow from three applications of the distributive law for real numbers. Identity for scalars follows from three applications of the identity for multiplication of reals. All that seems to be happening here is that each axiom of a vector space follows from three applications of the similar axiom for the reals. $\diamondsuit$

Example: $\mathbb{R}^n$

It is clear that the argument used to show that $\mathbb{R}^3$ (and for that matter $\mathbb{R}^2$) did not depend in any essential way on the fact that we took triples of real numbers. If we use n-tuples of real numbers $[a_1, ..., a_n]$ and define addition and scalar multiplication componentwise, then proving that the axioms of a vector space hold will be a tedious matter of applying the similar axioms for the reals n times. Having noticed the pattern, we claim to have proved the result by describing how we would go about proving it for any particular n.

It turns out that this is a very typical example of a vector space over the reals. Some of our later examples are merely $\mathbb{R}^n$ in disguise. $\diamondsuit$

5.1.3 Polynomial spaces

Example: $\mathbb{R}[x]_3$ Polynomials of degree 3 or less

For example consider polynomials of degree three or less in one variable x with coefficients in the reals. A typical such polynomial has the form ax^3+bx^2+cx+d, where any (or all) of the coefficients a, b, c, or d may be 0. It is clear that we may identify such a polynomial with the quadruple $[a, b, c, d]$ of its coefficients in order of decreasing powers of x. Addition of polynomials corresponds exactly to addition of vectors in $\mathbb{R}^4$ and multiplication of a vector by a real number corresponds to multiplication of a polynomial by a real number. Thus this example is just $\mathbb{R}^4$ in disguise. We know that $\mathbb{R}^4$ is a vector space over the reals so the set of polynomials of degree three or less is too. $\diamondsuit$

Example: $\mathbb{R}[x]$ All polynomials with real coefficients

Why stop at degree three? We know how to add two polynomials of different degrees– just put in 0 as the coefficient of higher powers of x in the polynomial of lower degree and treat it as a polynomial of higher degree. Note that this way of thinking about polynomial addition makes it clear that for any particular polynomials $a(x)$, $b(x)$, and $c(x)$ we can verify the axioms of a vector space by working in the set of polynomials of degree less than or equal to the largest of the three degrees. The zero polynomial is the identity. While we have used what we know about $\mathbb{R}^n$ to see that the set of all polynomials of one variable with real coefficients forms a vector space, it is clear that in some sense this is a "bigger" space than any of the $\mathbb{R}^n$'s. It is an example of an infinite dimensional vector space. The mind boggles only if it tries to view this situation geometrically. $\diamondsuit$

Example: $\mathbb{R}[x, y]$ Polynomials in two variables

Why stop with one variable? We know from algebra how to add polynomials in two (or more) variables and how to multiply them by constants. The behavior is very like the polynomials in one variable. In particular, they form a vector space over the reals. $\diamondsuit$

5.1.4 Sequences and function spaces

Example: $\mathbb{R}^{\mathbb{N}}$ Sequences of real numbers

An example of a vector space encountered in calculus is the set of all sequences of real numbers. Recall that a sequence may be thought of as a function from the natural numbers to the reals: we usually write them as (a_n), where a_n is describes the n^{th} term of the sequence. We are not concerned with convergence here, only with addition of sequences (done term by term) and multiplication of sequences by constants. It is clear that these operations give us sequences back again. The zero sequence is the identity. Use of the properties of the real numbers on each of the terms gives us the rest of the axioms. $\diamondsuit$

A more general class of examples arises from the following theorem.

Theorem 5.1.5 *If S is any set, then the set of all functions from S to $\mathbb{R}$ is a vector space over $\mathbb{R}$. The vector space operations are the sum $(f + g)(x) = f(x) + g(x)$, and the scalar product $(kf)(x) = kf(x)$. We will write this vector space as $\mathbb{R}^S$.*

PROOF:

We need to verify the axioms in Definition 5.1.2. The closure axioms follow directly from the definition of the operations on functions and the closure axioms for the real numbers. We know that commutativity for + is true because two functions are equal if they always have the same value and

$$\begin{aligned}
(f + g)(x) &= f(x) + g(x) \\
&= g(x) + f(x) \text{ by commutativity of } + \text{ in } \mathbb{R} \\
&= (g + f)(x)
\end{aligned}$$

Similarly associativity of $+$ for functions from S to $\mathbb{R}$ follows from associativity of $+$ for $\mathbb{R}$:

$$\begin{aligned}(f+(g+h))(x) &= f(x)+(g+h)(x)\\ &= f(x)+(g(x)+h(x))\\ &= (f(x)+g(x))+h(x)\\ &= (f+g)(x)+h(x)\\ &= ((f+g)+h)(x).\end{aligned}$$

The identity is the constant function with value 0. The inverse of a function f is the function $-f$ with value at x given by $-f(x)$. Since $f(x) + -f(x) = 0$ for all x, this gives the inverse. The absorption law follows from associativity for multiplication in $\mathbb{R}$:

$$h(kf)(x) = h(kf(x)) = (hk)f(x) = ((hk)f)(x).$$

The two distributive laws follow from the distributive law for $\mathbb{R}$:

$$(h+k)f(x) = hf(x) + kf(x) = (hf + kf)(x)$$

and

$$h(f+g)(x) = h(f(x)+g(x)) = (hf(x)+hg(x)) = (hf+hg)(x).$$

Identity for scalars follows from the fact that 1 is the identity for multiplication of real numbers. ∎

This gives us new ways to look at some of our earlier examples: $\mathbb{R}^3$ can be thought of as the space of all functions from the set $\{1, 2, 3\}$ to $\mathbb{R}$; sequences can be thought of as functions from the natural numbers $\mathbb{N}$ to $\mathbb{R}$. We can also get new examples by taking the set of functions from $\mathbb{R}$ to $\mathbb{R}$ (that is *all* functions, we will see in the next section how to restrict this to continuous functions, integrable functions, and differentiable functions).

5.1.5 Matrices form a vector space

Example: Matrices

Another very important example of a vector space which can be thought of as a function space is the space of $m \times n$ matrices. A matrix is a rectangular array of numbers. If we have an $m \times n$ matrix then there are m rows each with n entries which are arranged in neat columns. For example

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \end{bmatrix}$$

is a 3×4 matrix. When we refer to a general matrix we usually write something like $\mathbf{M} = [[m_{ij}]]$, indicating that the entry in the ith row and jth column is m_{ij}. This saves much space in writing things down and gives us a clear idea how to see that the set of $m \times n$ matrices is a vector space. This would be pictured as

$$\begin{bmatrix} m_{11} & \ldots & m_{1k} & \ldots & m_{1n} \\ \vdots & & \vdots & & \vdots \\ m_{i1} & \ldots & m_{ik} & \ldots & m_{in} \\ \vdots & & \vdots & & \vdots \\ m_{j1} & \ldots & m_{jk} & \ldots & m_{jn} \\ \vdots & & \vdots & & \vdots \\ m_{m1} & \ldots & m_{mk} & \ldots & m_{mn} \end{bmatrix}$$

Since we need two numbers to locate an entry in a matrix, we can think of an $m \times n$ matrix as a function from the set of pairs of numbers (i, j) to $\mathbb{R}$, where i runs from 1 to m and j runs from 1 to n. The domain in this function space is written as $\{1, ..., m\} \times \{1, \ldots, n\}$, a Cartesian product of sets.

The definition of the operations given in Theorem 5.1.5 boils down to adding matrices by adding coordinatewise. Similarly, multiplication by a scalar multiplies all of the entries in the matrix by that scalar. $\diamondsuit$

We can make one further step in generalizing Theorem 5.1.5 by noticing that the codomain of the functions need not be $\mathbb{R}$ but can also be any vector space over $\mathbb{R}$:

Theorem 5.1.6 : *If S is any set and $\mathcal{V}$ is a vector space over a field F, then the set of all functions from S to $\mathcal{V}$ is a vector space over F.*

The proof is just like the proof of Theorem 5.1.5 but uses the properties of $\mathcal{V}$ as a vector space where the properties of $\mathbb{R}$ were used in 5.1.5.

Exercises 5.1:

1. Show that the set of all formal power series in one variable with real coefficients is a vector space over $\mathbb{R}$. A formal power series is an expression of the form
$$\sum_{k=0}^{\infty} a_k x^k$$
There is no consideration of convergence. You add them term by term and also multiply by scalars term by term.

2. In computer science one often has occasion to use arrays of more than two dimensions. Spreadsheets sometimes have pages as well as rows and columns. Show that the set of all $3 \times 4 \times 2$ arrays of real numbers is a vector space over $\mathbb{R}$.

3. Show that the set of all polynomials with real coefficients and with only even powers of x appearing forms a vector space over $\mathbb{R}$.

4. Rational functions are those of the form $\frac{p(x)}{q(x)}$ with $p(x)$ and $q(x)$ polynomials with no common factors (a slightly different definition than is used in some calculus courses). Multiplication by scalars and addition are defined exactly as for rational numbers, with the insistence that common factors in numerator and denominator must be canceled. Does this form a vector space over $\mathbb{R}$?

5. Quadratic forms in x and y are expressions of the form $Ax^2+Bxy+Cy^2$. Show that quadratic forms in x and y are a vector space over $\mathbb{R}$.

6. Consider the set of sequences of real numbers with only finitely many non-zero members. Show how to make this into a vector space over $\mathbb{R}$.

7. Show that the space consisting only of the zero vector is a vector space.

8. Why can't the empty set be a vector space?

9. Show that $\mathbb{R}$ is a vector space over the reals. (See how the axioms for a vector space follow from those for a field.)

10. Show that $\mathbb{R}$ is a vector space over $\mathbb{Q}$ (the rational numbers) but not vice versa.

11. Show that $\mathbb{C}$ is a vector space over $\mathbb{R}$.

12. Show that $\{[x, y, x + 3y]\}$ is a vector space over $\mathbb{R}$.

13. Show that $\mathbb{R}^2$ with the usual addition but with the modified scalar multiplication $k * [x, y] = [kx, -ky]$ is not a vector space.

14. Show that $\mathbb{R}^2$ with the usual addition but with the modified scalar multiplication $k * [x, y] = [k^2x, k^2y]$ is not a vector space.

15. Show that $\mathbb{R}^2$ with the usual addition but with the modified scalar multiplication $k * [x, y] = [0, 0]$ is not a vector space.

16. Consider the following attempt to make $[0, 1)$ into a vector space:

    If $u, v \in [0, 1)$ and $k \in \mathbb{R}$ then let $u \oplus v$ be the fractional part of $u + v$ and $k \cdot v$ be the fractional part of kv.

    Which axioms for a vector space are satisfied and which are not?

17. Suppose we let $u, v \in [0, 1]$ and $k \in \mathbb{R}$ and define $u \oplus v = \min(1, u + v)$ and $k \cdot v = \max(\min(1, kv), 0)$. Which axioms for a vector space are satisfied and which are not?

5.2 Linear Transformations

5.2.1 Definitions

Linear algebra is only partly the study of vector spaces; indeed, the spaces are mostly of interest as the domains and codomains of maps. The appropriate kind of maps between vector spaces are those which preserve addition and scalar multiplication. They are called linear transformations. In later chapters we will see that linear transformations are closely related to matrices. Much of the power of linear algebra results from the interplay between concrete manipulation of matrices and the properties of linear transformations which give meaning to the manipulations.

Definition 5.2.1 *A function $L : \mathcal{V} \to \mathcal{W}$ is a* linear transformation *if $L(\vec{v}_1 + \vec{v}_2) = L(\vec{v}_1) + L(\vec{v}_2)$ and $L(k\vec{v}) = kL(\vec{v})$. The* domain *of L is $\mathcal{V}$; the* codomain *is $\mathcal{W}$.*

The properties of vector spaces make certain properties of linear transformations immediate. Since these properties make it easy to identify when maps are *not* linear transformations we start with them, and then give examples.

Proposition 5.2.1 *Any linear transformation must take $\vec{0}$ to $\vec{0}$.*

Proof:

$$L(\vec{0}) = L(0\vec{a}) = 0L(\vec{a}) = \vec{0}.$$

∎

Proposition 5.2.2 *Any linear transformation must preserve inverses.*

Proof:

$$L(-\vec{a}) = L(-1\vec{a}) = -1L(\vec{a}) = -L(\vec{a}).$$

∎

Next let us consider some examples of linear transformations between vector spaces. In order to be sure that we have made a reasonable definition we need to show that there are interesting examples of linear transformations.

5.2.2 Linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

Example: The function $f : \mathbb{R}^3 \to \mathbb{R}^2$ taking $[x, y, z]$ to $[x + y, z - 2x]$ is a linear transformation.

To see this we check that it preserves both addition and scalar multiplication:

$$\begin{aligned}
f([x_1, y_1, z_1] + [x_2, y_2, z_2]) &= f([x_1 + x_2, y_1 + y_2, z_1 + z_2]) \\
&= [(x_1 + x_2) + (y_1 + y_2), (z_1 + z_2) - 2(x_1 + x_2)] \\
&= [(x_1 + y_1) + (x_2 + y_2), (z_1 - 2x_1) + (z_2 - 2x_2)] \\
&= f([x_1, y_1, z_1]) + f([x_2, y_2, z_2])
\end{aligned}$$

and

$$\begin{aligned}
f(k[x, y, z]) &= f([kx, ky, kz]) \\
&= [kx + ky, kz - 2kx] \\
&= k[x + y, z - 2x] \\
&= kf([x, y, z])
\end{aligned}$$

$\diamondsuit$

Example: Projections

This example takes $\mathbb{R}^3$ to $\mathbb{R}^2$ by the map p taking $[x, y, z]$ to $[x, y]$. Again we must check that p preserves sums and scalar products.

$$\begin{aligned} p([x_1, y_1, z_1] + [x_2, y_2, z_2]) &= p([x_1 + x_2, y_1 + y_2, z_1 + z_2]) \\ &= [(x_1 + x_2), (y_1 + y_2)] \\ &= [x_1, y_1] + [x_2, y_2] \\ &= p([x_1, y_1, z_1]) + p([x_2, y_2, z_2]) \\ &\text{and} \\ p(k[x, y, z]) &= p([kx, ky, kz]) \\ &= [kx, ky] \\ &= k[x, y] \\ &= kf([x, y, z]) \end{aligned}$$

$\diamondsuit$

In Chapter 1 we studied many examples of linear transformations from $\mathbb{R}^2$ to itself. Recall that linear transformations can stretch the plane in either direction, rotate the plane, or do a combination of stretches and rotation.

Example: Multiplication by a matrix

When we worked with linear transformations from $\mathbb{R}^2$ to itself we found that it was often convenient to give the definition in terms of multiplication of a vector by a matrix. Suppose that the linear transformation from $\mathbb{R}^4$ to $\mathbb{R}^3$ takes $[w, x, y, z]$ to $[3w + 4x + y - z, x - y + 2z, 3w - 4y + z]$, we can write this as

$$\begin{bmatrix} 3 & 4 & 1 & -1 \\ 0 & 1 & -1 & 2 \\ 3 & 0 & -4 & 1 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3w + 4x + y - z \\ x - y + 2z \\ 3w - 4y + z \end{bmatrix}.$$

$\diamondsuit$

In general, left multiplication by an $m \times n$ matrix gives a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ using the following definition:

Definition 5.2.2 *If $\mathbf{A} = [[a_{ij}]]$ is an $m \times n$ matrix then*

$$\begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + \ldots + a_{1n}v_n \\ \vdots \\ a_{i1}v_1 + \ldots + a_{in}v_n \\ \vdots \\ a_{m1}v_1 \ldots a_{mn}v_n \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{j=1}^n a_{1j}v_j \\ \vdots \\ \sum_{j=1}^n a_{ij}v_j \\ \vdots \\ \sum_{j=1}^n a_{mj}v_j \end{bmatrix}.$$

Proposition 5.2.3 *Multiplication on the left by an $m \times n$ matrix $\mathbf{A}$ gives a linear transformation from $\mathbb{R}^n$ (thought of as columns) to $\mathbb{R}^m$ (again as columns).*

Proof:

We need to show that this operation preserves both addition and multiplication by scalars.

$$\begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} v_1 + w_1 \\ \vdots \\ v_n + w_n \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{j=1}^n a_{1j}(v_j + w_j) \\ \vdots \\ \sum_{j=1}^n a_{ij}(v_j + w_j) \\ \vdots \\ \sum_{j=1}^n a_{mj}(v_j + w_j) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}v_j \\ \vdots \\ \sum_{j=1}^n a_{ij}v_j \\ \vdots \\ \sum_{j=1}^n a_{mj}v_j \end{bmatrix} + \begin{bmatrix} \sum_{j=1}^n a_{1j}w_j \\ \vdots \\ \sum_{j=1}^n a_{ij}w_j \\ \vdots \\ \sum_{j=1}^n a_{mj}w_j \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

And similarly for multiplication by scalars:

$$\begin{aligned} \begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} kv_1 \\ \vdots \\ kv_n \end{bmatrix} &= \begin{bmatrix} \sum_{j=1}^n a_{1j} k v_j \\ \vdots \\ \sum_{j=1}^n a_{ij} k v_j \\ \vdots \\ \sum_{j=1}^n a_{mj} k v_j \end{bmatrix} \\ &= \begin{bmatrix} k\sum_{j=1}^n a_{1j} v_j \\ \vdots \\ k\sum_{j=1}^n a_{ij} v_j \\ \vdots \\ k\sum_{j=1}^n a_{mj} v_j \end{bmatrix} \\ &= k \begin{bmatrix} \sum_{j=1}^n a_{1j} v_j \\ \vdots \\ \sum_{j=1}^n a_{ij} v_j \\ \vdots \\ \sum_{j=1}^n a_{mj} v_j \end{bmatrix} \end{aligned}$$

∎

5.2.3 Linear transformations on polynomial spaces

Example: Differentiation of polynomials:

We know from basic calculus that the derivative of a polynomial is a polynomial, so the process of taking the first derivative is a function from the set of polynomials in x with real coefficients to itself. We observed in the last section that the set of polynomials is a vector space, which we called $\mathbb{R}[x]$. Differentiation gives a

function $D : \mathbb{R}[x] \to \mathbb{R}[x]$. Now for polynomials $\vec{a}$ and $\vec{b}$ we know that $D(\vec{a}+\vec{b}) = D(\vec{a}) + D(\vec{b})$ and for any $k \in \mathbb{R}$, $D(k\vec{a}) = kD(\vec{a})$. Thus D is a linear transformation. $\diamondsuit$

Example: Evaluation at a point

If we take the vector space $\mathbb{R}^S$ and an element $s \in S$ we get a linear transformation $E_s : \mathbb{R}^S \to \mathbb{R}$ which takes a function f to its value at s, $f(s)$. This is linear because of the way that we define addition and scalar multiplication of functions:

$$E_s(f+g) = (f+g)(s) = f(s) + g(s) = E_s(f) + E_s(g)$$

and

$$E_s(kf) = (kf)(s) = k(f(s)) = kE_s(f)$$

$\diamondsuit$

Example: Definite integral of polynomials

We can define a function A from $\mathbb{R}[x]$ to $\mathbb{R}$ by

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

It is an easy exercise in calculus to show that this is a linear transformation:

$$\begin{aligned} a(p(x)+q(x)) &= \int_0^1 p(x) + q(x)\ dx \\ &= \int_0^1 p(x)\ dx + \int_0^1 q(x)\ dx \\ &= A(p(x)) + A(q(x)) \end{aligned}$$

and

$$\begin{aligned} A(kp(x)) &= \int_0^1 k\ p(x)\ dx \\ &= k\int_0^1 p(x)\ dx \\ &= kA(p(x)) \end{aligned}$$

There was nothing special about the bounds 0 and 1.

$\diamondsuit$

Example: Integration of formal power series

When we study series representation of functions we see a theorem which states that inside the radius of convergence we can integrate power series term by term; with formal power series we do not have to worry about convergence, so we can define

$$L(\sum_{k=0}^{\infty} a_k x^k) = \sum_{k=1}^{\infty} \frac{a_{k-1}}{k} x^k.$$

Notice that we have chosen to make the constant term 0. This is necessary so that integration term by term preserves the zero vector. Term by term differentiation also gives a linear transformation. $\diamondsuit$

Exercises 5.2:

1. Show that the function L from $\mathbb{R}^2$ to $\mathbb{R}$ taking $[x, y]$ to $3x - 4y$ is a linear transformation.

2. Show that the projection map P from $\mathbb{R}^3$ to $\mathbb{R}^2$ taking $[x, y, z]$ to $[x, y]$ is a linear transformation.

3. Show that the inclusion $J : \mathbb{R} \to \mathbb{R}^2$ taking x to $[x, 0]$ is a linear transformation. Would this have been true if x went to $[x, 1]$ instead? What about $[x, 2x]$?

4. Show that the function $f : \mathbb{R}[x] \to \mathbb{R}[x]$ which takes a polynomial $p(x)$ to $p(x^2)$ is a linear transformation. What about the map taking $p(x)$ to $p(3x + 2)$?

5. In Chapter 1 we saw how to multiply a 2×2 matrix by a column vector. Define the map $M : \{2\times 2 \text{ matrices }\} \to \mathbb{R}^2$ which multiplies the matrix by the column vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Show this is a linear transformation.

For each of the maps in problems 6–11 either prove that the map is linear by checking the definition or give an example to show how it fails to be linear.

6. $f : \mathbb{R}^2 \to \mathbb{R}^2$ with $f[x, y] = [3x + 2y, x + y + 1]$

7. $g : \mathbb{R}^3 \to \mathbb{R}$ with $g[x, y, z] = [x + y - 2z]$

8. $h : \mathbb{R}[x] \to \mathbb{R}$ with $h(p) = p(1)$

9. $k : \mathbb{R}[x] \to \mathbb{R}[x]$ with $k(p) = p(x + 1)$

10. $p : \mathbb{R}^2 \to \mathbb{R}^3$ with $p[x, y] = [0, x + y, x - y]$

11. $m : \mathbb{R}^2 \to \mathbb{R}^3$ with $m[x, y] = [1, x + y, x - y]$

12. Recall that the equation for a line in $\mathbb{R}^3$ can be written in the form $(x, y, z) = (x_0, y_0, z_0) + t(d_1, d_2, d_3)$ where (x_0, y_0, z_0) is a point on the line and (d_1, d_2, d_3) is a direction vector. Suppose that $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation. Prove that the image under T of a line is a line or a point.

13. Show that if $L : \mathcal{V} \to \mathcal{W}$ is linear and $L(\vec{a})$ and $L(\vec{b})$ both equal $\vec{0}$ then $L(\vec{a} + \vec{b})$ and $L(k\vec{a})$ are $\vec{0}$ too.

14. Prove that if $L : \mathcal{V} \to \mathcal{W}$ is linear then $L(\vec{a} - \vec{b}) = L(\vec{a}) - L(\vec{b})$.

15. Show that $L : \mathcal{V} \to \mathcal{W}$ is 1-1 if $\vec{0}$ is the only vector with $L(\vec{a}) = \vec{0}$. (Hint: if $L\vec{a} = L\vec{b}$ consider $L(\vec{a} - \vec{b})$.)

16. Prove that L is linear if and only if $L(x\vec{a} + y\vec{b}) = xL(\vec{a}) + yL(\vec{b})$ for all scalars x and y and vectors $\vec{a}$ and $\vec{b}$.

17. (Project Problem) There is a calculus of finite differences which closely parallels the integral and differential calculus. It is based on four operations on sequences:

Forward difference	$\triangle : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$	with	$\triangle(a)(n) = a(n + 1) - a(n)$
Shift	$\sigma : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$	with	$\sigma(a)(n) = a(n + 1)$
Partial sums	$S : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$	with	$S(a)(n) = \sum_{k=0}^{n} a(k)$
Sum of first m+1	$S_m : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$	with	$S_m(a) = \sum_{k=0}^{m} a(k)$

(a) Show that all of these operators are linear transformations.

(b) Then prove that

$$(\triangle \circ S)(a) = \sigma(a)$$

and

$$(S_m \circ \triangle)(a) = a(m+1) - a(0)$$

for any sequence a, the analogs of the fundamental theorem of calculus.

5.3 Vector spaces over $\mathbb{Z}_2$ and $\mathbb{C}$

Our next examples are over somewhat less familiar fields. We will start with the complex numbers.

5.3.1 $\mathbb{C}$

The complex numbers are obtained from the reals by formally adding i, the square root of -1. This is done to guarantee that the equation $x^2+1=0$ has two roots (i and $-i$). In order to make the complex numbers a field we need to add many more points than just i and $-i$, however. The general form of a complex number is $a+bi$, where a and b are real numbers. The number a is called the real part of $a+bi$ and the number bi is called the imaginary part. Two complex numbers $a+bi$ and $c+di$ are equal if $a=c$ and $b=d$. We define the sum by

$$(a+bi)+(c+di) = (a+c)+(b+d)i$$

and the product of two complex numbers by

$$(a+bi)(c+di) = ac+bci+adi+bd(i^2)$$

multiplying the same way we multiply polynomials. Since $i^2 = -1$ this simplifies to

$$(a+bi)(c+di) = (ac-bd)+(bc+ad)i$$

It takes a bit of checking to see that this gives a commutative and associative multiplication and that the distributive law holds. The calculations involved are left to the exercises. The multiplicative identity is $1+0i$. To find inverses

we use the fact that every complex number $a+bi$ has a conjugate $(a+bi)^* = a-bi$ with the property that $(a+bi)(a-bi) = a^2+b^2$, a real number. If we think of the multiplicative inverse of $a+bi$ as $\frac{1}{a+bi}$ and then multiply both top and bottom by $a-bi$, then we get the expression $\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i$ for the inverse of $a+bi$. With these definitions the complex numbers form a field.

Complex numbers are used extensively in electrical engineering in the study of waves. In that setting vector spaces of functions of a complex variable are the most natural place to work.

Example: $\mathbb{C}^{\mathbb{R}}$

The space of all complex valued functions of a real variable is a vector space over $\mathbb{C}$. ♢

Example: Waves

The set of functions of the form

$$f(t) = \sum_{k=1}^{n} c_k e^{i\omega_k t}$$

where $c_k \in \mathbb{C}$ consists of the functions you can get by superposition of simple waves. It also forms a vector space over $\mathbb{C}$.

♢

The other place that complex vector spaces are used is in eigenvalue problems. The most useful property of the complex numbers is that any polynomial of degree n has n roots over the complex numbers (this is called algebraic closure and the fact that the complex numbers are algebraically closed is called the fundamental theorem of algebra). The vector spaces involved in eigenvalue problems are ones of the form $\mathbb{C}^n$. As in the case of $\mathbb{R}^n$, the axioms of a vector space over $\mathbb{C}$ follow directly from n applications of the field axioms for $\mathbb{C}$.

5.3.2 $\mathbb{Z}_2$

The field of integers modulo 2, $\mathbb{Z}_2$, is particularly easy to work with and has close ties to computations using digital circuitry. The easy way to think of

$\mathbb{Z}_2$ is as a set $\{0, 1\}$ with operations defined by the tables

$$\begin{array}{c|c|c|} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ \hline 1 & 1 & 0 \\ \hline \end{array} \text{ and } \begin{array}{c|c|c|} \times & 0 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 0 & 1 \\ \hline \end{array}$$

This is one of the few examples where one can reasonably verify a set of axioms by checking all of the possible cases. For instance, to verify associativity of addition we calculate eight sums both ways:

a	b	c	a + b	b + c	(a+b)+c	a+(b+c)
0	0	0	0	0	0	0
0	0	1	0	1	1	1
0	1	0	1	1	1	1
0	1	1	1	0	0	0
1	0	0	1	0	1	1
1	0	1	1	1	0	0
1	1	0	0	1	0	0
1	1	1	0	0	1	1

These operations have close ties to logic ($\times$ is *and*, $+$ is *exclusive or*) and have simple circuits which compute them.

As with the reals and the complex numbers, we can form a vector space of ordered n-tuples of elements of $\mathbb{Z}_2$. The vectors correspond exactly to the bit patterns in a digital signal.

Exercises 5.3:

1. Prove that the multiplication of complex numbers is associative.
2. A Laurent series in z expanded around a point $c \in \mathbb{C}$ is a series of the form
$$\sum_{k=-m}^{\infty} a_k (z-c)^k$$
Show that Laurent series form a vector space over $\mathbb{C}$.
3. Show complex conjugation $((a+b\,i)^* = (a-b\,i))$ preserves both addition and multiplication.

4. Recalling the examples of vector spaces over the reals, give two more examples of vector spaces over the complex numbers.

5. Show that any vector space over $\mathbb{C}$ can also be thought of as a vector space over $\mathbb{R}$.

6. Show that a linear transformation $L : \mathcal{V} \to \mathcal{W}$ between complex vector spaces is also linear if the vector spaces are thought of as being over $\mathbb{R}$.

7. Prove the distributive law for $\mathbb{Z}_2$.

8. We noted that $\mathbb{Z}_2$ has operations with logical meaning: $+$ is *exclusive or* and $\times$ is *and*. Show how to express the other logical connectives given in the following table using $+$, $\times$, and constants 0 and 1.

a	b	$\neg a$	$a \vee b$	$a \longrightarrow b$	$a \leftrightarrow b$	$a \vert b$	$a \downarrow b$
0	0	1	0	1	1	1	1
0	1	1	1	1	0	1	0
1	0	0	1	0	0	1	0
1	1	0	1	1	1	0	0

Chapter 6

Subspaces

Our examples of vector spaces in the last chapter provide the starting place for building vector spaces, but a much larger class of examples results from considering subsets of known vector spaces which are themselves vector spaces.

6.1 Definition and Examples

Example: A plane in $\mathbb{R}^3$

If we consider the subset W of $\mathbb{R}^3$ in which the third component is always 0,

$$W = \{\vec{a} \in \mathbb{R}^3 \,|\, \vec{a} = [a_1, a_2, 0]\}$$

we observe that we have a subset of $\mathbb{R}^3$ which is essentially the same as $\mathbb{R}^2$. Since we know that $\mathbb{R}^2$ is a vector space, $\mathcal{W}$ is a subset of $\mathbb{R}^3$ which is itself a vector space. We shall call $\mathcal{W}$ a *subspace* of $\mathbb{R}^3$. $\diamondsuit$

The general definition follows.

Definition 6.1.1 *A subset W of a vector space $\mathcal{V}$ is a* **subspace** *of $\mathcal{V}$ if and only if the vectors in W satisfy all the axioms for a vector space with respect to the same operations of addition and multiplication by a scalar as used for $\mathcal{V}$. When W is a subspace of $\mathcal{V}$ we write $\mathcal{W} \leq \mathcal{V}$.*

It is immediately obvious that $\mathcal{V}$ is itself always a subspace of $\mathcal{V}$, and, at the other extreme, it is almost as obvious that $\{\vec{0}\}$ is a subspace of $\mathcal{V}$. Thus, we always have $\{\vec{0}\} \leq \mathcal{V}$ and $\mathcal{V} \leq \mathcal{V}$. We refer to these as the *trivial* subspaces.

For a nontrivial subspace we consider the subset W of $\mathbb{R}^2$ with the property that the sum of the two components is 0: $W = \{[a_1, a_2] \mid a_1 + a_2 = 0\}$. Then $W \leq \mathbb{R}^2$. The proof of this assertion requires checking that all ten axioms of Definition 5.1.2 hold for vectors in W.

As a start we look at the two closure axioms. If $\vec{a} = [a_1, a_2] \in W$, $\vec{b} = [b_1, b_2] \in W$, and $k \in \mathbb{R}$ then

$$\vec{a} + \vec{b} = [a_1 + b_1, a_2 + b_2]$$

and

$$a_1 + b_1 + a_2 + b_2 = (a_1 + a_2) + (b_1 + b_2) = 0 + 0 = 0,$$

since $\vec{a}, \vec{b} \in W$. Also,

$$k\vec{a} = k[a_1, a_2] = [ka_1 ka_2],$$

and

$$ka_1 + ka_2 = k(a_1 + a_2) = k0 = 0.$$

Only eight more axioms to go! But in some sense the remaining axioms don't require any work, since we already know that they hold for vectors all vectors in $\mathbb{R}^2$, the only parts which could fail for the more restrictive collection W are the identity and inverse axioms (which require that specific vectors exist). In fact these cause no difficulty either, as the following theorem shows:

Theorem 6.1.1 *A nonempty subset W of a vector space $\mathcal{V}$ is a subspace if and only if W satisfies the two closure axioms.*

PROOF:

> The "if" part of the assertion is immediate, for if W is a subspace then *all* the axioms must be satisfied; thus, in particular, the two closure axioms must hold.
>
> To show the "only if" part, we must verify the validity of all ten axioms of a vector space for vectors in W. Our hypothesis is that the closure axioms hold. Also, since we are working with a subset of a vector space $\mathcal{V}$ and since commutativity, associativity, the two distributive laws, absorption, and identity for scalar

multiplication are valid for *all* vectors in $\mathcal{V}$ and all scalars in $\mathbb{R}$, we know that, in particular, they must be valid for vectors in W. Thus it remains only to show that $\vec{0} \in W$ and that along with every $\vec{a} \in W$ the inverse $-\vec{a}$ is also in W. But these both follow easily from the hypothesized closure under scalar multiples: if $\vec{a} \in W$ (and such exists because W is nonempty), then

$$\vec{0} = 0\vec{a} \in W$$

and

$$-\vec{a} = (-1)\vec{a} \in W.$$

∎

We now consider a few other examples of subspaces.

Example: A plane in space

As an extension of our first example, let $W \subseteq \mathbb{R}^2$ be defined by

$$W = \{[a_1, a_2] \,|\, k_1 a_1 + k_2 a_2 = 0\}$$

where k_1 and k_2 are arbitrary real numbers fixed in advance. We assert that the *subset* W is a *subspace* of $\mathbb{R}^2$. All we need to do is prove that closure holds for both operations: let $\vec{a} = [a_1, a_2], \vec{b} = [b_1, b_2] \in W$, $c \in \mathbb{R}$, then $\vec{a} + \vec{b} = [a_1 + b_1, a_2 + b_2]$ has

$$\begin{aligned} k_1(a_1 + b_1) + k_2(a_2 + b_2) &= (k_1 a_1 + k_2 a_2) + (k_1 b_1 + k_2 b_2) \\ &= 0 + 0 = 0. \end{aligned}$$

Thus the sum is in W. Similarly we can take a scalar product $c\vec{a} = [ca_1, ca_2]$ where $\vec{a} \in W$. Then

$$k_1(ca_1) + k_2(ca_2) = c(k_1 a_1 + k_2 a_2) = c0 = 0.$$

Thus $\vec{a} + \vec{b} \in W$ and $c\vec{a} \in W$, and W is indeed a subspace of $\mathbb{R}^2$. ♢

A comment is in order about the geometric interpretation of these examples. In each case $\mathcal{V}$ is simply the Euclidean plane. In the first the subspace

$$W = \{[x, y] \,|\, x + y = 0\};$$

in other words, W is the line through the origin with equation $y = -x$. For the second we generalized this to

$$W = \{[x, y] \mid k_1x + k_2y = 0\}$$

where k_1 and k_2 are any real numbers chosen in advance. This again gives a line through the origin (e.g., if $k_1 = 2$ and $k_2 = -1$, we have $y = 2x$). In general, letting k_1 and k_2 take on all possible values in $\mathbb{R}$, not both zero, we get all the straight lines through the origin (the x-axis appears as $k_1 = 0, k_2 = 1$). If $k_1 = k_2 = 0$, then $W = \mathcal{V}$ one of the trivial subspaces.

We note in passing that a line which does not pass through the origin cannot be a subspace, since any subspace must contain the zero vector.

Example: Polynomials of degree $\leq n$

> Let $\mathcal{V} = \mathbb{R}[x]$, the vector space of all polynomials in one variable with real coefficients, and consider $\mathbb{R}[x]_n$, the set of polynomials of degree at most n. Now the sum of two polynomials, each of degree at most n, is a polynomial of degree at most n. Also, the product of a number $k \in \mathbb{R}$ and a polynomial of degree $m \leq n$ is either the 0 polynomial (if $c = 0$; recall that the zero polynomial is considered to be of degree less than any n) or a polynomial of degree $m \leq n$ (if $k \neq 0$). Thus $\mathbb{R}[x]_n \leq \mathbb{R}[x]$. $\diamondsuit$

Notice that, for this example, the verification of the closure axioms depends upon knowledge of properties of polynomial functions. Similarly, in our next example we must draw on facts about continuous functions.

Example: Continuous functions from $\mathbb{R}$ to $\mathbb{R}$ form a subspace of $\mathbb{R}^{\mathbb{R}}$:

> In calculus we prove two limit theorems which are relevant to this example: that the limit of a sum of two functions is the sum of the limits, and that the limit of a constant times a function is the constant times the limit. These limit theorems tell us that the sum of continuous functions is again a continuous function and that a constant times a continuous function is a continuous function. $\diamondsuit$

Example: Differentiable functions from $\mathbb{R}$ to $\mathbb{R}$ form a subspace of $\mathbb{R}^{\mathbb{R}}$:

Also in calculus we consider the more restrictive class of differentiable functions from $\mathbb{R}$ to $\mathbb{R}$. One of the first theorems we prove is that the derivative of a sum is the sum of the derivatives, so the set of differentiable functions is closed under addition. Shortly thereafter (usually in the same lecture) we prove that a constant times a differentiable function is differentiable and indeed the derivative of kf is k times the derivative of f. Thus we have closure, so we have a subspace. $\diamondsuit$

Example: Spaces of linear transformations

The set of linear transformations from $\mathcal{V}$ to $\mathcal{W}$, written as $\mathrm{Hom}(\mathcal{V}, \mathcal{W})$, is a subspace of $\mathcal{W}^{\mathcal{V}}$ The map which takes everything to $\vec{0}$ is linear, so $\mathrm{Hom}(\mathcal{V}, \mathcal{W})$ is not empty. To show closure we need to show that the sum of two linear transformations is linear and that a constant times a linear transformation is a linear transformation. Both of these are straightforward: if $L_1 : \mathcal{V} \to \mathcal{W}$ and $L_2 : \mathcal{V} \to \mathcal{W}$ are linear then

$$\begin{aligned}
(L_1 + L_2)(\vec{a} + \vec{b}) &= L_1(\vec{a} + \vec{b}) + L_2(\vec{a} + \vec{b}) \\
&= L_1(\vec{a} + L_1(\vec{b}) + L_2(\vec{a}) + L_2(\vec{b}) \\
&= (L_1 + L_2)(\vec{a}) + (L_1 + L_2)(\vec{b}) \\
&\text{and} \\
(L_1 + L_2)(k\vec{a}) &= L_1(k\vec{a}) + L_2(k\vec{a}) \\
&= kL_1(\vec{a}) + kL_2(\vec{a}) \\
&= k(L_1 + L_2)(\vec{a})
\end{aligned}$$

This shows that $L_1 + L_2$ is linear.

To see that rL_1 is linear involves a similar calculation:

$$\begin{aligned}
(rL_1)(\vec{a} + \vec{b}) &= r(L_1(\vec{a} + \vec{b})) \\
&= r(L_1(\vec{a}) + L_1(\vec{b})) \\
&= rL_1(\vec{a}) + rL_1(\vec{b}) \\
&\text{and}
\end{aligned}$$

$$\begin{array}{rcl} (rL_1)(k\vec{a}) & = & r(L_1(k\vec{a})) \\ & = & r(kL_1(\vec{a}) \\ & = & (rk)L(\vec{a} \\ & = & k((rL_1)(\vec{a})) \end{array}$$

◊

The special case $\mathrm{Hom}(\mathcal{V}, \mathbb{R})$ is called the **dual space** of $\mathcal{V}$ and is written $\mathcal{V}^*$.

6.1.1 Images and kernels of linear transformations

If $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation, then the set of all vectors in the codomain ($\mathcal{W}$) which are values of L at elements of $\mathcal{V}$ forms a subspace of $\mathcal{W}$. This example is important enough to warrant formal inclusion as a definition and a theorem:

Definition 6.1.2 *If $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation then the set $Im(L) = \{\vec{w} \in \mathcal{W} \mid$ There is a $\vec{v} \in \mathcal{V}$ such that $\vec{w} = L(\vec{v})\}$.*

Theorem 6.1.2 *If $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation then the set $Im(L)$ is a subspace of $\mathcal{W}$.*

PROOF:

> We need to show that $\mathrm{Im}(L)$ is closed under both sums and scalar products. Now if $\vec{w}_1$ and $\vec{w}_2$ are in $\mathrm{Im}(L$ then there are vectors $\vec{v}_1$ and $\vec{v}_2$ in $\mathcal{V}$ with $\vec{w}_1 = L(\vec{v}_1)$ and $\vec{w}_2 = L(\vec{v}_2)$ so
>
> $$\begin{array}{rcl} \vec{w}_1 + \vec{w}_2 & = & L(\vec{v}_1) + L(\vec{v}_2) \\ & = & L(\vec{v}_1 + \vec{v}_2) \end{array}$$
>
> so $\vec{w}_1 + \vec{w}_2 \in \mathrm{Im}(L)$. Similarly $\mathrm{Im}(L)$ is closed under scalar multiples because linear transformations preserve scalar multiplication.
>
> ∎

The other important subspace associated with a linear transformation is its kernel:

Definition 6.1.3 *If $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation then the set of all vectors in $\mathcal{V}$ which are taken to $\vec{0}$ by L is called the* **kernel** *of L and is written Ker(L).*

Theorem 6.1.3 *If $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation then Ker(L) is a subspace of $\mathcal{V}$.*

PROOF:

Since $L(\vec{0}) = \vec{0}$, we know that the kernel is not empty. Thus to show it is a subspace all we need to do is show that it is closed under sums and scalar multiples. Now if $\vec{v}_1$ and $\vec{v}_2$ are in Ker(L) then

$$L(\vec{v}_1 + \vec{v}_2) = L(\vec{v}_1) + L(\vec{v}_2) = \vec{0} + \vec{0} = \vec{0}$$

so $\vec{v}_1 + \vec{v}_2 \in \text{Ker}(L)$ as well. To see that $k\vec{v}_1 \in \text{Ker}(L)$ we note that

$$L(k\vec{v}_1) = kL(\vec{v}_1) = k\vec{0} = \vec{0}$$

as needed. ∎

Important examples of kernels are given by the sets of solutions to various kinds of linear problems.

Example: Solutions to a single homogeneous linear equation $ax + by + cz = 0$

By a homogeneous linear equation we mean a linear equation in which the constant term is 0. The set of solutions to such an equation can be thought of as the kernel of the linear transformation $L : \mathbb{R}^3 \to \mathbb{R}$ taking $[x, y, z]$ to $ax + by + cz$. It is easy to check that this map is a linear transformation and that the kernel is precisely the set of solutions to the linear equation. $\diamondsuit$

Example: Solutions to a pair of simultaneous homogeneous equations in three variables:

For concreteness let us consider the solutions to

$$\begin{aligned} x+y-z &= 0 \\ x-2y+3z &= 0. \end{aligned}$$

We would like to use the same definitions of addition and scalar multiplication that we used for $\mathbb{R}^3$, since after all the solutions are ordered triples. We need to check to see that the sum of two solutions, say $[a,b,c]$ and $[d,e,f]$, is again a solution. Thus we assume that

$$\begin{aligned} a+b-c &= 0 \\ a-2b+3c &= 0 \\ e+f-g &= 0 \\ e-2f+3g &= 0 \end{aligned}$$

Adding the equations gives

$$\begin{aligned} (a+e)+(b+f)-(c+g) &= 0 \\ (a+e)-2(b+f)+3(c+g) &= 0 \end{aligned}$$

so the sum gives another solution. Similarly if we multiply each term of the equations by k we get

$$\begin{aligned} k(a+b-c) &= k0 \\ k(a-2b+3c) &= k0 \\ ka+kb-kc &= 0 \\ ka-2kb+3kc &= 0 \end{aligned}$$

so multiplying a solution by a constant gives another solution. Thus we get a subspace of $\mathbb{R}^3$.

We can recognize this as an example of a kernel of a linear transformation if we observe that the function taking $\mathbb{R}^3$ to $\mathbb{R}^2$ by taking $[x,y,z]$ to $[x+y-z, x-2y+3z]$ is a linear transformation. The set of vectors taken to $[0,0]$ is precisely the solution to the system of homogeneous equations. $\diamondsuit$

Example: Solution to homogeneous linear differential equations

The most important application of the fact that differentiable functions form a vector space comes in the attempt to describe the set of all solutions of a homogeneous (i.e. constant term = 0) linear differential equation. As an example consider the solutions of the equation $dy = y\ dx$, or $y' - y = 0$. If f is a solution then

$$f'(x) - f(x) = 0.$$

If f is a solution and k is a constant then

$$(kf)'(x) - kf(x) = kf'(x) - kf(x) = k0 = 0$$

so kf is again a solution. Similarly if both f and g are solutions then

$$\begin{aligned} f'(x) - f(x) &= 0 \\ g'(x) - g(x) &= 0 \\ (f+g)'(x) - (f+g)(x) &= f'(x) + g'(x) - f(x) - g(x) \\ &= f'(x) - f(x) + g'(x) - g(x) \\ &= 0 \end{aligned}$$

Thus the sum and scalar product of solutions are again solutions. This gives a subspace of the differentiable functions from $\mathbb{R}$ to itself.

This example too can be thought of as the kernel of a linear transformation from differentiable functions to functions: take f to $f' - f$. $\diamondsuit$

6.1.2 Some further linear transformations

With the addition of these subspaces to our vocabulary of vector spaces we can now identify a few more examples of linear transformations:

Example: A definite integral

The set $C[0,1]$ of all continuous functions from the closed unit interval $[0,1]$ to the reals is a vector space. The map $L : C[0,1] \to$

$\mathbb{R}$ which takes a function f to the definite integral $\int_0^1 f(x)dx$ is a linear transformation. This summarizes two handy lemmas from calculus:

$$\int_0^1 (f+g)(x)dx = \int_0^1 f(x)dx + \int_0^1 g(x)dx$$

and

$$\int_0^1 kf(x)dx = k\int_0^1 f(x)dx$$

for k a real number. $\diamondsuit$

Example: Inclusions of subspaces

If $\mathcal{S}$ is a subspace of the vector space $\mathcal{V}$, then the map from $\mathcal{S}$ to $\mathcal{V}$ taking a vector to itself, called the inclusion map, is a linear transformation because the definition of the operations is the same in $\mathcal{S}$ as it is in $\mathcal{V}$. $\diamondsuit$

Exercises 6.1:

1. Determine whether the subset A of the vector space $\mathcal{V}$ is a subspace. If not, why not?

   (a) $\mathcal{V} = \mathbb{R}^2$, $A = \{[a,b] \mid a + 2b = 0\}$

   (b) $\mathcal{V} = \mathbb{R}^2$, $A = \{[a,b] \mid a + 2b = 1\}$

   (c) $\mathcal{V} = \mathbb{R}^3$, $A = \{[x,y,z] \mid 2x + 3y - 5z = 0\}$

   (d) $\mathcal{V} = \mathbb{R}^3$, $A = \{[x,y,z] \mid 2x + 3y - 5z = 4\}$

   (e) $\mathcal{V} = \mathbb{R}[x]_3$ (polynomials of degree at most 3), $A = \{p \mid p(1) = 0\}$

   (f) $\mathcal{V} = \mathbb{R}[x]_3, A = \{p(x) \mid \text{the coefficient of } x^2 \text{ is } 0\}$

   (g) $\mathcal{V} =$ differentiable functions from $\mathbb{R}$ to $\mathbb{R}$, $\mathcal{A} = \{f \mid f'(x) = 0\}$

   (h) $\mathcal{V} =$ differentiable functions from $\mathbb{R}$ to $\mathbb{R}$, $\mathcal{A} = \{f \mid f'(x) - (x - 1)f(x) = 0\}$

2. Consider $\mathbb{R}^2$ as a vector space. Which of the following curves have graphs which are subspaces?
   (a) $y = 2x$
   (b) $y = 2x + 1$
   (c) $y = x^2$
   (d) $x^2 + y^2 = 0$
   (e) $x^2 + y^2 = 25$
   (f) $x = 0$

3. Consider $\mathbb{R}^3$ as a vector space. Which of the following are subspaces:
   (a) The z-axis
   (b) the xy-plane
   (c) the plane $z = 1$
   (d) the sphere $x^2 + y^2 + z^2 = 4$
   (e) the line $x = 1 + t, y = 2t, z = t - 4$

4. Prove that if $\mathcal{W} \leq \mathcal{V}$ and $\mathcal{U} \leq \mathcal{W}$, then $\mathcal{U} \leq \mathcal{V}$.

5. Let $\mathcal{V}$ be a vector space, $\mathcal{W}$ and $\mathcal{U}$ subspaces of $\mathcal{V}$. Prove that if $\mathcal{U}$ is a subset of $\mathcal{W}$, then $\mathcal{U}$ is a subspace of $\mathcal{W}$.

6. Describe the kernel of the following linear transformations:
   (a) $L : \mathbb{R}^2 \to \mathbb{R}$ with $L([x, y]) = 2x - 3y$
   (b) differentiation of polynomials
   (c) $L : \mathbb{R}^3 \to \mathbb{R}^4$ with $L([x, y, z]) = [x+y, y-z, x-2y+z, 3x-2y+z]$

7. Show that the set of all functions from $\mathbb{R}$ to $\mathbb{R}$ which have derivative at 0 equal to 0 is a vector space. Would the same be true if we asked for derivative at 0 equal to 1?

8. Show that the set of all functions which have second derivative equal to the zero function is a vector space.

9. Show that the set of polynomials of degree exactly 3 is not a vector space.

10. Show that the set of periodic functions with period p forms a subspace of $\mathbb{R}^{\mathbb{R}}$. Recall that f is periodic with period p if $f(x+p) = f(x)$ for all $x \in \mathbb{R}$.

11. (Based on problem 17 of section 5.2) Find the image and the kernel of each of the operators $\triangle : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, $\sigma : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, $S : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, and $S_m : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$.

12. (Project Problem) In this problem you will show that $\mathcal{V}$ can naturally be thought of as isomorphic to a subspace of $\mathcal{V}^{**}$.

    (a) Show that for each $\vec{a} \in \mathcal{V}$ the function "evaluate at $\vec{a}$" denoted by $E_{\vec{a}}$ which takes a linear transformation L to its value at $\vec{a}$, $L(\vec{a}) \in \mathbb{R}$, is in $\mathcal{V}^{**}$.

    (b) Show that the function $E : \mathcal{V} \to \mathcal{V}^{**}$ taking $\vec{a}$ to $E_{\vec{a}}$ is linear.

    (c) Show that the function E is one to one: if $E_{\vec{a}} = E_{\vec{b}}$ then $\vec{a} = \vec{b}$.

6.2 Subspaces spanned by a set of vectors

We have seen that the trivial subspaces of $\mathcal{V}$, $\{\vec{0}\}$ and $\mathcal{V}$, constitute the extreme cases of subspaces of a vector space $\mathcal{V}$, the smallest and the largest, respectively. In our next example we consider what might be referred to as a smallest non-trivial subspace.

Example: The subspace spanned by a single vector:

Let $\vec{a} \neq \vec{0}$ be a fixed but arbitrary vector of $\mathcal{V}$, let $\mathrm{Span}(\{\vec{a}\}) = \{\vec{b} \,|\, \vec{b} = k\vec{a}, k \in \mathbb{R}\}$. Then $\mathrm{Span}(\{\vec{a}\}) \leq \mathcal{V}$. The proof is straightforward. If

$$\begin{aligned} \vec{b}_1 &= k_1\vec{a}, \\ \vec{b}_2 &= k_2\vec{a} \end{aligned}$$

then

$$\begin{aligned} \vec{b}_1 + \vec{b}_2 &= k_1\vec{a} + k_2\vec{a} \\ &= (k_1 + k_2)\vec{a} \end{aligned}$$

So sums of elements of $\text{Span}(\{\vec{a}\})$ are in $\text{Span}(\{\vec{a}\})$. Similarly we can show that $\text{Span}(\{\vec{a}\})$ is closed under scalar multiplication:

$$\begin{aligned} c\vec{b}_1 &= c(k_1\vec{a}) \\ &= (ck_1)\vec{a} \in \text{Span}(\{\vec{a}\}). \end{aligned}$$

Thus $\text{Span}(\{\vec{a}\})$ is a subspace of $\mathcal{V}$.

As a particular instance, suppose $\mathcal{V} = \mathbb{R}^2$, the Euclidean plane, and $\vec{a} = [a_1, a_2] \neq [0, 0]$. Then $\text{Span}(\{\vec{a}\}) = \{k\vec{a} \mid k \in \mathbb{R}\}$ is simply the straight line through 0 and the point (a_1, a_2). Thus every point in the plane determines a subspace of $\mathbb{R}^2$, these subspaces being, for points other than the origin, the line through the origin and the point. It is clear from this remark that two points on the same line through the origin will determine the same subspace. $\diamondsuit$

The subspace illustrated in the preceding example is a particular case of a more general concept. As we saw, we could generate a subspace by taking all scalar multiples of a fixed nonzero vector. Why not try the same technique using more than one vector?

Definition 6.2.1 *Let A be a nonempty subset of vectors in $\mathcal{V}$. A* **linear combination** *of vectors in A is a vector $\vec{b}$ of the form*

$$\vec{b} = c_1\vec{a}_1 + \ldots + c_m\vec{a}_m,$$

where $c_1, \ldots, c_m$ are scalars and $\vec{a}_1, \ldots, \vec{a}_m \in A$. (Note that A can be an infinite subset of $\mathcal{V}$, but the vectors used in a linear combination are a finite subset of A.)

As an illustration, consider the following example:

Example: A linear combination in $\mathbb{R}^3$

Let $\mathcal{V} = \mathbb{R}^3$, $A = \{\vec{a}_1 = [1, 3, -1], \vec{a}_2 = [2, 0, 3]\}$. Taking $c_1 = 2$ and $c_2 = 3$ we find that

$$\begin{aligned} \vec{b} &= c_1\vec{a}_1 + c_2\vec{a}_2 \\ &= 2[1, 3, -1] + 3[2, 0, 3] \\ &= [2, 6, -2] + [6, 0, 9] \\ &= [8, 6, 7] \end{aligned}$$

is a linear combination of $\vec{a}_1$ and $\vec{a}_2$. $\diamondsuit$

Notice that if $\vec{a}$ is in A then $\vec{a}$ is itself a linear combination of vectors in A, by writing $\vec{a} = 1\vec{a}$.

Example: A set spanning $\mathbb{R}^3$

> Let $\mathcal{V} = \mathbb{R}^3$, $A = \{\vec{e}_1 = [1,0,0], \vec{e}_2 = [0,1,0], \vec{e}_3 = [0,0,1]\}$. Then *every* vector in $\mathbb{R}^3$ is a linear combination of vectors in A, since if $\vec{a} = [a_1, a_2, a_3]$ then $\vec{a} = a_1\vec{e}_1 + a_2\vec{e}_2 + a_3\vec{e}_3$. $\diamondsuit$

Definition 6.2.2 *Let S be a subset of $\mathcal{V}$. If S is not empty then the span of S, written Span(S), is the set of all linear combinations of vectors in S. If S is empty Span(S) is the vector space $\{\vec{0}\}$.*

Example: Finding the span of a set

> Let $\mathcal{V}$ be $\mathbb{R}^3$, $A = \{[1,3,-1],[2,0,3]\}$. Then Span(A) is
>
> $$\{\vec{b} \,|\, \vec{b} = c_1[1,3,-1] + c_2[2,0,3]\}$$
> $$= \{\vec{b} \,|\, \vec{b} = [c_1, 3c_1, -c_1] + [2c_2, 0, 3c_2]\}$$
> $$= \{\vec{b} \,|\, \vec{b} = [c_1 + 2c_2, 3c_1, -c_1 + 3c_2]\}$$
>
> $\diamondsuit$

A careful look at these examples leads us to conjecture that the span of a set might be a subspace of $\mathcal{V}$. Before we can prove this assertion we must attend to a minor technical detail. Suppose A is a non-empty subset of $\mathcal{V}$ and suppose $\vec{b}$ is a linear combination of $\vec{a}_1, \vec{a}_2$, and $\vec{c}$ a linear combination of $\vec{a}_2, \vec{a}_3, \vec{a}_4$, where $\vec{a}_1, \ldots, \vec{a}_4 \in A$:

$$\begin{aligned} \vec{b} &= k_1\vec{a}_1 + k_2\vec{a}_2 \\ \vec{c} &= h_2\vec{a}_2 + h_3\vec{a}_3 + h_4\vec{a}_4. \end{aligned}$$

Using zeroes as coefficients, where appropriate, we can then write

$$\begin{aligned} \vec{b} &= k_1\vec{a}_1 + k_2\vec{a}_2 + 0\vec{a}_3 + 0\vec{a}_4 \\ \vec{c} &= 0\vec{a}_1 + h_2\vec{a}_2 + h_3\vec{a}_3 + h_4\vec{a}_4 \end{aligned}$$

Since any linear combination from A involves at most a finite subset of A, we can always use a similar device in writing vectors in the span of A.

Proposition 6.2.1 *Let A be a subset of $\mathcal{V}$. Then the span of A is a subspace of $\mathcal{V}$.*

PROOF:

If A is empty, then by definition $\text{Span}(A) = \{\vec{0}\}$, a trivial subspace of $\mathcal{V}$.

If A is nonempty, use the remark preceding the statement of the proposition to show closure. Let $\vec{u}, \vec{v} \in \text{Span}(A)$, $c \in \mathbb{R}$; then we can write for some positive integer m,

$$\begin{aligned}\vec{u} &= k_1\vec{a}_1 + \ldots + k_m\vec{a}_m\\ \vec{v} &= h_1\vec{a}_1 + \ldots + h_m\vec{a}_m.\end{aligned}$$

Consequently,

$$\begin{aligned}\vec{u} + \vec{v} &= (k_1\vec{a}_1 + \ldots + k_m\vec{a}_m) + (h_1\vec{a}_1 + \ldots + h_m\vec{a}_m)\\ &= (k_1 + h_1)\vec{a}_1 + \ldots + (k_m + h_m)\vec{a}_m \in \text{Span}(A),\end{aligned}$$

and

$$\begin{aligned}c\vec{u} &= c(k_1\vec{a}_1 + \ldots + k_m\vec{a}_m)\\ &= (ck_1)\vec{a}_1 + \ldots + (ck_m)\vec{a}_m \in \text{Span}(A).\end{aligned}$$

Thus $\text{Span}(A) \leq \mathcal{V}$. ∎

This result gives us a way of creating or generating subspaces of a vector space. For this reason the span of a subset A, $\text{Span}(A)$, is also referred to as the subspace generated by A. If $\mathcal{W} = Span(A)$ then we say that A is a generating set for $\mathcal{W}$.

Example: A plane through the origin is the span of a set of two vectors.

For example if $A = \{[1, 3, -1], [2, 0, 3]\} \subset \mathbb{R}^3$, then $\text{Span}(A)$ is exactly the plane through the origin and the points (1,3,-1), (2,0,3). ♢

Each matrix gives rise to two significant subspaces:

Definition 6.2.3 *If* $\mathbf{M}$ *is a* $m \times n$ *matrix then the* **row space** *of* $\mathbf{M}$ *is the subspace of* $\mathbb{R}^n$ *spanned by the row vectors of* $\mathbf{M}$.

Definition 6.2.4 *If* $\mathbf{M}$ *is a* $m \times n$ *matrix then the* **column space** *of* $\mathbf{M}$ *is the subspace of* $\mathbb{R}^m$ *spanned by the column vectors of* $\mathbf{M}$.

Example: Row and column spaces

For the matrix

$$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 \\ 2 & 1 & 5 & 3 \end{bmatrix}$$

the row space is spanned by $\{[1,0,2,1],[0,1,1,1],[2,1,5,3]\}$. Now $[2,1,5,3] = 2[1,0,2,1]+[0,1,1,1]$ so we can actually get the whole row space by looking at the span of the first two rows. The row space is $\text{Span}(\{[1,0,2,1],[0,1,1,1]\})$.

The column space is spanned by

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} \right\}.$$

It is easy to see that the last two vectors on this list can be obtained as linear combinations of the first two:

$$2\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

Notice that the row space and the column space both needed the same number of vectors in a minimal spanning set. $\diamondsuit$

The subspace generated by a set of vectors is the smallest subspace which contains those vectors: Suppose $A \subseteq \mathcal{V}$ and $\mathcal{W}$ is a subspace of $\mathcal{V}$ which contains A. Then $\text{Span}(A) \leq \mathcal{W}$. Thus, any subspace which contains the vectors in A must be at least as big as $\text{Span}(A)$.

Proposition 6.2.2 *If $A \subseteq \mathcal{W} \leq \mathcal{V}$ then Span$(A) \leq \mathcal{W}$.*

PROOF:

To show this we let $\vec{b} \in \text{Span}(A)$. Then

$$\vec{b} = k_1\vec{a}_1 + \ldots + k_m\vec{a}_m, \vec{a}_1 \ldots, \vec{a}_m \in A.$$

Since $A \subseteq \mathcal{W}$ and $\mathcal{W}$ is a subspace, it follows that $\vec{b} \in \mathcal{W}$. Thus every vector in Span(A) is also in $\mathcal{W}$. ∎

A little reflection about the span of a set of vectors should produce some questions which deserve consideration. We turn to a few of these. The first has to do with economy or efficiency. To illustrate the point we have in mind, consider the case where $A = \{\vec{a}_1, \vec{a}_2\}$ but $\vec{a}_2 = c\vec{a}_1$ for some $c \in \mathbb{R}$. Then, if $\vec{b} \in \text{Span}(A)$, $\vec{b} = k_1\vec{a}_1 + k_2\vec{a}_2$, we see that we can write

$$\vec{b} = k_1\vec{a}_1 + k_2(c\vec{a}_1) = (k_1 + ck_2)\vec{a}_1 \in \text{Span}(\{\vec{a}_1\}).$$

We have that $\text{Span}(A) = \text{Span}(\{\vec{a}_1\})$. In other words, the inclusion of $\vec{a}_2$ in A is unnecessary as far as the generation of Span(A) is concerned.

More generally, we can assert the following.

Proposition 6.2.3 *Let $A = \{\vec{a}_1, \ldots, \vec{a}_m\} \subseteq \mathcal{V}$. Suppose that some vector in A, say $\vec{a}_m$, is a linear combination of the others, $\vec{a}_m = c_1\vec{a}_1 + \ldots + c_{m-1}\vec{a}_{m-1}$. If $A_1 = \{\vec{a}_1 \ldots, \vec{a}_{m-1}\}$, then Span$(A_1)$ = Span(A).*

The proof is left as an exercise.

Notice that it is always the case that $A \subseteq \text{Span}(A)$. Could it ever happen that $A = \text{Span}(A)$? Obviously, since Span(A) is a subspace of $\mathcal{V}$, a necessary condition for this equality to hold is that A itself be a subspace. As it happens, this condition is also sufficient.

Proposition 6.2.4 *Let $\mathcal{A} \subseteq \mathcal{V}$. Then Span$(A) = A$ if and only if A is a subspace of $\mathcal{V}$.*

PROOF:

If Span(A) $= A$, we must show that A is a subspace. But this is trivial, since Span(A) is a subspace and Span(A) $= A$. On the other hand, suppose A is a subspace. We always have $A \subseteq$ Span(A). We also have $A \subseteq A$, and in this case A is a subspace of $\mathcal{V}$. Thus, Span(A) $\leq A$. Consequently, $A =$ Span(A).
∎

Exercises 6.2:

1. Find the subspace of $\mathbb{R}^3$ spanned by the following sets of vectors:
   (a) $\{[1,2,3],[1,2,4]\}$
   (b) $\{[3,-2,1],[3,2,1],[-3,2,1]\}$
   (c) $\{[1,1,2],[-1,0,3],[1,2,7]\}$
2. In $\mathbb{R}[\mathrm{x}]_3$
   (a) is $x+1$ in the subspace Span($\{x^2-1, x^2+x+2, 3\}$)
   (b) is x^2+2x+1 in Span($\{1, x, x^2-1\}$)
   (c) find a spanning set for the whole space (bonus if you find an elegant, efficient solution).
3. Prove Proposition 6.2.3.
4. Describe the row space and the column space of the following matrices:
   (a) $\begin{bmatrix} 2 & 1 & 4 & 3 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 3 & 3 \end{bmatrix}$
   (b) $\begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 3 \\ 2 & 2 \end{bmatrix}$
   (c) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

5. Let $S \subseteq \mathbb{R}^{\mathbb{N}}$ be the set of strictly increasing sequences (that is $a_{k+1} > a_k$ for all k). What is $\text{Span}(S)$?

6. Let $S \subseteq \mathbb{R}^{\mathbb{R}}$ be the set of strictly increasing functions (that is $f(x_1) > f(x_2)$ whenever $x_1 > x_2$). What is $\text{Span}(S)$?

7. Let $S \subseteq \mathbb{R}^{\mathbb{N}}$ be the set of eventually constant sequences. What is $\text{Span}(S)$?

8. Is $Span(S \cap T) = Span(S) \cap Span(T)$? Give either a proof or a counterexample.

9. Is $Span(S \cup T) = Span(S) \cup Span(T)$? Give either a proof or a counterexample.

10. Prove that if S is a spanning set for $\mathcal{V}$ and $S \subseteq T$ then T is also a spanning set for $\mathcal{V}$.

11. Show that if S is a set of vectors in $\mathcal{V}$, $\vec{v} \in \mathcal{V}$, and there are two different linear combinations of vectors in S giving $\vec{v}$, then there are two different linear combinations of vectors in S giving $\vec{0}$.

12. (Project Problem) The brick function associated with the interval $[a, b)$ is

$$\beta_{[a,b)} = \begin{cases} 1 & \text{if } a \leq x < b \\ 0 & \text{otherwise} \end{cases}$$

We aim to describe the subspace $\mathcal{S}$ of $\mathbb{R}^{\mathbb{R}}$ spanned by the brick functions.

(a) Show that if $f \in \mathcal{S}$ then f has bounded support. (The support of a function is the set of all x where $f(x) \neq 0$.)

(b) What can you say about the range of a function in $\mathcal{S}$?

(c) It is clear that

$$\int_{-\infty}^{\infty} \beta_{[a,b)}(x)\ dx = b - a.$$

Use this and the linearity of the integral to define

$$\int_{-\infty}^{\infty} f(x)\ dx$$

for any f in $\mathcal{S}$.

6.3 Error Correcting Codes

A useful application of vectors over $\mathbb{Z}_2$ is the theory of error correcting codes. When a stream of binary signals is sent along a transmission line, or when a bit pattern is retrieved from a computer memory, occasional errors will occur. If the communication line is not too noisy, or the memory chip is only suffering from transient disruption, the errors should be relatively infrequent, mostly happening singly, rather than in bursts. Early computers shut down when such an error occurred, leading to much frustration among early computer users. In 1948 Richard W. Hamming started the theory of error correcting codes by adding checksums to the message. The idea was to break the message into patterns of four bits and then intersperse new bits so that of seven bits sent the sums of the positions 1,3,5, and 7 add up to 0, as do the sums of bits 2,3,6 and 7 and the sum of bits 4,5,6, and 7. the check bits are put in positions 1,2, and 4. The resulting code is easy to decode: the message is in bits 3,5,6, and 7. It has enough checking that, by means of some linear algebra over $\mathbb{Z}_2$, if a single bit out of the seven is transmitted in error, it can be detected and corrected.

We can think of this process as in terms of linear transformations on vector spaces over $\mathbb{Z}_2$. The four bit words are precisely the members of the vector space $\mathbb{Z}_2^4$. Encoding is done by the linear transformation $C : \mathbb{Z}_2^4 \to \mathbb{Z}_2^7$ which takes a message word to a codeword by

$$[m_1, m_2, m_3, m_4] \mapsto [m_1{+}m_2{+}m_4, m_1{+}m_3{+}m_4, m_1, m_2{+}m_3{+}m_4, m_2, m_3, m_4]$$

Its image gives the subspace of $\mathbb{Z}_2^7$ consisting of the code words. Decoding uses the linear transformation $D : \mathbb{Z}_2^7 \to \mathbb{Z}_2^4$ with

$$D([c_1, c_2, c_3, c_4, c_5, c_6, c_7]) = [c_3, c_5, c_6, c_7]$$

Notice that $DC = \mathrm{id}_{\mathbb{Z}_2^4}$

An essential feature of the Hamming (7,4) code thus described is that the subspace of code words can also be described as a kernel. The code words can be described as the set of 7-tuples $[a_1, a_2, a_3, a_4, a_5, a_6, a_7]$ in $\mathbb{Z}_2^7$ with

$$\begin{aligned} a_1 + a_3 + a_5 + a_7 &= 0 \\ a_2 + a_3 + a_6 + a_7 &= 0 \\ a_4 + a_5 + a_6 + a_7 &= 0 \end{aligned}$$

This is just the solutions to a homogeneous system of linear equations in $\mathbb{Z}_2$. It can be thought of as the set of vectors that the linear transformation $S : \mathbb{Z}_2^7 \to \mathbb{Z}_2^3$, called the syndrome, with

$$S([a_1, a_2, a_3, a_4, a_5, a_6, a_7]) = [a_4+a_5+a_6+a_7, a_2+a_3+a_6+a_7, a_1+a_3+a_5+a_7]$$

takes to $[0, 0, 0]$.

The particular choice of equations is a clever one (it was patented in 1951 along with the circuitry to do the coding and decoding) which allows us to identify where a single error in transmission has occurred with a small amount of calculation. Transmitted codewords are elements of $\mathbb{Z}_2^7$ and since the syndrome is a linear transformation we can see what the effect of an error in a single bit would be by looking at seven possible single bit error patterns:

Bit pattern	Syndrome
1000000	$[0, 0, 1]$
0100000	$[0, 1, 0]$
0010000	$[0, 1, 1]$
0001000	$[1, 0, 0]$
0000100	$[1, 0, 1]$
0000010	$[1, 1, 0]$
0000001	$[1, 1, 1]$

If an error has occurred in bit 5 of the transmitted codeword, for instance, then S of the result will be $[1, 0, 1]$, the binary representation of 5. If the single error is in bit 6, then S will give $[1, 1, 0]$, the binary representation for 6.

Example: Using Hamming code

Suppose we encode the message "HELP" by first converting the characters to ASCII code, written in binary,

$$\text{HELP} \to [1001000][1000101][1001100][1010000]$$

Then break this up into 4-bit words instead of the seven bits that the ASCII code uses:

$$[1001][0001][0001][0110][0110][0101][0000])$$

Next apply C to each of these words to get

[0011001][1101001][1101001][1100110][110110][0100101][0000000]

Now the point of an error correcting code is to correct errors, so lets add the following bit stream to get errors in transmission in the bits where the 1's occur:

$$\text{errors} = [0001000][0100000][0010000][1100000]$$

$$[0000001][0010000][0001000]$$

Notice that except for one cluster each of these error patterns only has one error in a transmitted 7-bit word. If we add this error, then the received message is:

[0010001][1001001][1111001][0000110][1100111][0110101][0001000]

Applying S tells us which bits to change to correct for the errors:

$$[100][010][011][011][111][011][100]$$

Notice that this identifies the single errors correctly, but does not correct the double error in the fourth transmitted word. Our corrected received words would then be

[0011001][1101001][1101001][0010110][1100110][0100101][0000000]

Extracting bits 3,5,6, and 7 of each of these words gives

$$[1001][0001][0001][1110][0110][0101][0000].$$

Regrouping into seven bit words gives

$$[1001000][1000111][1001100][1010000]$$

which can be converted back to the character string "HGLP". Notice that we have one character incorrectly transmitted as a result of the burst of errors in the transmission. ♢

Exercises 6.3:

1. Show that the code words for the Hamming (7,4) code all differ from each other in at least 3 bits.

2. The following bit streams were received by a communications device. It is known that the Hamming (7,4) code was used and that there was noise on the line. Correct the errors. Then give the decoded string.

   (a) 0011010001111101010111100110

   (b) 1010101010101010101010101010

   (c) 1000000110000011100001111000

3. Hamming also developed an (8,4) code by taking a 4 bit message and adjoining 4 more bits so that the sums of the 1,2, and 5 bits, the 3,4, and 6 bits, the 1,3, and 7 bits, and the 2,4, and 8 bits all give 0. Show that the code words for this code form a vector space over $\mathbb{Z}_2$. Can you see how to use this scheme to correct a single error in any bit?

6.4 Sums and Intersections of Subspaces

We have seen how to create subspaces of a vector space using the span of a set of vectors. Given subspaces, say $\mathcal{W}_1 \leq \mathcal{V}, \mathcal{W}_2 \leq \mathcal{V}$, can we produce new subspaces from these? In Set theory we used intersection and union to produce new subsets from old; will they work for subspaces as well? If, perhaps naively, we try these operations on subspaces, we find that one of them "works" and the other one does not. As an illustration of the successful case, let $\mathcal{V} = \mathbb{R}^3, \mathcal{W}_1 =$ the xy-plane and $\mathcal{W}_2 =$ the zx-plane. Then $\mathcal{W}_1 \cap \mathcal{W}_2$ is simply the x-axis, a line through the origin, hence a subspace (here $\mathcal{W}_1 \cap \mathcal{W}_2 = \text{Span}(\{[1,0,0]\})$).

Proposition 6.4.1 *Let $\mathcal{W}_1, \mathcal{W}_2 \leq \mathcal{V}$. Then $\mathcal{W}_1 \cap \mathcal{W}_2 \leq \mathcal{V}$.*

PROOF:

All we need to do is show that the two closure axioms hold for $\mathcal{W}_1 \cap \mathcal{W}_2$. If $\vec{a}, \vec{b} \in \mathcal{W}_1 \cap \mathcal{W}_2$, $c \in \mathbb{R}$, then by definition of intersection, $\vec{a}$ and $\vec{b} \in \mathcal{W}_1$ and $\vec{a}$ and $\vec{b} \in \mathcal{W}_2$; since $\mathcal{W}_1$ and $\mathcal{W}_2$ are subspaces,

$$\vec{a} + \vec{b} \in \mathcal{W}_1,$$

$$\vec{a} + \vec{b} \in \mathcal{W}_2,$$

and

$$c\vec{a} \in \mathcal{W}_1 \text{ and } c\vec{a} \in \mathcal{W}_2.$$

Thus,

$$\vec{a} + \vec{b} \in \mathcal{W}_1 \cap \mathcal{W}_2$$

and

$$c\vec{a} \in \mathcal{W}_1 \cap \mathcal{W}_2.$$

∎

We remark that $\mathcal{W}_1 \cap \mathcal{W}_2 \neq \emptyset$, since the vector $\vec{0}$ is in every subspace and thus must always be in $\mathcal{W}_1 \cap \mathcal{W}_2$. However, in some cases, it may be that $\mathcal{W}_1 \cap \mathcal{W}_2 = \{\vec{0}\}$. For example, if $\mathcal{V} = \mathbb{R}^2$, the only nontrivial subspaces are the lines through the origin, and the intersection of two distinct subspaces is precisely the origin. To indicate other possibilities, we note that the nontrivial subspaces of $\mathbb{R}^3$ are the lines through the origin and the planes through the origin. Clearly, the intersection of two distinct lines or of a plane and a line not in the plane will be just the origin; the intersection of two distinct planes, as indicated above, will be the line of intersection.

In order to see what might happen with the union of two subspaces, consider $\mathcal{V} = \mathbb{R}^2$, $\mathcal{W}_1 =$ x-axis, $\mathcal{W}_2 =$ y-axis. Then $\vec{e}_1 = [1, 0] \in \mathcal{W}_1$, $\vec{e}_2 = [0, 1] \in \mathcal{W}_2$, but $\vec{e}_1 + \vec{e}_2 = [1, 1] \notin \mathcal{W}_1 \cup \mathcal{W}_2$, this last being the set of vectors on either x-axis or y-axis. But this example of how things go wrong gives a suggestion for getting round the difficulty: we must, at the very least, include in our new set the sum of any two vectors in the given subspaces. As it turns out, that is all we need do.

Definition 6.4.1 *Let $\mathcal{W}_1$ and $\mathcal{W}_2$ be subspaces of $\mathcal{V}$, then the sum $\mathcal{W}_1 + \mathcal{W}_2$ is the set of all vectors which can be written as the sum of a vector in $\mathcal{W}_1$ and a vector in $\mathcal{W}_2$: $\mathcal{W}_1 + \mathcal{W}_2 = \{\vec{a} \in \mathcal{V} \mid \vec{a} = \vec{b}_1 + \vec{b}_2, \vec{b}_1 \in \mathcal{W}_1, \vec{b}_2 \in \mathcal{W}_2\}$.*

Example:

As an example consider $\mathcal{V} = \mathbb{R}^2$, $\mathcal{W}_1 = \{[a_1, 0] \mid a_1 \in \mathbb{R}\}$, $\mathcal{W}_2 = \{[0, a_2] \mid a_2 \in \mathbb{R}\}$, i.e., $\mathcal{W}_1$ is the x-axis and $\mathcal{W}_2$ the y-axis in the Euclidean plane. Then a vector in $\mathcal{W}_1 + \mathcal{W}_2$ is the sum of a vector in $\mathcal{W}_1$ and a vector in $\mathcal{W}_2$; thus $\vec{a} \in \mathcal{W}_1 + \mathcal{W}_2$ is of the form $\vec{a} = [a_1, 0] + [0, a_2] = [a_1, a_2]$. Clearly $\mathcal{W}_1 + \mathcal{W}_2 = \mathcal{V} = \mathbb{R}^2$. ♢

Example:

Similarly, it is easy to see that if $\mathcal{V} = \mathbb{R}^3$, $\mathcal{W}_1 = \{[a_1, a_2, 0] \mid a_1, a_2 \in \mathbb{R}\}$, $\mathcal{W}_2 = \{[0, 0, a_3] \mid a_3 \in \mathbb{R}\}$, then $\mathcal{W}_1 + \mathcal{W}_2 = \mathbb{R}^3 = \mathcal{V}$. $\diamondsuit$

These examples show that, just as $\mathcal{W}_1 \cap \mathcal{W}_2$ may turn out to be the trivial subspace $\{\vec{0}\}$, so might it happen that $\mathcal{W}_1 + \mathcal{W}_2$ is the trivial subspace $\mathcal{V}$. We now have a way of creating a new ("bigger") subspace from two subspaces.

Proposition 6.4.2 *If $\mathcal{W}_1, \mathcal{W}_2 \leq \mathcal{V}$, then $\mathcal{W}_1 + \mathcal{W}_2 \leq \mathcal{V}$ (the sum of two subspaces is again a subspace).*

PROOF:

We must verify that $\mathcal{W}_1 + \mathcal{W}_2$ is closed with respect to addition and scalar multiplication. Suppose $\vec{a}$ and $\vec{b} \in \mathcal{W}_1 + \mathcal{W}_2$, $c \in \mathbb{R}$. Then

$$\begin{aligned}
\vec{a} &= \vec{a}_1 + \vec{a}_2 \text{ with } \vec{a}_1 \in \mathcal{W}_1 \text{ and } \vec{a}_2 \in \mathcal{W}_2 \\
\vec{b} &= \vec{b}_1 + \vec{b}_2 \text{ with } \vec{b}_1 \in \mathcal{W}_1 \text{ and } \vec{b}_2 \in \mathcal{W}_2, \\
\vec{a} + \vec{b} &= (\vec{a}_1 + \vec{a}_2) + (\vec{b}_1 + \vec{b}_2) \\
&= (\vec{a}_1 + \vec{b}_1) + (\vec{a}_2 + \vec{b}_2) \in \mathcal{W}_1 + \mathcal{W}_2 \\
c\vec{a} &= c(\vec{a}_1 + \vec{a}_2) \\
&= c\vec{a}_1 + c\vec{a}_2 \in \mathcal{W}_1 + \mathcal{W}_2,
\end{aligned}$$

since $\mathcal{W}_1$ and $\mathcal{W}_2$ are subspaces and thus satisfy the closure axioms. ∎

Notice, by the way, that we always have, for $\mathcal{W}_1$ and $\mathcal{W}_2$ subspaces of $\mathcal{V}$,

$$\begin{aligned}
\mathcal{W}_1 \cap \mathcal{W}_2 &\leq \mathcal{W}_1 \\
\mathcal{W}_1 \cap \mathcal{W}_2 &\leq \mathcal{W}_2 \\
\mathcal{W}_1 &\leq \mathcal{W}_1 + \mathcal{W}_2 \\
\mathcal{W}_2 &\leq \mathcal{W}_1 + \mathcal{W}_2
\end{aligned}$$

It is also true–and very easy to show–that $\mathcal{W}_1 \cap \mathcal{W}_2$ is the largest subspace common to both $\mathcal{W}_1$ and $\mathcal{W}_2$ and that $\mathcal{W}_1 + \mathcal{W}_2$ is the smallest subspace which contains both $\mathcal{W}_1$ and $\mathcal{W}_2$.

When two subspaces have the smallest possible intersection and have sum giving all of $\mathcal{V}$ we have a particularly nice situation:

Definition 6.4.2 *Let $\mathcal{V}$ be a vector space and let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of $\mathcal{V}$ which satisfy*

$$\mathcal{V} = \mathcal{U} + \mathcal{W} \text{ and } \mathcal{U} \cap \mathcal{W} = \{\vec{0}\},$$

then $\mathcal{V}$ is said to be the **direct sum** *of $\mathcal{U}$ and $\mathcal{W}$; this relation is indicated by writing*

$$\mathcal{V} = \mathcal{U} \oplus \mathcal{W}.$$

Example:

> In $\mathcal{V} = \mathbb{R}^3$ if we take $\mathcal{U}$ =the xy-plane and $\mathcal{W}$ = any line through the origin not contained in the xy-plane, then $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$.
> $\diamondsuit$

Decomposing spaces as direct sums of subspaces on which a linear transformation has a nice form is an important part of advanced linear algebra.

Exercises 6.4:

For problems 1-10 given the subspaces $\mathcal{U}$ and $\mathcal{W}$ of $\mathcal{V}$,describe the subspaces $\mathcal{U} + \mathcal{W}$ and $\mathcal{U} \cap \mathcal{W}$:

1. $\mathcal{V} = \mathbb{R}^3, \mathcal{U} = \{[a_1, 0, 0] \mid a_1 \in \mathbb{R}\}$ and $\mathcal{W} = \{[0, a_2, 0] \mid a_2\mathbb{R}\}$
2. $\mathcal{V} = \mathbb{R}^3, \mathcal{U} = \{[a_1, a_2, 0] \mid a_1$ and $a_2 \in \mathbb{R}\}$ and $\mathcal{W} = \{[0, 0, a_3] \mid a_3 \in \mathbf{R}\}$
3. $\mathcal{V} = \mathbb{R}^3, \mathcal{U} = \{[a_1, a_2, 0 \mid a_1$ and $a_2 \in \mathbb{R}\}$ and $\mathcal{W} = \{[0, a_3, a_4] \mid a_3$ and $a_4 \in \mathbb{R}\}$
4. $\mathcal{V} = \mathbb{R}^3, \mathcal{U} = \{[a_1, a_2, a_3] \mid a_1 + a_2 + a_3 = 0\}$ and $\mathcal{W} = \{[0, 0, a_4] \mid a_4 \in \mathbb{R}\}$
5. $\mathcal{V} = \mathbb{R}^3, \mathcal{U} = \{[a_1, a_2, a_3] \mid a_1 + a_2 + a_3 = 0\}$ and $\mathcal{W} = \{[a_4, a_5, a_6] \mid 2a_4 - a_5 + a_6 = 0\}$
6. $\mathcal{V} = \mathbb{R}[x]$, $\mathcal{U} = \mathbb{R}[x]_3$ and $\mathcal{W} = \{$ polynomials involving only even powers$\}$.
7. $\mathcal{V} = \mathbb{R}[x], \mathcal{U} = \{p | p(-x) = p(x)\}$ and $\mathcal{W} = \{p | p(-x) = -p(x)\}$

8. $\mathcal{V} = \mathbb{R}[x]$, $\mathcal{U} = \{p|p(1) = 0\}$ and $\mathcal{W} = \{p|p'(1) = 0\}$

9. $\mathcal{V} = \mathbb{R}^{\mathbb{N}}$, $\mathcal{U} = \{a|a_{2n+1} = -a_{2n}\}$ and $\mathcal{W} = \{b|b_n = b_0 2^n\}$

10. $\mathcal{V} = \mathbb{R}^{\mathbb{N}}$, $\mathcal{U} = \{a|a$ converges to $0\}$ and $\mathcal{W} = \{a|\sum_{k=0}^{\infty} a_k$ converges$\}$

11. Let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of $\mathcal{V}$. Is it ever true that $\mathcal{U} \cup \mathcal{W}$ is a subspace of $\mathcal{V}$?

12. Prove that if a subspace $\mathcal{U}$ contains both subspaces $\mathcal{W}_1$ and $\mathcal{W}_2$ then it must contain $\mathcal{W}_1 + \mathcal{W}_2$.

Chapter 7

Systems of Linear Equations

In Chapter 6 and in later chapters circumstances come up in which we needed to find the solution (or the general form for all solutions) of a system of linear equations. For small systems, say three equations in three unknowns, most students can muddle through with *ad hoc* methods. Many applications and much of our work with spaces more complicated than $\mathbb{R}^3$ will require a more efficient and systematic approach. In this chapter we will work with Gaussian elimination with backsolving, a computationally efficient algorithm which has the advantage of having clear goals at each step. This algorithm will arise later in the course for other uses. It is easy to program on a computer, though somewhat sensitive to round off error when floating point arithmetic is used. The algorithms in this chapter can be carried out using arithmetic in any field we choose. To avoid problems with round off error we will usually work over $\mathbb{Q}$, though $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}_2$ also work the same way.

7.1 Gaussian Elimination with Backsolving

Let us start by considering a system of three equations in three unknowns. As an example we will use :

$$\begin{aligned} x + 3y + z &= 3 \\ -2x - 4y + 2z &= 12 \\ 3x + 8y + 2z &= 5. \end{aligned}$$

It is unlikely that the solution to this system will be obtained easily by inspection. We will use the first equation to eliminate x from the other two

equations. This could be done by substituting $x = 3 - 3y - z$ into the other equations but that would make work in simplification. Instead we will use the fact that both sides of an equality can be multiplied by the same number without changing the validity of the equality and the fact that "equals added to equals yield equals" to eliminate x from the later equations. We start by adding 2 times the first equation to the second:

$$\begin{array}{rrcl} 2(x+3y+z=3) \text{ is} & 2x+6y+2z & = & 6 \\ \text{add to} & -2x-4y+2z & = & 12 \\ \text{to get} & 2y+4z & = & 18 \end{array}$$

which then replaces the second equation. We can then rewrite the system as:

$$\begin{array}{rcl} x+3y+z & = & 3 \\ 2y+4z & = & 18 \\ 3x+8y+2z & = & 5. \end{array}$$

If we now add -3 times the first equation to the third we can eliminate x from that equation as well without changing the triples (x, y, z) which satisfy the system. This gives

$$\begin{array}{rcl} x+3y+z & = & 3 \\ 2y+4z & = & 18 \\ -y-z & = & -4. \end{array}$$

We can multiply both sides of the second equation by $\frac{1}{2}$ to get a more convenient form for using the second to eliminate y:

$$\begin{array}{rcl} x+3y+z & = & 3 \\ y+2z & = & 9 \\ -y-z & = & -4. \end{array}$$

Adding the second equation to the third gives:

$$\begin{array}{rcl} x+3y+z & = & 3 \\ y+2z & = & 9 \\ z & = & 5. \end{array}$$

We can now substitute 5 in for z in the second equation and solve for y to get $y = -1$, and then use the values for both y and z to get $x = 1$ from the

first equation. These operations can be accomplished by adding -2 times the third row to the second and then adding -1 times the third row and -3 times the (new) second row to the first.

As described this is a cumbersome, though organized, procedure. It is made less cumbersome by using matrices to let position of numbers keep track of the variables and the equal sign. A matrix is a rectangular array of numbers. For the time being we are using matrices much as they are used for storage in computer science. The system of equations

$$\begin{aligned} x + 3y + z &= 3 \\ -2x - 4y + 2z &= 12 \\ 3x + 8y + 2z &= 5 \end{aligned}$$

can be recaptured from the matrix of coefficients

$$\begin{bmatrix} 1 & 3 & 1 \\ -2 & -4 & 2 \\ 3 & 8 & 2 \end{bmatrix}$$

and the matrix of constants

$$\begin{bmatrix} 3 \\ 12 \\ 5 \end{bmatrix}.$$

It is most convenient to write this as one matrix

$$\begin{bmatrix} 1 & 3 & 1 & 3 \\ -2 & -4 & 3 & 12 \\ 3 & 8 & 2 & 5 \end{bmatrix}$$

called the augmented matrix corresponding to the system. This may be thought of as the system with all of the variables and the equal signs erased.

The operations used on the system of equations will be replaced with operations on a matrix called *elementary row operations*. There are three types of elementary row operations. Since we will be using them repeatedly we will give a shorthand notation for specifying what you do to get from one matrix to the next.

Interchange rows i and j:

$$\begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{j1} & \dots & a_{jk} & \dots & a_{jn} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix} \overset{R_i \leftrightarrow R_j}{\leadsto} \begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{j1} & \dots & a_{jk} & \dots & a_{jn} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix}$$

Multiply row i by r (where $r \neq 0$)

$$\begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix} \overset{rR_i}{\leadsto} \begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ ra_{i1} & \dots & ra_{ik} & \dots & ra_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix}$$

Add r times row i to row j (row i is unchanged)

$$\begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{j1} & \dots & a_{jk} & \dots & a_{jn} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix} \overset{R_j + rR_i}{\leadsto} \begin{bmatrix} a_{11} & \dots & a_{1k} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ik} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{j1} + ra_{i1} & \dots & a_{jk} + ra_{ik} & \dots & a_{jn} + ra_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mk} & \dots & a_{mn} \end{bmatrix}$$

The strategy used is as follows: first, switch rows if necessary to get a non-zero entry in the first row, first column position. Next divide the first row by the 1,1-entry to make that entry 1. Then use the operation of adding a multiple of the first row to the later rows to get zeros in the first column

below the first entry. In our example this goes as follows:

$$\begin{bmatrix} 1 & 3 & 1 & 3 \\ -2 & -4 & 2 & 12 \\ 3 & 8 & 2 & 5 \end{bmatrix} \quad \underset{\rightsquigarrow}{R_2 + 2R_1} \quad \begin{bmatrix} 1 & 3 & 1 & 3 \\ 0 & 2 & 4 & 18 \\ 3 & 8 & 2 & 5 \end{bmatrix}$$

$$\underset{\rightsquigarrow}{R_3 - 3R_1} \quad \begin{bmatrix} 1 & 3 & 1 & 3 \\ 0 & 2 & 4 & 18 \\ 0 & -1 & -1 & 14 \end{bmatrix}.$$

Next we repeat this procedure using the second row to eliminate all entries in the second column below it. This continues until the matrix of coefficients has 1 as the first nonzero entry in each row a 1 and all entries below that 1 in the same column are 0. In our example we continue as follows:

$$\begin{bmatrix} 1 & 3 & 1 & 3 \\ 0 & 2 & 4 & 18 \\ 0 & -1 & -1 & -4 \end{bmatrix} \quad \underset{\rightsquigarrow}{\tfrac{1}{2}R_2} \quad \begin{bmatrix} 1 & 3 & 1 & 3 \\ 0 & 1 & 2 & 9 \\ 0 & -1 & -1 & -4 \end{bmatrix}$$

$$\underset{\rightsquigarrow}{R_3 + R_2} \quad \begin{bmatrix} 1 & 3 & 1 & 3 \\ 0 & 1 & 2 & 9 \\ 0 & 0 & 1 & 5 \end{bmatrix}$$

Next we get zeros above the ones using the same operations and working our way back from right to left. This means that at each step we are working with a row with the minimum possible number of nonzero entries, thus minimizing the arithmetic.

$$\begin{bmatrix} 1 & 3 & 1 & 3 \\ 0 & 1 & 2 & 9 \\ 0 & 0 & 1 & 5 \end{bmatrix} \quad \underset{\rightsquigarrow}{\begin{matrix} R_1 - R_3 \\ R_2 - 2R_3 \end{matrix}} \quad \begin{bmatrix} 1 & 3 & 0 & -2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 5 \end{bmatrix}$$

$$\underset{\rightsquigarrow}{R_1 - 3R_2} \quad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 5 \end{bmatrix}$$

Putting the variables and the equal sign back in we have

$$\begin{aligned} x &= 1 \\ y &= -1 \\ z &= 5. \end{aligned}$$

Example: Solving a system

Find the solutions to

$$\begin{aligned} x+2y+z &= 2 \\ 2x+6y-z &= -15 \\ -x+2y+4z &= 15 \end{aligned}$$

The augmented matrix is

$$\begin{bmatrix} 1 & 2 & 1 & 2 \\ 2 & 6 & -1 & -15 \\ -1 & 2 & 4 & 15 \end{bmatrix}$$

Our algorithm gives

$$\begin{bmatrix} 1 & 2 & 1 & 2 \\ 2 & 6 & -1 & -15 \\ -1 & 2 & 4 & 15 \end{bmatrix} \begin{matrix} R_2-2R_1 \\ R_3+R_1 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & 2 & -3 & -19 \\ 0 & 4 & 5 & 17 \end{bmatrix}$$

$$\begin{matrix} \frac{1}{2}R_2 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & 1 & -\frac{3}{2} & -\frac{19}{2} \\ 0 & 4 & 5 & 17 \end{bmatrix} \begin{matrix} R_3-4R_2 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & 1 & -\frac{3}{2} & -\frac{19}{2} \\ 0 & 0 & 11 & 55 \end{bmatrix}$$

$$\begin{matrix} \frac{1}{11}R_3 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & 1 & -\frac{3}{2} & -\frac{19}{2} \\ 0 & 0 & 1 & 5 \end{bmatrix} \begin{matrix} R_1-R_3 \\ R_2+\frac{3}{2}R_3 \\ \rightsquigarrow \end{matrix}$$

$$\begin{bmatrix} 1 & 2 & 0 & -3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 5 \end{bmatrix} \begin{matrix} R_1-2R_2 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 5 \end{bmatrix}$$

While this example isn't too bad, it does show how fractions can creep into and back out of the work. $\diamondsuit$

Example: A system over $\mathbb{Z}_2$

Solve the system

$$\begin{aligned} x_1+x_3+x_4 &= 1 \\ x_1+x_2+x_3+x_4 &= 0 \\ x_2+x_3+x_4 &= 1 \\ x_1+x_2+x_3 &= 0 \end{aligned}$$

Applying the Gaussian elimination algorithm using arithmetic in $\mathbb{Z}_2$ gives

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix} \quad \begin{matrix} R_2 + R_1 \\ R_4 + R_1 \\ \rightsquigarrow \end{matrix} \quad \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$\begin{matrix} R_3 + R_2 \\ R_4 + R_2 \\ \rightsquigarrow \end{matrix} \quad \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\begin{matrix} R_1 + R_4 \\ R_3 + R_4 \\ \rightsquigarrow \end{matrix} \quad \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\begin{matrix} R_1 + R_3 \\ \rightsquigarrow \end{matrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Thus

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 1 \\ x_3 &= 0 \\ x_4 &= 0 \end{aligned}$$

$\diamondsuit$

So far we have pretended that all systems have solutions. What happens to the algorithm if we try it on one which does not?

Example: A system with no solutions

Find the solutions (if any) of

$$\begin{aligned} x + y &= 2 \\ x + y &= 3 \end{aligned}$$

These are two parallel lines, so we know that there are no solutions. Let us follow our algorithm until we reach a point where this lack of solutions becomes apparent.

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 3 \end{bmatrix} \begin{array}{c} R_2 - R_1 \\ \leadsto \end{array} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

at which point we are stuck. If we put the variables and the equal sign back in we get

$$\begin{aligned} x + y &= 2 \\ 0x + 0y &= 1. \end{aligned}$$

The second equation tells us something we know to be false, so whatever values we put in for x and y we cannot get simultaneous solutions. Anytime we get a row with all zero entries except for the last column and with a nonzero entry there we know that the system of equations has no solutions. Such systems are called **inconsistent**. $\diamondsuit$

Example: A system with many solutions

We can also get systems which have many solutions instead of one. As an example let us consider what our algorithm does for the system

$$\begin{aligned} x + y + z &= 3 \\ x + 2y - z &= 4 \\ 2x + 3y &= 7 \end{aligned}$$

Setting up the matrix and applying the algorithm gives:

$$\begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & 2 & -1 & 4 \\ 2 & 3 & 0 & 7 \end{bmatrix} \begin{array}{c} R_2 - R_1 \\ R_3 - 2R_1 \\ \leadsto \end{array} \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & -2 & 1 \\ 0 & 1 & -2 & 1 \end{bmatrix}$$

$$\begin{array}{c} R_1 - R_2 \\ R_3 - R_2 \\ \leadsto \end{array} \begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

At this point we can put the variables and equal signs back in to get

$$\begin{aligned} x + 3z &= 2 \\ y - 2z &= 1 \end{aligned}$$

The variable z can assume any value we wish, but once it is chosen x and y are determined. Since one of the equations gives us no new information, such systems are called **redundant**. $\diamondsuit$

Given any system of equations, even one with different numbers of equations and unknowns, we can apply elementary row operations according to our algorithm to reduce the matrix of coefficients to what is called row-reduced echelon form:

Definition 7.1.1 *A matrix is in* **row-reduced echelon form** *if and only if*

1. *The first nonzero entry in each row is a 1*
2. *The first nonzero entry in a row appears to the right of the first nonzero entry in the row above it*
3. *All other entries in the column of that first nonzero entry in the row are 0.*
4. *All rows with only 0 entries are at the bottom.*

Example: The following matrices are in row-reduced echelon form:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 56 \\ 0 & 0 & 1 & 2 & \frac{3}{4} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 56 \\ 0 & 0 & 1 & 2 & \frac{3}{4} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 4 \\ 0 & 1 & 2 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & -1 & \pi \end{bmatrix}$$

$\diamondsuit$

Example: The following matrices are not in row-reduced echelon form

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 4 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 56 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & \frac{3}{4} \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$\diamondsuit$

The approach used in solving systems of equations using elementary row operations also works for reducing a matrix to row reduced echelon form. Our strategy is to use interchange and division of rows to get the first nonzero entry in a row to be a 1 and to make that entry as far to the left as possible. We then use addition of a multiple of that row to rows below it to get zeros

below it in the column. Then move on to the next row and repeat the process, being careful not to mess up previous work by interchanging with rows above the current one. Once we have all of the first nonzero entries equal to 1 and all the entries below them 0, we work backwards from right to left to get the zeros above the first nonzero entries in each row. Often we can avoid use of fractions by judicious use of row exchange instead of dividing to get the 1 as the first entry in a row.

To solve any system of equations: form the augmented matrix, reduce it to echelon form, and put the variables and equal signs back in. The result will be a description of the solutions. As we have described it this process is called Gaussian elimination with backsolving. A formal description in pseudocode is given in Figure 7.1. It is the best algorithm for systems not known to have nice structure (like lots of zero entries for instance). There are other algorithms which are better for very large systems with many coefficients zero. Such systems do occur, for instance, in modeling secondary oil recovery: many systems of 50,000 equations with 50,000 unknowns must be solved. The work involved is considerable.

While use of elementary row operations makes sense and is not difficult, it does get tedious for large problems. We can repackage the operations into a macro called **pivot** to combine several steps into one action. In obtaining the row reduced echelon form we look at the first non-zero entry in a row, divide that row by that entry, and then use addition of multiples of that row to other rows to get 0's in the rest of the column.

Definition 7.1.2 *If* $\mathbf{M}$ *is a matrix with* $m_{ij} \neq 0$ *then the pivot on the* ij *position of* $\mathbf{M}$ *is the sequence of row operations* $\frac{1}{m_{ij}} R_i$ *then* $R_k - m_{kj} R_i$ *for all* $k \neq i$*. It results in a new matrix with a 1 in the* ij *position and 0 in the rest of the* j^{th} *column.*

Example: A pivot operation

If

$$\mathbf{M} = \begin{bmatrix} 2 & 1 & 3 & 4 \\ 2 & 4 & 2 & 6 \\ 1 & 2 & 4 & 3 \\ 1 & 1 & -1 & 2 \end{bmatrix}$$

Suppose that A is the matrix obtained by augmenting the coefficient matrix with the column of constants.

```
Clearing below the diagonal
Set R=1
Set C=1
While (R < Number of Rows) and (C ≤ Number of Columns of coefficient matrix) Do
    If A(R,C)≠ 0
        Then Do
            Multiply row R by 1/A(R,C) (to get 1 in leading position of row R)
            For i= R+1 to Number of Rows (to clear entries below the (R,C) position)
                Add -A(i,C) times row R to row i
            Next i
            Increment R by adding 1
            Increment C by adding 1

        Else Do (search for next row and column to use)
            Set NewRow=R+1
            Do until (NewRow > Number of Rows) or (A(NewRow,C)≠ 0)
                Increment NewRow by adding 1
            If (NewRow > Number of Rows)
                Then Increment C by adding 1
                Else Swap rows R and NewRow
            End If
        End Else
    End If
End While

Backsolving
If (R = Number of Rows) then Multiply row R by 1/A(R,C)
While (R > 0) Do
    Set C = 1
    While (A(i,C) = 0) increment C (Find first nonzero entry in row R)
    For i=1 to R-1 (clear entries above that nonzero entry)
        Add -A(i,C) times row R to row i
    Next i
    Decrement R by subtracting 1
End While
```

Figure 7.1: Gaussian Elimination with Backsolving

then performing a pivot on the 2,3 position leads to the matrix

$$\begin{bmatrix} -1 & -5 & 0 & -5 \\ 1 & 2 & 1 & 3 \\ -3 & -6 & 0 & -21 \\ 2 & 3 & 0 & 8 \end{bmatrix}.$$

It is the same as the sequence of elementary row operations

$$\begin{bmatrix} 2 & 1 & 3 & 4 \\ 2 & 4 & 2 & 6 \\ 1 & 2 & 4 & 3 \\ 1 & 1 & -1 & 2 \end{bmatrix} \quad \begin{matrix} \frac{1}{2}R_2 \\ \rightsquigarrow \end{matrix} \quad \begin{bmatrix} 2 & 1 & 3 & 4 \\ 1 & 2 & 1 & 3 \\ 1 & 2 & 4 & 3 \\ 1 & 1 & -1 & 2 \end{bmatrix}$$

$$\begin{matrix} R_1 - 3R_2 \\ R_3 - 4R_2 \\ R_4 + R_2 \\ \rightsquigarrow \end{matrix} \quad \begin{bmatrix} -1 & -5 & 0 & -5 \\ 1 & 2 & 1 & 3 \\ -3 & -6 & 0 & -21 \\ 2 & 3 & 0 & 8 \end{bmatrix}$$

$\diamondsuit$

For solving systems of equations we have been using the Gaussian elimination with backsolving algorithm. A closely related algorithm, the Gauss-Jordan elimination algorithm, makes all of the entries above *and* below the diagonal zero. This algorithm has only one phase, and it takes a little more arithmetic, but has a more compact description if we use the pivot operation. We give the Gauss-Jordan algorithm for row reduction in a more formal pseudocode in Figure 7.2

Example: Gauss-Jordan Row Reduction

Let us use this algorithm to reduce the matrix

$$\begin{bmatrix} 2 & 4 & 6 & 8 & 10 \\ 2 & 3 & 4 & 1 & 1 \\ 4 & 6 & 8 & 5 & 5 \end{bmatrix}$$

to row reduced echelon form. We will indicate a pivot on position i,j by $P(i, j)$.

$$\begin{bmatrix} 2 & 4 & 6 & 8 & 10 \\ 2 & 3 & 4 & 1 & 1 \\ 4 & 6 & 8 & 5 & 5 \end{bmatrix} \quad \begin{matrix} P(1,1) \\ \rightsquigarrow \end{matrix} \quad \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & -1 & -2 & -7 & -9 \\ 0 & -2 & -4 & -11 & -15 \end{bmatrix}$$

```
Set Row=1
Set Column=1
While (Row ≤ Number of Rows) and (Column ≤ Number of Columns) Do
    If A(Row,Column)≠ 0
        Then do
            Pivot on position (Row,Column) in A
            Increment Row by adding 1
            Increment Column by adding 1
        Else do
            Set NewRow=Row+1
            Do until (NewRow > Number of Rows) or (A(NewRow,Column)≠ 0)
                Increment NewRow by adding 1
            If (NewRow > Number of Rows)
                Then Increment Column by adding 1
                Else Swap rows Row and NewRow
            End If
        End Else
    End If
End While
```

Figure 7.2: Gauss-Jordan Algorithm for Row Reduction Using Pivot

$$\begin{array}{c} P(2,2) \\ \rightsquigarrow \end{array} \quad \begin{bmatrix} 1 & 0 & -1 & -10 & -13 \\ 0 & 1 & 2 & 7 & 9 \\ 0 & 0 & 0 & 3 & 3 \end{bmatrix}$$
$$\begin{array}{c} P(3,4) \\ \rightsquigarrow \end{array} \quad \begin{bmatrix} 1 & 0 & -1 & 0 & -3 \\ 0 & 1 & 2 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$\diamondsuit$

7.1.1 Operation Counts

One way to measure the efficiency of the algorithm for solving a system of equations is to count the number of multiplications which it needs to solve a system. When we do this we assume that no special structure shows up in the process of the algorithm and that no unnecessary operations are performed. Let us count how many multiplications are needed to solve a system of n equations in n unknowns using Gaussian elimination with backsolving.

Getting the 1 on the diagonal in the i^{th} row requires a multiplication in each position after the i^{th}, so $n+1-i$ multiplications.

Once we have the 1 on the diagonal we will, in general, need to use the row operation adding a multiple of the i^{th} row to each of the $n-i$ rows below it to make the entries below the diagonal 0. There are $n+1-i$ entries in each row whose value we need to calculate (we don't worry about the zeros at the beginning or about the i^{th} column since we know what those values will be without calculating them). This makes $(n-i)(n+1-i)$ multiplications to get the zeros below the diagonal in the i^{th} column.

Backsolving requires one multiplication for each row above the diagonal entry being used. So it involves $n-i+1$ multiplications for each of the columns $i=2$ to $i=n$.

For the whole algorithm we need

$$\sum_{i=1}^{n} n+1-i = \frac{n(n+1)}{2}$$

multiplications to get all of the 1's on the diagonals,

$$\sum_{i=1}^{n-1}(n-i)(n+1-i) = \sum_{k=1}^{n-1} k^2 + k = \frac{(n-1)n(2n-2)}{6} + \frac{(n-1)n}{2}$$

multiplications to get the zeros below the diagonal, and

$$\sum_{i=2}^{n}(n-i+1) = \sum_{i=1}^{n-1} i = \frac{(n-1)n}{2}$$

multiplications for the backsolving. All together this gives (after simplifying)

$$\frac{n^3}{3} + n^2 - \frac{n}{3}$$

multiplications for the Gaussian elimination with backsolving algorithm.

An analysis like the one above gives

$$\frac{n^3}{2} + \frac{n^2}{2}$$

multiplications for the solution of a system of equations using the Gauss-Jordan algorithm.

Some students may be familiar with solution of equations using determinants (Cramer's rule). The reason we do not study this method in this course is made clear by the following table showing the number of multiplications needed to solve an n by n system of equations:

n	Gauss	Gauss-Jordan	Cramer
2	6	6	6
3	12	18	36
4	36	40	200
5	65	75	1230

Exercises 7.1:

For problems 1–10 solve the following systems of linear equations. If the system is inconsistent, say so. If the system has multiple solutions, describe the general form for a solution.

1.

$$\begin{aligned} x_1 + x_2 - x_3 &= 3 \\ 2x_1 - x_2 + 3x_3 &= 0 \\ -x_1 - 2x_2 + x_3 &= -5 \end{aligned}$$

2.

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_1 - x_2 + x_3 &= 3 \\ x_1 + 2x_2 + x_3 &= 7 \end{aligned}$$

3.

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 2 \\ 3x_1 + 2x_2 - x_3 &= 1 \\ 4x_2 + x_3 &= 5 \end{aligned}$$

4.

$$\begin{aligned} x_1 + x_2 + x_3 &= 6 \\ x_2 + x_3 + x_4 &= 2 \\ x_3 + x_4 + x_5 &= 1 \\ x_1 - x_5 &= 2 \\ 2x_2 + 2x_4 &= 4 \end{aligned}$$

5.

$$\begin{aligned} -x_1 + x_2 + 3x_3 + 2x_4 &= -7 \\ -6x_1 + 3x_2 + 8x_3 + 2x_4 &= -24 \\ -2x_1 + x_2 + 3x_3 + x_4 &= -9 \end{aligned}$$

6.

$$\begin{aligned} 2x_1 + x_2 + 5x_3 &= 1 \\ x_1 + 2x_2 + 5x_3 &= 0 \\ x_1 + x_2 + 5x_3 &= 1 \\ -x_1 + 2x_2 + 2x_3 &= -1 \end{aligned}$$

7.

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 6 \\ 2x_1 + x_2 &= 9 \\ 3x_1 + x_2 + x_3 &= 8 \end{aligned}$$

8.

$$\begin{aligned} 2x_1 + 8x_2 - x_3 &= 6 \\ -4x_1 - 8x_2 + 3x_3 &= -5 \\ 4x_1 + 4x_2 - 2x_3 &= 3 \end{aligned}$$

9.

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 10 \\ 3x_1 - x_2 + 2x_3 &= -9 \\ x_1 - x_2 + x_3 &= -9 \end{aligned}$$

10.

$$\begin{aligned} 5x_1 + 2x_2 + x_3 &= -4 \\ 2x_1 + x_2 + x_3 &= -5 \\ 4x_1 + 2x_2 + x_3 &= -7 \end{aligned}$$

For problems 11–16, how do you interpret these final results of the Gaussian elimination algorithm:

11. $\begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

12. $\begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

13. $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 0 & 4 \end{bmatrix}$

14. $\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 6 \end{bmatrix}$

15. $\begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 6 \end{bmatrix}$

16. $\begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

For problems 17–20, reduce to row-reduced echelon form:

17. $\begin{bmatrix} 0 & 1 & 1 \\ 2 & 4 & 6 \\ 3 & 1 & 0 \end{bmatrix}$

18. $\begin{bmatrix} 1 & 0 & 3 & 4 \\ 2 & 4 & 1 & 3 \\ 0 & 1 & 1 & 1 \end{bmatrix}$

19. $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

20. $\begin{bmatrix} 1 & 3 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 & 4 \end{bmatrix}$

21. Find the solution in $\mathbb{C}$ if any exists:

$$\begin{aligned} x_1 + (3+i)x_2 + x_3 &= i \\ ix_2 + 2x_2 - (1-2i)x_3 &= 1 \\ X_1 + x_2 + 3x_3 &= (2-i) \end{aligned}$$

22. Find the solution in $\mathbb{Z}_2$ if any exists:

$$\begin{aligned} x_1 + x_2 + x_4 + x_5 &= 1 \\ x_1 + x_3 + x_5 &= 0 \\ x_2 + x_3 + x_4 &= 0 \\ x_1 + x_3 + x_4 &= 1 \\ x_3 + x_4 + x_5 &= 0 \end{aligned}$$

7.2 Spanning Sets Revisited

At this point it becomes much easier to answer some of the questions raised in Chapter 6. One such question is whether or not a specific vector is an element of the subspace spanned by a set of vectors.

Example: Is $[1,1,1]$ in the subspace spanned by $[1,2,3]$ and $[3,2,1]$?

To answer this we ask if there are numbers x and y with $[1,1,1]=x[1,2,3]+y[3,2,1]$. This gives a system of three equations in two unknowns:

$$\begin{aligned} x+3y &= 1 \\ 2x+2y &= 1 \\ 3x+y &= 1 \end{aligned}$$

Solving this we find

$$\begin{bmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \\ 3 & 1 & 1 \end{bmatrix} \begin{matrix} R_2-2R_1 \\ R_3-3R_1 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 3 & 1 \\ 0 & -4 & -1 \\ 0 & -8 & -2 \end{bmatrix} \begin{matrix} \frac{1}{4}R_2 \\ \rightsquigarrow \end{matrix}$$

$$\begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & \frac{1}{4} \\ 0 & -8 & -2 \end{bmatrix} \begin{matrix} R_3+8R_2 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{matrix} R_1-3R_2 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 0 & \frac{1}{4} \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 \end{bmatrix}$$

This shows $[1,1,1]=\frac{1}{4}[1,2,3]+\frac{1}{4}[3,2,1]$ so $[1,1,1]\in \mathrm{Span}(\{[1,2,3],[3,2,1]\})$. $\diamondsuit$

Example: Is $[1,1,1]$ in the subspace spanned by $[0,1,2]$ and $[2,1,1]$?

Again, we ask for x and y with $x[0,1,2]+y[2,1,1]=[1,1,1]$ or

$$\begin{aligned} 0x+2y &= 1 \\ x+y &= 1 \\ 2x+y &= 1 \end{aligned}$$

While it is clear by inspection that this system has no solutions we can use our algorithm to get

$$\begin{bmatrix} 0 & 2 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} \overset{R_1 \leftrightarrow R_2}{\rightsquigarrow} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$$

$$\overset{R_3 - 2R_1}{\rightsquigarrow} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & -1 & -1 \end{bmatrix} \overset{\frac{1}{2}R_2}{\rightsquigarrow}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & -1 & -1 \end{bmatrix} \overset{R_3 + R_2}{\rightsquigarrow} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & -\frac{1}{2} \end{bmatrix}$$

At this point we can tell that the system has no solutions. Thus $[1, 1, 1]$ is not in the subspace spanned by $[0, 1, 2]$ and $[2, 1, 1]$. $\diamondsuit$

Homogeneous systems (those where all of the constant terms are 0) have a particularly nice theory.

Proposition 7.2.1 *The solutions to a homogeneous system of equations in n unknowns form a subspace of $\mathbb{R}^n$.*

PROOF:

The solutions to a single homogeneous equation

$$a_1x_1 + a_2x_2 + ... + a_nx_n = 0$$

can be thought of as elements of the kernel of the linear transformation taking $[x_1, ..., x_n]$ to $a_1x_1 + ... + a_nx_n$. Thus the set of all solutions to a single homogeneous equation forms a subspace of $\mathbb{R}^n$. The set of solutions to a system of homogeneous equations is the intersection of the sets of solutions of the equations taken individually. Thus the set of solutions to a system of homogeneous linear equations is the intersection of subspaces, and thus is itself a subspace. ∎

Since the zero vector will always give a solution to a homogeneous system of equations we will be interested in knowing when there are other, nontrivial, solutions.

Theorem 7.2.2 *A homogeneous system of linear equations with more unknowns than equations always has a nontrivial solution.*

PROOF:

When we apply Gaussian elimination with backsolving to the augmented matrix we end up clearing at most as many columns as there are equations since we use a different row each time we clear a column, and there are only as many rows as there are equations. This means that our final answer has a column corresponding to a variable which can be chosen freely. Hence we can choose that variable to be nonzero. ∎

Example: A homogeneous system with nontrivial solutions

The system

$$\begin{aligned} x+y+z+w &= 0 \\ x+y+2z+3w &= 0 \end{aligned}$$

has nontrivial solutions:

We start by applying the Gaussian Elimination with backsolving algorithm:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 3 & 0 \end{bmatrix} \underset{\rightsquigarrow}{R_2 - R_1} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \end{bmatrix}$$

$$\underset{\rightsquigarrow}{R_1 - R_2} \begin{bmatrix} 1 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 \end{bmatrix}$$

Since we have cleared only the columns corresponding to the variables x and z in this case, both y and w can be chosen arbitrarily. Letting $y = 1$ and $w = 1$, we get $x = w - y = 0$ and $z = -2w = -2$. Thus $(0, 1, -2, 1)$ is a nontrivial solution. ♢

Exercises 7.2:
For problems 1–10, is the vector $\vec{a}$ in the subspace of $\mathcal{V}$ spanned by S?

1. $\vec{a} = [1,2,3]$, $S = \{[1,1,3],[-2,1,-3]\}$, $\mathcal{V} = \mathbb{R}^3$
2. $\vec{a} = [1,1,1]$, $S = \{[0,1,2],[1,2,1],[1,3,3]\}$, $\mathcal{V} = \mathbb{R}^3$
3. $\vec{a} = [6,5,6]$, $S = \{[1,2,4],[0,1,2],[1,1,1]\}$, $\mathcal{V} = \mathbb{R}^3$
4. $\vec{a} = [6,2,3,1]$, $S = \{[1,2,3,4],[0,1,2,3],[0,0,1,2]\}$, $\mathcal{V} = \mathbb{R}^4$
5. $\vec{a} = [1,0,1,0]$, $S = \{[1,1,1,1],[1,-1,-1,1],[1,1,-1,-1]\}$, $\mathcal{V} = \mathbb{R}^4$
6. $\vec{a} = [1,0,1,2]$, $S = \{[1,1,1,1],[1,-1,-1,1],[1,1,-1,-1]\}$, $\mathcal{V} = \mathbb{R}^4$
7. $\vec{a} = [0,1,0,1]$, $S = \{[1,1,0,0],[1,0,1,0],[0,0,1,1]\}$, $\mathcal{V} = \mathbb{R}^4$
8. $\vec{a} = 3x^3 + 4x^2 + x - 2$, $S = \{x+1, x^2+1, x^3+1\}$, $\mathcal{V} = \mathbb{R}[x]$
9. $\vec{a} = x^4 + x^2 - 1$, $S = \{1, x, x^2+x+1, x^4+x^3+1, x^3+1\}$, $\mathcal{V} = \mathbb{R}[x]$
10. $\vec{a} = \sin(x)$, $S = \{\cos(x), x^2, e^x\}$, $\mathcal{V} = \mathbb{R}^{\mathbb{R}}$

7.3 Applications of Systems of Linear Equations

In the last section we saw a situation in which systems of linear equations arises in linear algebra. Many practical problems involve systems of linear equations. Now that we have an efficient means to solve such systems problems giving rise to systems of many equations in many unknowns are more tractable.

7.3.1 Fitting a curve through points using the method of undetermined coefficients:

Find the equation of the circle passing through the three points (2,3), (1,4), and (5,6). We know that the equation of a circle has the form

$$x^2 + y^2 + Ax + By + C = 0$$

We need to find A, B, and C.

Since (2,3) is to be on the circle

$$2^2 + 3^2 + 2A + 3B + C = 0.$$

Since (1,4) is to be on the circle

$$1^2 + 4^2 + 1A + 4B + C = 0.$$

Since (5,6) is to be on the circle

$$5^2 + 6^2 + 5A + 6B + C = 0.$$

This gives the system

$$\begin{aligned} 2A + 3B + C &= -13 \\ A + 4B + C &= -17 \\ 5A + 6B + C &= -61. \end{aligned}$$

This is solved as follows:

$$\begin{bmatrix} 2 & 3 & 1 & -13 \\ 1 & 4 & 1 & -17 \\ 5 & 6 & 1 & -61 \end{bmatrix} \begin{matrix} R_1 \leftrightarrow R_2 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 4 & 1 & -17 \\ 2 & 3 & 1 & -13 \\ 5 & 6 & 1 & -61 \end{bmatrix} \begin{matrix} R_2 - 2R_1 \\ R_3 - 5R_1 \\ \rightsquigarrow \end{matrix}$$

$$\begin{bmatrix} 1 & 4 & 1 & -17 \\ 0 & -5 & -1 & 21 \\ 0 & -14 & -4 & 24 \end{bmatrix} \begin{matrix} \frac{-1}{5}R_2 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 4 & 1 & -17 \\ 0 & 1 & \frac{1}{5} & -\frac{21}{5} \\ 0 & -14 & -4 & 24 \end{bmatrix} \begin{matrix} R_3 + 14R_2 \\ \rightsquigarrow \end{matrix}$$

$$\begin{bmatrix} 1 & 4 & 1 & -17 \\ 0 & 1 & \frac{1}{5} & -\frac{21}{5} \\ 0 & 0 & -\frac{6}{5} & -\frac{174}{5} \end{bmatrix} \begin{matrix} -\frac{5}{6}R_3 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 4 & 1 & -17 \\ 0 & 1 & \frac{1}{5} & -\frac{21}{5} \\ 0 & 0 & 1 & 29 \end{bmatrix} \begin{matrix} R_1 - R_3 \\ R_2 - \frac{1}{5}R_3 \\ \rightsquigarrow \end{matrix}$$

$$\begin{bmatrix} 1 & 4 & 0 & -46 \\ 0 & 1 & 0 & -10 \\ 0 & 0 & 1 & 29 \end{bmatrix} \begin{matrix} R_1 - 4R_2 \\ \rightsquigarrow \end{matrix} \begin{bmatrix} 1 & 0 & 0 & -6 \\ 0 & 1 & 0 & -10 \\ 0 & 0 & 1 & 29 \end{bmatrix}$$

so $A = -6$, $B = -10$, and $C = 29$.

This same technique can be used to fit a polynomial of degree n to a set of $n + 1$ points. It can also be used to fit surfaces through points in space, provided that the general form of the surface is known.

7.3.2 Deriving the error for Simpson's rule:

In finding formulas to approximate integrals numerically one approach is to specify the points to be used and find coefficients so that the resulting formula gives the exact answer for appropriate powers of x. For Simpson's rule we use the points $(-h, f(-h))$, $(0, f(0))$ and $(h, f(h))$ and ask that the result be accurate for polynomials of degree 3 or less. If we postulate a formula of the form

$$\int_{-h}^{h} f(x)dx = Af(-h) + Bf(0) + Cf(h) + Df^{iv}(z)$$

where $-h < z < h$ then we can find A, B, C, and D by making different choices for f and using the precision of the formula:

$$\begin{array}{ccccll}
\int_{-h}^{h} 1dx & = & 2h & = & A + B + C & \text{using } f(x) = 1 \\
\int_{-h}^{h} xdx & = & 0 & = & A(-h) + C(h) & \text{using } f(x) = x \\
\int_{-h}^{h} x^2dx & = & \frac{2}{3}h^3 & = & Ah^2 + Ch^2 & \text{using } f(x) = x^2 \\
\int_{-h}^{h} x^3dx & = & 0 & = & A(-h)^3 + Ch^3 & \text{using } f(x) = x^3 \\
\int_{-h}^{h} x^4dx & = & \frac{2}{5}h^5 & = & Ah^4 + Ch^4 + D4! & \text{using } f(x) = x^4
\end{array}$$

This gives the system of equations

$$\begin{aligned}
A + B + C &= 2h \\
-hA + hC &= 0 \\
h^2A + h^2C &= \frac{2}{3}h^3 \\
-h^3A + h^3C &= 0 \\
h^4A + h^4C + 24D &= \frac{2}{5}h^5.
\end{aligned}$$

The second and fourth equations give exactly the same information so this system is redundant.

Solving it gives

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 2h \\ -h & 0 & h & 0 & 0 \\ h^2 & 0 & h^2 & 0 & \frac{2}{3}h^3 \\ -h^3 & 0 & h^3 & 0 & 0 \\ h^4 & 0 & h^4 & 24 & \frac{2}{5}h^5 \end{bmatrix} \begin{array}{c} R_2 + hR_1 \\ R_3 - h^2R_1 \\ R_4 + h^3R_1 \\ R_5 - h^4R_1 \\ \rightsquigarrow \end{array} \begin{bmatrix} 1 & 1 & 1 & 0 & 2h \\ 0 & h & 2h & 0 & 2h^2 \\ 0 & -h^2 & 0 & 0 & -\frac{4}{3}h^3 \\ 0 & h^3 & 2h^3 & 0 & 2h^4 \\ 0 & -h^4 & 0 & 24 & -\frac{8}{5}h^5 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 1 & 1 & 0 & 2h \\ 0 & 1 & 2 & 0 & 2h \\ 0 & 0 & 2h^2 & 0 & \frac{2}{3}h^3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2h^4 & 24 & \frac{2}{5}h^5 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 1 & 1 & 0 & 2h \\ 0 & 1 & 2 & 0 & 2h \\ 0 & 0 & 1 & 0 & \frac{h}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 24 & -\frac{4}{15}h^5 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 0 & h^3 \\ 0 & 1 & 0 & 0 & \frac{4}{3}h \\ 0 & 0 & 1 & 0 & \frac{h}{3} \\ 0 & 0 & 0 & 1 & -\frac{h^5}{90} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This shows that

$$\int_{-h}^{h} f(x)dx = \frac{h}{3}(f(-h) + 4f(0) + f(h)) - \frac{h^5}{90}f^{iv}(z)$$

for some $z \in (-h, h)$.

7.3.3 Random Walks:

In the game of matching pennies two players each flip a coin. If both come up heads or both come up tails, the first player wins both. If they don't match the second player wins both. If the first player starts with 2 cents and the second starts with 3 cents what is the probability that the first player gets 5 cents before the second player does? After one toss the game will change to (1 cents,4 cents) with probability .5 and to (3 cents,2 cents) with probability .5. If we let P_i be the probability of the first player winning if he starts with i cents this tells us that $P_i = .5P_{i-1} + .5P_{i+1}$. Combining this with the obvious cases $P_0 = 0$ and $P_5 = 1$ we get a system of 6 equations in 6 unknowns:

$$\begin{array}{rrrrrrcl} P_0 & & & & & & = & 0 \\ -\frac{1}{2}P_0 & +P_1 & -\frac{1}{2}P_2 & & & & = & 0 \\ & -\frac{1}{2}P_1 & +P_2 & -\frac{1}{2}P_3 & & & = & 0 \\ & & -\frac{1}{2}P_2 & +P_3 & -\frac{1}{2}P_4 & & = & 0 \\ & & & -\frac{1}{2}P_3 & +P_4 & -\frac{1}{2}P_5 & = & 0 \\ & & & & & P_5 & = & 1 \end{array}$$

This is not difficult to solve because there are so many zero entries:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \rightsquigarrow$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{2}{5} \\ 0 & 0 & 0 & 1 & 0 & 0 & \frac{3}{5} \\ 0 & 0 & 0 & 0 & 1 & 0 & \frac{4}{5} \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Thus the probability that the first player wins if he starts out with 2 cents is $\frac{2}{5}$.

This same kind of model can be used to approximate temperature distributions and diffusion of gases or liquids. The systems of equations which result are often very large but tend to have many entries equal to zero and have the nonzero entries clustered in a band near the diagonal. This extra structure greatly reduces the amount of arithmetic needed to solve the system.

7.3.4 Partial Fractions

In technique of integration we develop a general method for dealing with integrals of rational functions. First we factor the denominator into a product of linear and irreducible quadratic factors. Then we use partial fractions to reduce the problem to dealing with fractions whose denominators involve powers of a single linear or quadratic factor. The resulting problems are then integrated using u-substitution or trigonometric substitutions. In a calculus course we often restrict the problem to one using denominators which are products of distinct linear factors because the systems of linear equations which result are easy to solve without special technique.

Now that we have powerful methods for solving big systems of linear equations we can tackle the more difficult systems that result when the denominator involves repeated factors.

As an example let us consider the problem of breaking

$$\frac{2x^5 + 13x^4 + 35x^3 + 43x^2 + 35x + 16}{(x+2)^2(x^2+x+1)^2}$$

into a sum of simpler fractions. First notice that the degree of the denominator is 6, so the degree of the most general possible numerator will be 5 and there will be 6 pieces of information contained in the coefficients. Thus we expect to have a system of six equations in six unknowns. The standard way to break up this fraction is as

$$\frac{A}{x+2} + \frac{B}{(x+2)^2} + \frac{Cx+D}{x^2+x+1} + \frac{Ex+F}{(x^2+x+1)^2}$$

which has six unknowns (A,B,C,D,E, and F). Adding fractions in this expression gives a fraction with numerator

$$\begin{array}{l} (2A + B + 4D + 4F) + (5A + 2B + 4C + 8D + 4E + 4F)x + \\ (8A + 3B + 8C + 9D + 4E + F)x^2 + (7A + 2B + 9C + 5D + E)x^3 + \\ (4A + B + 5C + D)x^4 + (A + C)x^5 \end{array}$$

and denominator $(x+2)^2(x^2+x+1)^2$. Now this fraction will be equal to our original fraction if the polynomials in the numerator are equal. This happens if the constant term and the coefficients of x, x^2, x^3, x^4, and x^5 are equal. This gives us the system of equations

$$\begin{aligned} 2A + B + 4D + 4F &= 16 \\ 5A + 2B + 4C + 8D + 4E + 4F &= 35 \\ 8A + 3B + 8C + 9D + 4E + F &= 43 \\ 7A + 2B + 9C + 5D + E &= 35 \\ 4A + B + 5C + D &= 13 \\ A + C &= 2 \end{aligned}$$

Solving this system is done by row reducing the matrix

$$\begin{bmatrix} 2 & 1 & 0 & 4 & 0 & 4 & 16 \\ 5 & 2 & 4 & 8 & 4 & 4 & 35 \\ 8 & 3 & 8 & 9 & 4 & 1 & 43 \\ 7 & 2 & 9 & 5 & 1 & 0 & 35 \\ 4 & 1 & 5 & 1 & 0 & 0 & 13 \\ 1 & 0 & 1 & 0 & 0 & 0 & 2 \end{bmatrix}$$

to get

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

so

$$\frac{2x^5+13x^4+35x^3+43x^2+35x+16}{(x+2)^2(x^2+x+1)^2} = \frac{-1}{x+2} + \frac{-2}{(x+2)^2} + \frac{3x+4}{x^2+x+1} + \frac{-1x+1}{(x^2+x+1)^2}$$

Exercises 7.3:

Several of these problems involve large systems; use of a symbolic computer algebra system is recommended.

1. Show that a polynomial of degree 3 with at least 4 distinct roots must be the zero polynomial.

2. The trapezoid rule for approximating $\int_0^h f(x)\,dx$ gives the exact answer for polynomials of degree less than or equal to 1. The error term is a multiple of $f''(z)$ for some $z \in (0,h)$. Use this information to derive the trapezoid rule with its error term.

3. Simpson's 3/8 rule approximates integrals using

$$\int_0^{3h} f(x)\,dx = a_1 f(0) + a_2 f(h) + a_3 f(2h) + a_4 f(3h) + c f^{iv}(\xi)$$

   where $\xi \in (0,3h)$. It gives exact answers for polynomials of degree 3 or less. Find the coefficients a_i and the coefficient for the error c.

4. Find the equation of the circle through the points (1,2), (3,4), (10,1).

5. Find the polynomial of degree 3 through the points (0,0),(1,0),(2,4), and (3,0).

6. A plane in $\mathbb{R}^3$ has an equation of the form $Ax + By + C = z$. Find the equation of the plane through the points $(1,2,3),(2,1,3),(1,-3,-5)$.

7. In the model of matching pennies, suppose that the participants are betting on an unfair event, so that the probability that the first player's fortune goes from i to $i-1$ is $\frac{1}{4}$ instead of $\frac{1}{2}$ and the probability that it goes from i to $i+1$ is $\frac{3}{4}$. Rework the example for this random walk with drift assuming that there are a total of 5 pennies at stake.

8. (Computer problem) A slot machine is designed so that it pays 3 times your bet with probability $\frac{1}{4}$ and nothing with probability $\frac{3}{4}$. You have \$5 and decide to bet until you either have \$15 or nothing. Which strategy gives you a better chance of avoiding going broke: betting the whole \$5 at once, or betting in \$1 increments until you either reach \$15 or go broke?

9. Give a partial fractions expansion for

$$\frac{-2+x-4x^2-x^3+2x^5}{(x-1)^2(x^2+1)^2}$$

10. Give a partial fractions expansion for

$$\frac{254+96x+51x^2+65x^3-10x^4+14x^5-2x^6+x^7}{(x-2)^2(x^2+5)^3}$$

Chapter 8

Linear Independence and Bases

In Chapter 6 we discussed the subspace determined by the set of all linear combinations of a set A of vectors, the linear span of A. Turning it around, we called A a spanning set for the subspace Span(A). If A contains an element $\vec{a}$ which is a linear combination of other elements of A then $\text{Span}(A \setminus \{\vec{a}\} = \text{Span}(A)$. In this chapter we want to consider the idea of an efficient or minimal spanning set for a vector space; we call such a set a basis. A basis will give us a convenient handle with which we can "get hold" of a vector space. First, however, we must introduce several new ideas.

8.1 Linear Independence

Let $\mathcal{V}$ be a vector space and suppose we have a set $A = \{\vec{a}_1, \ldots, \vec{a}_n\}$ of vectors in $\mathcal{V}$. We consider the possibility that a linear combination of the vectors in A could equal the zero vector:

$$c_1\vec{a}_1 + \ldots + c_n\vec{a}_n = \vec{0}.$$

Obviously this can always be done by taking all the scalars to be zero: $c_1 = \ldots = c_n = 0$. In some cases this is the *only* way the equation will hold, but in others it is possible to obtain $\vec{0}$ as a linear combination of the vectors in A by using some nonzero scalars. (As trivial examples of these two possibilities, let $A = \{\vec{a}\}$ where $\vec{a} \neq \vec{0}$, then $c\vec{a} = \vec{0}$ only if $c = 0$. Next let $A = \{\vec{a}_1, \vec{a}_2\}$, where $\vec{a}_2 = -\vec{a}_1$; then $1\vec{a}_1 + 1\vec{a}_2 = 1\vec{a}_1 + 1(-\vec{a}_1) = \vec{0}$.) The distinction between these two possibilities turns out to be of paramount importance.

Definition 8.1.1 *A set A of vectors in $\mathcal{V}$ is* **linearly dependent** *if and only if $\vec{0}$ can be written as a linear combination of vectors in A in which there are nonzero coefficients. A is* **linearly independent** *if whenever a linear combination of elements of A has $c_1\vec{a}_1 + \ldots + c_n\vec{a}_n = \vec{0}$ all of the coefficients must be 0 (we have $c_1 = c_2 = \ldots = c_n = 0$).*

Example: A linearly independent set

Let $A = \{[1,0,0],[0,1,0],[0,0,1]\} \subseteq \mathbb{R}^3$. Then $c_1[1,0,0] + c_2[0,1,0] + c_3[0,0,1] = [c_1, c_2, c_3]$, and thus $c_1[1,0,0] + c_2[0,1,0] + c_3[0,0,1] = [0,0,0]$ if and only if $c_1 = c_2 = c_3 = 0$. Thus A is independent (we shall frequently omit the adverb "linearly"). $\diamondsuit$

Example: A dependent set

Let $A = \{\vec{a}_1, \vec{a}_2\} \subseteq \mathcal{V}$, where $\vec{a}_2 = k\vec{a}_1$, $k \neq 0$. Then

$$\begin{aligned} k\vec{a}_1 + (-1)\vec{a}_2 &= k\vec{a}_1 + (-1)k\vec{a}_1 \\ &= k\vec{a}_1 - k\vec{a}_1 \\ &= 0, \end{aligned}$$

and this set A is dependent. $\diamondsuit$

Example: Any set containing $\vec{0}$ is dependent

Let $A = \{\vec{0}, \vec{a}_1, \ldots, \vec{a}_n\}$. Then A is dependent, for

$$1\vec{0} + 0\vec{a}_1 + \ldots + 0\vec{a}_n = \vec{0}.$$

In other words, any set which contains the zero vector is dependent. (By way of contrast, as we have already seen, any set A which contains exactly one vector is independent if that vector is nonzero.) $\diamondsuit$

Example: Examples in $\mathbb{R}^3$

Let $\mathcal{V} = \mathbb{R}^3,\ A = \{[1,2,3],[2,4,6]\},\ B = \{[1,2,3],[2,1,0]\}$. Then A is dependent, since

$$2[1,2,3] + (-1)[2,4,6] = [2,4,6] + [-2,-4,-6] = \vec{0}$$

but B is independent. For

$$\begin{aligned} c_1[1,2,3] + c_3[2,1,0] &= [c_1, 2c_1, 3c_1] + [2c_3, c_3, 0] \\ &= [c_1 + 2c_3, 2c_1 + c_3, 3c_1] \\ &= [0,0,0] \end{aligned}$$

can hold only if $c_1 = 0$, and this implies that $c_3 = 0$ also. $\diamondsuit$

Example:

Let $\mathcal{V} = \mathbb{R}^3$, let $A = \{\vec{a}_1 = [1,2,-1], \vec{a}_2 = [2,0,1], \vec{a}_3 = [-1,1,0], \vec{a}_4 = [1,6,-4]\}$. We consider first the subset $A_1 = \{\vec{a}_1, \vec{a}_2, \vec{a}_3\}$. If

$$c_1\vec{a}_1 + c_2\vec{a}_2 + c_3\vec{a}_3 = 0,$$

then

$$[c_1 + 2c_2 - c_3, 2c_1 + 3c_3, -c_1 + c_2] = [0,0,0],$$

or

$$\begin{aligned} c_1 + 2c_2 - c_3 &= 0 \\ 2c_1 + 3c_3 &= 0 \\ -c_1 + c_2 &= 0. \end{aligned}$$

It is easy to show, either by the methods of Chapter 7 or by solving for c_2 and c_3 in terms of c_1, that the only solution to this homogeneous system is the trivial one. Thus A_1 is an independent subset. Next, however, consider $A_2 = \{\vec{a}_1, \vec{a}_2, \vec{a}_4\}$. If

$$c_1\vec{a}_1 + c_2\vec{a}_2 + c_4\vec{a}_4 = 0,$$

then we are led to the system

$$\begin{aligned} c_1 + 2c_2 + c_4 &= 0 \\ 2c_1 + 6c_4 &= 0 \\ -c_1 + c_2 - 4c_4 &= 0 \end{aligned}$$

Using Gaussian elimination, we find that $c_1 = -3c_4, c_2 = c_4$, and c_4 is arbitrary; or, letting $c_4 = -1$, the linear combination $3\vec{a}_1 - \vec{a}_2 - \vec{a}_4 = 0$. Thus the set A_2 is dependent. $\diamondsuit$

Example: Independence in polynomial spaces

Let $\mathcal{V} = \mathbb{R}[x]_n$, then the set $A = \{1, x, x^2, \ldots x^n\}$ is independent. To investigate independence of A we need to see if there are any non-trivial choices of c_i so that

$$c_0 + c_1 x + c_2 x^2 + \ldots + c_n x^n = 0,$$

where the right-hand side is the zero polynomial. The question of independence does not present us with the problem of finding the n roots of an algebraic equation, but of identifying the zero polynomial as one which can be represented only by $c_0 = c_1 = \ldots = c_n = 0$. Since the zero polynomial has all coefficients 0, the set A is independent. $\diamondsuit$

Example: Any set of three vectors in $\mathbb{R}^2$ is dependent

Let $A = \{\vec{a}, \vec{b}, \vec{c}\} \subseteq \mathbb{R}^2$. Then A is linearly dependent, for, if

$$h\vec{a} + k\vec{b} + l\vec{c} = \vec{0},$$

then

$$h[a_1, a_2] + k[b_1, b_2] + l[c_1, c_2] = [0, 0],$$

or

$$\begin{aligned} a_1 h + b_1 k + c_1 l &= 0 \\ a_2 h + b_2 k + c_2 l &= 0, \end{aligned}$$

and this homogeneous system of two equations in three unknowns will have nontrivial solutions. In a similar way, any set $A = \{\vec{a}_1, \ldots \vec{a}_n, \vec{a}_{n+1}\} \subseteq \mathbb{R}^n$ can be shown to be dependent. $\diamondsuit$

Proposition 8.1.1 *The empty set is linearly independent.*

PROOF:

Since there is no way to write $\vec{0}$ as a linear combination of vectors in the empty set (there being no such vectors to work with), it is vacuously true that every such linear combination has any property we wish. In particular we can conclude that it must be trivial. ∎

We now look at a few alternative but equivalent ways of describing a dependent set.

Proposition 8.1.2 *A set $A = \{\vec{a}_1, \ldots, \vec{a}_n\} \subseteq \mathcal{V}$ is dependent if and only if some $\vec{a}_j \in A$ is a linear combination of the other vectors in A.*

PROOF:

If A is dependent, then scalars $c_1, \ldots c_n$ can be found, not all zero, such that

$$c_1\vec{a}_1 + \ldots + c_n\vec{a}_n = \vec{0}.$$

If $c_j \neq 0$, then

$$c_j\vec{a}_j = -c_1\vec{a}_1 - \ldots - c_{j-1}\vec{a}_{j-1} - c_{j+1}\vec{a}_{j+1} - \ldots - c_n\vec{a}_n$$

or,

$$\vec{a}_j = \frac{-c_1}{c_j}\vec{a}_1 - \ldots + \frac{-c_{j-1}}{c_j}\vec{a}_{j-1} + \frac{-c_{j+1}}{c_j}\vec{a}_{j+1} + \ldots + \frac{-c_n}{c_j}\vec{a}_n.$$

On the other hand, if

$$\vec{a}_j = k_1\vec{a}_1 + \ldots + k_{j-1}\vec{a}_{j-1} + k_{j+1}\vec{a}_{j+1} + \ldots + k_n\vec{a}_n,$$

then

$$k_1\vec{a}_1 + \ldots + k_{j-1}\vec{a}_{j-1} + (-1)\vec{a}_j + k_{j+1}\vec{a}_{j-1} + \ldots + k_n\vec{a}_n = \vec{0},$$

and, as $-1 \neq 0$, A is dependent. ∎

The next description is very similar to the preceding one, but its special form will be useful later.

Proposition 8.1.3 *An ordered set $A = (\vec{a}_1, \ldots, \vec{a}_n)$ of vectors in $\mathcal{V}$ is dependent if and only if some $\vec{a}_j \in A$ is a linear combination of the preceding $\vec{a}_i$.*

PROOF:

If A is dependent, then, by definition there are scalars such that

$$c_1\vec{a}_1 + \ldots + c_n\vec{a}_n = \vec{0},$$

where the c_i are not all zero. Let j be the largest integer such that $c_j \neq 0$ (thus $c_{j+1} = \ldots = c_n = 0$). Then $c_1\vec{a}_1 + \ldots + c_j\vec{a}_j = \vec{0}, c_j \neq 0$, and we can write

$$\vec{a}_j = \frac{-c_1}{c_j}\vec{a}_1 + \ldots + \frac{-c_{j-1}}{c_j}\vec{a}_{j-1}.$$

Going the other way, if $\vec{a}_j = k_1\vec{a}_1 + \ldots + k_{j-1}\vec{a}_{j-1}$, then

$$k_1\vec{a}_1 + \ldots + k_{j-1}\vec{a}_{j-1} + (-1)\vec{a}_j + 0\vec{a}_{j+1} + \ldots + 0\vec{a}_n = 0,$$

and A is dependent. ∎

Exercises 8.1:

For problems 1 through 13 Determine whether the following sets of vectors are independent or dependent:

1. $\{[1, 0, 3], [3, -4, 3], [2, 9, 0]\}$ in $\mathbb{R}^3$
2. $\{[1, 2, 3, 4], [2, 4, 6, 1], [3, 3, 1, 2], [-1, 2, -3, 1]\}$ in $\mathbb{R}^4$
3. $\{[1, 1, 3], [3, 1, 2], [1, 5, -2], [1, 0, 3]\}$ in $\mathbb{R}^3$
4. $\{x + 3, x^2 - x - 4, -x^2 + 2\}$ in $\mathbb{R}[x]_3$
5. $\{\sin x, \cos x, \sin 2x\}$ in C[0,2π]
6. $\{[4, 6, -2, 3], [2, 1, 3, -2], [1, 0, 1, 0]\}$ in $\mathbb{R}^4$
7. $\{x + 1, x - 1, x^2 + x, x^2 - 1, x^3 + x^2\}$ in $\mathbb{R}$[x]
8. $\{f_k | f_k(x) = 1 + kx, k \in \mathbb{N}\}$ in $\mathbb{R}^{\mathbb{R}}$
9. $\{g_k | g_k(x) = x^k, k \in \mathbb{N}\}$ in $\mathbb{R}^{\mathbb{R}}$

10. $\{[0,1,1,1,0,1,1],[1,1,0,0,1,1,0],[1,0,1,0,1,0,1],[1,1,0,0,0,1,1]\}$ in $\mathbb{Z}_2^7$.

11. $\{[0,0,1,1,0],[1,1,1,0,0],[1,0,1,0,1],[0,1,1,1,0],[0,0,0,1,1]\}$ in $\mathbb{Z}_2^5$

12. $\{[1,i],[i,1]\}$ in $\mathbb{C}^2$

13. $\{[1+i,1-i],[2,-i]\}$ in $\mathbb{C}^2$

14. Show that if $\{\vec{a}_1,\ldots,\vec{a}_n\}\subset\mathcal{U}$ is linearly independent as a subset of $\mathcal{U}$ and $\mathcal{U}$ is a subspace of $\mathcal{V}$ then if $\{\vec{a}_1,\ldots,\vec{a}_n\}\subset\mathcal{U}$ is linearly independent as a subset of $\mathcal{V}$.

15. Show that:

    (a) If A is dependent and $A\subseteq B$ then B is dependent

    (b) If B is independent and $A\subseteq B$ then A is independent

16. Prove the following: if $A=\{\vec{a}_1,\ldots,\vec{a}_m\}\subseteq\mathcal{V}$, $\vec{b}\notin\mathrm{Span}(A)$, and A is linearly independent then $A^*=\{\vec{a}_1,\ldots,\vec{a}_m,\vec{b}\}$ is linearly independent.

17. Show that if S_1 and S_2 are linearly independent sets of vectors and
$$\mathrm{Span}(S_1)\cap\mathrm{Span}(S_2)=\{\vec{0}\}$$
then $S_1\cup S_2$ is linearly independent. Give an example to show that it is not sufficient to have only $S_1\cap S_2=\emptyset$.

18. An ascending chain of sets is a sequence S_i with
$$S_1\subseteq S_2\subseteq\cdots\subseteq S_n\subseteq\cdots$$
Prove that if each of the S_i in an ascending chain is linearly independent then so is $\bigcup S_i$.

19. Prove that a subset of the columns in a matrix is linearly dependent if and only if that same set of columns is dependent after any elementary row operation has been performed.

20. Prove that if $\mathbf{A}$ is an $m\times n$ matrix with $m\geq n$, then the columns of $\mathbf{A}$ are linearly independent if and only if the row reduced echelon form $\mathbf{M}$ of $\mathbf{A}$ has $m_{ii}=1$ for all $1\leq i\leq n$.

21. Prove that the set of nonzero rows of a matrix in row reduced echelon form is linearly independent.

8.2 Bases

We are now in a position to define and elaborate on the idea of an *economical* or *efficient* spanning set of vectors for a vector space $\mathcal{V}$. The property we need to add to that of a spanning set to make it efficient is independence.

Definition 8.2.1 *A* **basis** *for a vector space $\mathcal{V}$ is a set B of vectors which spans $\mathcal{V}$ and is linearly independent. The plural of basis is bases; we will often be concerned with several at once.*

Example: Bases

The following are bases:

1. $\mathcal{V} = \mathbb{R}^3$, $B = \{[1,0,0],[0,1,0],[0,0,1]\}$, the **standard** basis
2. $\mathcal{V} = \mathbb{R}^3$, $B = \{[1,1,-1],[1,0,1],[0,1,1]\}$
3. $\mathcal{V} = \mathbb{R}[x]_n$, $B = \{1, x, x^2, \ldots, x^n\}$
4. $\mathcal{V} = \mathbb{R}[x]$, $B = \{1, x, x^2, \ldots, x^n, \ldots\}$
5. $\mathcal{V} = \mathbb{R}^n$, $B = \{[1,0,0,\ldots,0],[0,1,0,\ldots,0],\ldots,[0,0,\ldots,0,1]\}$.
6. $\mathcal{V} = \{\vec{0}\}$, $B = \emptyset$

$\diamondsuit$

Theorem 8.2.1 *If B is a basis for a vector space $\mathcal{V}$, then each vector $\vec{v}$ in $\mathcal{V}$ can be written as a linear combination of elements of B in exactly one way.*

PROOF:

Since $\text{Span}(B) = \mathcal{V}$ we know that every vector can be written in *at least* one way as a linear combination of elements of B. What we really need to show is uniqueness. Suppose that $\vec{v}$ can be written as a linear combination in two ways. Since only a finite number of elements of B are involved in each linear combination, we could list all elements of B that are used in both linear combinations by using 0 as a coefficient as needed:

$$\begin{aligned} \vec{v} &= \sum_{i=1}^{n} h_i \vec{b}_i \\ &= \sum_{i=1}^{n} k_i \vec{b}_i \end{aligned}$$

Subtracting and combining terms gives

$$\vec{0} = \sum_{i=1}^{n} (h_i - k_i)\vec{b}_i.$$

Since the set B is linearly independent this tells us that $h_i - k_i = 0$ for every i. Thus our linear combinations giving $\vec{v}$ were not, in fact, different after all. ∎

Remark: Although by definition a basis is a set of vectors, which means that the order of listing the vectors should not be relevant, it turns out that we often do consider the order important. In other words, we will usually assume that we are working with *ordered* bases.

If we have a finite ordered basis we can use it to think of any vector as a column of numbers.

Definition 8.2.2 *If $B = (\vec{b}_1, \vec{b}_n)$ is an ordered basis for $\mathcal{V}$, then the B-coordinate representation of $\vec{a} = k_1\vec{b}_1 + \ldots + k_n\vec{b}_n$ is the column vector* $\begin{bmatrix} k_1 \\ \vdots \\ k_n \end{bmatrix}$. *Because column vectors are awkward within text, we often write this column as $[k_1, \ldots, k_2]^t$, where the ^{t} indicates transpose, an operation which turns rows into columns.*

Example: Finding B-coordinates

We noted above that $B = ([1, 1, -1], [1, 0, 1], [0, 1, 1])$ is an ordered basis for $\mathbb{R}^3$. We can use it to find B-coordinates for the vector $[4, 5, 6]$ by solving the system of equations

$$\begin{aligned} 1x + 1y + 0z &= 4 \\ 1x + 0y + 1z &= 5 \\ -1x + 1y + 1z &= 6 \end{aligned}$$

Using row reduction on the augmented matrix for this we get

$$\begin{bmatrix} 1 & 1 & 0 & 4 \\ 1 & 0 & 1 & 5 \\ -1 & 1 & 1 & 6 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 4 \end{bmatrix}$$

so $x = 1$, $y = 3$, and $z = 4$. This tells us that the B-coordinates of $[4, 5, 6]$ give the column

$$\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix},$$

the same as the rightmost column in the final matrix in the row reduction. $\diamondsuit$

In our examples above, most of the bases were finite sets. This case deserves special mention.

Definition 8.2.3 *A vector space $\mathcal{V}$ is* **finite dimensional** *if and only if $\mathcal{V}$ has a finite basis. If $\mathcal{V}$ is not finite dimensional it is infinite dimensional.*

The basis given for $\mathbb{R}[x]$ above is not finite. A little reflection will suggest that $\mathbb{R}[x]$ cannot have a finite basis. Thus $\mathbb{R}[x]$ is infinite dimensional. Referring again to our examples, we observe that both bases given for $\mathbb{R}^3$ have 3 vectors (and that, more generally, the basis for $\mathbb{R}^n$ has n vectors). Is this coincidental? The answer is "No": all bases for a given finite-dimensional space must have the same number of vectors. But before we prove it we must prove a related assertion.

Theorem 8.2.2 *Let $\mathcal{V}$ be finite dimensional. Let $B = \{\vec{b}_1, \ldots, \vec{b}_n\}$ span $\mathcal{V}$, let $A = \{\vec{a}_1, \ldots, \vec{a}_m\}$ be a linearly independent set. Then $n \geq m$.*

In other words, any spanning set of vectors must have at least as many vectors as any independent set.
PROOF:

The proof involves introducing, one at a time, the vectors in A into the set B, but then removing, one at a time, a vector from this new set. The desired result is obtained by some simple counting. In the following two columns the properties listed at

the top of each will hold at each stage, as you should verify.

$$\begin{array}{cc}
\text{dependent} & \text{spans } \mathcal{V} \\
n+1 \text{ vectors} & n \text{ vectors} \\
B_1 = (\vec{a}_1, \vec{b}_1, \ldots, \vec{b}_n) & B_1{*} = \{\vec{a}_1, \vec{b}_1, \vec{b}_2, \ldots, \vec{b}_{n-1}\} \\
B_2 = (\vec{a}_2, \vec{a}_1, \vec{b}_1, \ldots, \vec{b}_{n-1}) & B_2{*} = \{\vec{a}_2, \vec{a}_1, \vec{b}_1, \vec{b}_2, \ldots, \vec{b}_{n-2}\} \\
B_3 = (\vec{a}_3, \vec{a}_2, \vec{a}_1, \vec{b}_1, \ldots, \vec{b}_{n-2}) & B_3{*} = \{\vec{a}_3, \vec{a}_2, \vec{a}_1, \vec{b}_1, \vec{b}_2, \ldots, \vec{b}_{n-3}\} \\
\vdots & \vdots \\
B_m = (\vec{a}_m, \ldots, \vec{a}_1, \vec{b}_1, \ldots, \vec{b}_{n-m+1}) &
\end{array}$$

Explanation: Since B spans $\mathcal{V}$ every vector, hence, $\vec{a}_1$ must be a linear combination of the vectors in B; thus $B_1 = \{\vec{a}_1\} \cup B$ is linearly dependent. Clearly B_1 has $n+1$ vectors. Order B_1 so that a_1 comes first. Now since B_1 is dependent, some vector in B_1 is a linear combination of the preceding vectors. This vector is clearly not $\vec{a}_1$; it is therefore one of the $\vec{b}_i$'s. We renumber, if necessary, to make it $\vec{b}_n$. Then $\vec{b}_n$ can be deleted from B_1 to give $B_1{*}$, a set of n vectors which spans $\mathcal{V}$. We now obtain B_2 by adding $\vec{a}_2$ to the beginning of $B_1{*}$. Then B_2 is a set with $n+1$ vectors, dependent for the same reason B_1 is dependent. Again, as with B_1, some vector must be a linear combination of the preceding ones. This vector cannot be any of the $\vec{a}_i$, since A is independent, and thus must be a $\vec{b}_i$. Again, we renumber if necessary to make it $\vec{b}_{n-1}$ which can then be eliminated, to give $B_2{*}$, a spanning set with n vectors. And so on.

How will the story end? Continue to

$$B_m = \{\vec{a}_m, \ldots, \vec{a}_2, \vec{a}_1, \vec{b}_1, \ldots, \vec{b}_{n-m+1}\}.$$

All m of the vectors in A have been introduced and $m-1$ of the $\vec{b}_i$ have been deleted. Will B_m still include any of the $\vec{b}_i$? It must, for the sets B_i are always dependent, and the absence of all of the $\vec{b}_i$ would mean $B_m = A$, an independent set. Thus we must have $n - m + 1 \geq 1$, or $n - m \geq 0$, so $n \geq m$. ∎

Now we can easily prove the following fundamental result.

Corollary 8.2.3 *If $\mathcal{V}$ is finite-dimensional then any two bases for $\mathcal{V}$ have the same number of vectors.*

PROOF:

Let $B = \{\vec{b}_1, \ldots, \vec{b}_m\}$ and $B' = \{\vec{b}'_1, \ldots, \vec{b}'_n\}$ be bases. Then since B is a spanning set and B' is independent, we must have $m \geq n$. But, reversing the roles, B' is a spanning set and B is independent, so $n \geq m$. It follows that $m = n$. ∎

With the knowledge provided by this corollary we can make more precise the concept of dimension for finite-dimensional vector spaces.

Definition 8.2.4 *The* **dimension** *of a finite-dimensional vector space $\mathcal{V}$ is the number of vectors in a basis for $\mathcal{V}$. This will be denoted by* $\dim(\mathcal{V})$.

Example: Dimension of some familiar spaces

We can now say that $\dim(\mathbb{R}^3) = 3$, $\dim(\mathbb{R}^n) = n$, and $\dim(\mathbb{R}[x]_n) = n + 1$. ♢

Corollary 8.2.4 *If $\mathcal{V}$ is a vector space of dimension n then any set of $n+1$ or more vectors in $\mathcal{V}$ is dependent.*

PROOF:

If we had a set of $n + 1$ vectors which was independent then we would have an independent set with more vectors in it than there are in a basis, which is a spanning set. This is impossible by the theorem, so any set of $n + 1$ (or more) vectors must be dependent. ∎

Exercises 8.2:

1. Is $B = ([1, 0, 0], [1, 1, 0], [1, 2, 3])$ an ordered basis for $\mathbb{R}^3$? If so, what are the B-coordinates of $[4, 5, 6]$?

2. Is $B = ([2, -1, 3], [0, 1, 4], [2, 0, 0])$ an ordered basis for $\mathbb{R}^3$? If so, what are the B-coordinates of $[4, 5, 6]$?

3. Is $B = ([2, -1, 3], [0, 1, -4], [2, 0, -1], [1, 1, 1])$ an ordered basis for $\mathbb{R}^3$? If so, what are the B-coordinates of $[4, 5, 6]$?

4. Is $B = ([1, 1, 3], [0, 1, 4])$ an ordered basis for $\mathbb{R}^3$? If so, what are the B-coordinates of $[4, 5, 6]$?

5. Is $\{1, x + 1, x^2 + 2x + 1, x^3 + 3x^2 + 3x + 1\}$ a basis for $\mathbb{R}[\mathrm{x}]_3$?

6. Is $\{x - 1, x + 1, x^2 + 2x + 1, 3x^2 + 3x + 1, x^3\}$ a basis for $\mathbb{R}[\mathrm{x}]_3$?

7. Is $\{x, x^2 - 3x, 5x^3 + 3x^2 + 1\}$ a basis for $\mathbb{R}[\mathrm{x}]_3$?

8. Is $\{x, x^2 + x + 1, x + 1\}$ a basis for $\mathbb{R}[x]_2$?

9. Is $\{[1 + i, 0], [i, i]\}$ a basis for $\mathbb{C}^2$?

10. Is $\{[1, 0, 1, 0, 1], [1, 1, 1, 0, 0], [0, 1, 1, 1, 0], [1, 1, 0, 1, 1], [1, 1, 1, 1, 1]\}$ a basis for $\mathbb{Z}_2^5$?

11. Show that if $\mathcal{V}$ is a vector space over $\mathbb{C}$ of dimension n then it is a vector space over $\mathbb{R}$ of dimension $2n$.

12. Suppose that $\{\vec{b}_1, \ldots, \vec{b}_n\}$ is a basis for $\mathbb{Q}^n$. Is it also a basis for $\mathbb{R}^n$?

13. Bases for function spaces can be quite difficult to find. For example, consider $\mathbb{R}^{\mathbb{N}}$, the space of sequences of real numbers. We might try to get a basis by taking $\{f_i\}$ where

$$f_i(n) = \begin{cases} 0 & \text{if } i \neq n \\ 1 & \text{if } i = n \end{cases}$$

Show that this family of sequences is independent but that it does not span the space of all sequences.

14. Show that if $\{\vec{b}_1, \ldots, \vec{b}_n\}$ is a basis then

$$\mathcal{V} = \mathrm{Span}(\{\vec{b}_1, \ldots, \vec{b}_k\}) \oplus \mathrm{Span}(\{\vec{b}_{k+1}, \ldots, \vec{b}_n\})$$

(Recall that $\mathcal{V} = \mathcal{W} \oplus \mathcal{U}$ if and only if $\mathcal{V} = \mathcal{W} + \mathcal{U}$, and $\mathcal{W} \cap \mathcal{U} = \{\vec{0}\}$.)

15. Conversely, show that if $\mathcal{V} = \mathcal{W} \oplus \mathcal{U}$ and $\{\vec{b}_1, \ldots, \vec{b}_n\}$ and $\{\vec{c}_1, \ldots, \vec{c}_m\}$ are bases for $\mathcal{W}$ and $\mathcal{U}$, respectively, then $\{\vec{b}_1, \ldots, \vec{b}_n, \vec{c}_1, \ldots, \vec{c}_m\}$ is a basis for $\mathcal{V}$.

8.3 Finding Bases

There are essentially two different approaches to constructing a basis: start small with an independent set and build up until you get a spanning set, or start big with a spanning set and then reduce the size until you get independence. Both approaches work. An important tool appeared as Exercise 4 of Section 5.1. We repeat it here for easy reference.

Proposition 8.3.1 *Let $A = \{\vec{a}_1, \ldots, \vec{a}_n\}$ be an independent set of vectors in $\mathcal{V}$ and let $\vec{b}$ be a vector not in Span(A). Then $A \cup \{\vec{b}\} = \{\vec{a}_1, \ldots, \vec{a}_n, \vec{b}\}$ is independent.*

This is our main tool for building up a basis by finding ever larger independent sets. As an example suppose we want to extend the set $\{[2,1]\}$ to a basis for $\mathbb{R}^2$. Since $[2,1]$ is not the zero vector, this is an independent set. To enlarge it without losing independence we need to find another vector not in Span$(\{[2,1]\})$, that is, not of the form $[2c, c]$. We choose $[1,-1]$. The set $\{[2,1],[1,-1]\}$ will be independent. Since it has two elements and $\mathbb{R}^2$ has dimension 2, it should be a basis. Does it span $\mathbb{R}^2$?

Suppose that $\vec{b} = [b_1, b_2]$ is any vector in $\mathbb{R}^2$. Can we find c_1 and c_2 such that $c_1[2,1] + c_2[1,-1] = [b_1, b_2]$? This amounts to finding a solution to the system of equations

$$\begin{aligned} 2c_1 + c_2 &= b_1 \\ c_1 - c_2 &= b_2. \end{aligned}$$

It is a routine matter to see that $c_1 = \frac{1}{3}(b_1 + b_2)$ and $c_2 = \frac{1}{3}(b - 1 - 2b_2)$ is the unique solution. Thus B is a basis.

This was almost too simple: $\mathbb{R}^2$ isn't quite big enough; suppose we try $\mathbb{R}^3$. Again we begin with a single vector, say $[1,1,-1]$. We now want to find a vector not in Span$(\{[1,1,-1]\})$. The vector $[1,0,1]$ will do since it is not a multiple of $[1,1,-1]$. Thus the set $A = \{[1,1,-1],[1,0,1]\}$ is independent, but it has only two vectors and we know that the dimension of $\mathbb{R}^3$ is three. We must find a vector not in Span(A). The general form for a vector in Span(A) is $[c_1 + c_2, c_1, -c_1 + c_2]$. One way to find a vector which is not of this form is to notice that the difference of the first and third components is $2c_1$, twice the second component. The vector $[2,-1,2]$ does not have this property so it is not in Span(A) and $A \cup \{[2,-1,2]\}$ is independent. It remains to verify that $\{[1,-1,1],[1,0,1],[2,-1,2]\}$ spans $\mathbb{R}^3$. Since we have an independent

set of three vectors and we know that the dimension of the space is three, it seems like this step shouldn't be necessary. The next proposition shows that it isn't.

Proposition 8.3.2 *If $\mathcal{V}$ is of dimension n and $A = \{\vec{a}_1, \ldots, \vec{a}_n\}$ is a linearly independent set of n vectors, then A is a basis.*

PROOF:

> We need to show that A spans $\mathcal{V}$. Suppose that it doesn't. Then there is a vector $\vec{v}$ in $\mathcal{V}$ which is not in Span(A). This tells us that $A \cup \{\vec{v}\}$ is a linearly independent set with n+1 elements. Since $\mathcal{V}$ has dimension n it has a basis, and hence a spanning set, with n elements. But any linearly independent set has no more elements than any spanning set, so $n + 1 \leq n$ in our case. This is false, so our assumption that A didn't span $\mathcal{V}$ must have been wrong. ∎

Using this proposition we can formalize the procedure we used in the two examples to describe how to extend any linearly independent set to a basis.

Theorem 8.3.3 *If A is a linearly independent set of vectors in a finite dimensional vector space $\mathcal{V}$, then there is a basis B with $A \subseteq B$.*

PROOF:

> If A is not a basis then there is a vector in $\mathcal{V}$ which is not in Span(A). Adjoining it to A gives us a linearly independent set. We can continue this process until our linearly independent set has dim($\mathcal{V}$) vectors. We then have a basis. ∎

Another approach to building a basis is to start with a spanning set and then remove vectors until we get a basis. As an example of this process consider the set

$$C = \{[1, 2, 3], [1, 0, 1], [2, 2, 4], [0, 1, 1], [0, 2, 3]\}$$

in $\mathbb{R}^3$. To see that this is a spanning set for $\mathbb{R}^3$ notice that $[1, 0, 0] = [1, 2, 3] - [0, 2, 3]$, $[0, 1, 0] = 3[0, 1, 1] - [0, 2, 3]$, and $[0, 0, 1] = [0, 2, 3] - 2[0, 1, 1]$. The set C, however, has five vectors in it and the dimension of the space is three. We

need to remove two vectors. By Proposition 8.1.3, a set is dependent if and only if some vector can be written as a linear combination of the previous vectors. We check in succession: $[1,0,1]$ is not a multiple of $[1,2,3]$ so we don't need to remove it;

$$[2,2,4] = [1,2,3] + [1,0,1]$$

so we can remove $[2,2,4]$ and not lose the spanning property. Next we see if $[0,1,1]$ is a linear combination of $[1,2,3]$ and $[1,0,1]$. It is since

$$[0,1,1] = \frac{1}{2}([1,2,3] - [1,0,1]).$$

This leaves us with the spanning set

$$\{[1,2,3],[1,0,1],[0,2,3]\}.$$

If we then look to see if the system

$$x[1,2,3] + y[1,0,1] + z[0,2,3] = [0,0,0]$$

has any nontrivial solutions. The system becomes

$$\begin{aligned} x + y &= 0 \\ 2x + 2z &= 0 \\ 3x + y + 3z &= 0. \end{aligned}$$

From the first two equations we conclude that $x = -y = -z$. Putting this into the third we see that all three variables must be 0. Again it seems like we shouldn't have had to go to the trouble to show that we had a linearly independent set, since we had a spanning set with $\dim(\mathcal{V})$ vectors.

Proposition 8.3.4 *If $\mathcal{V}$ has dimension n and A is a spanning set with n vectors, then A is a basis.*

PROOF:

> We need to show that A is linearly independent. Suppose not, then there is an element $\vec{a}$ of A which can be removed leaving a spanning set with $n-1$ elements. Now a basis is a linearly independent set with n elements, and any spanning set has no fewer elements than any linearly independent set. This tells us that $n-1 > n$, which is false, so A must have been linearly independent. ∎

This tells us what we need to do to cut a spanning set down to a basis: remove elements which can be written as linear combinations of earlier elements until we have $\dim(\mathcal{V})$ elements, at which point we are done.

Theorem 8.3.5 *Any finite spanning set S in a finite dimensional vector space $\mathcal{V}$ contains a basis.*

Proof:

> Suppose our spanning set S has m elements. Then $m \geq \dim(\mathcal{V})$ since any spanning set has at least as many elements as any linearly independent set. If $m = \dim(\mathcal{V})$ then we are done by the previous proposition. If $m > \dim(\mathcal{V})$ then S must be dependent, so one of its elements is a linear combination of the others and can be removed to give a spanning set with $m - 1$ members. Continuing in this fashion we will eventually obtain a spanning set contained in S which has $\dim(\mathcal{V})$ members. That subset must be a basis. ∎

A word of caution: these propositions require the hypothesis that $\dim(\mathcal{V}) = n$. To show that $\dim(\mathcal{V}) = n$ we need to exhibit a basis with n vectors.

Combining these theorems with some observations about row reduction will give us a prescription for how to find bases for vector spaces of the form $\mathbb{R}^n$: to cut a finite spanning set S of vectors down to a linearly independent set of vectors, first make a matrix using the set S as columns. Then reduce the matrix to row reduced echelon form. This does not change the independence of sets of columns, so the vectors in S which were put into columns which contain the leading 1's in each row will form a linearly independent set. Since the S spans $\mathbb{R}^n$, there will be no rows of 0's in the row reduced echelon form, thus there will be n rows with leading 1's. Thus the independent set formed by the subset of S consisting of vectors in the columns of the leading 1's will form a basis.

Example: Using row reduction to reduce a spanning set to a basis

> We attempt to find a basis in the set
>
> $$\{[1, 2, 3], [1, 1, 4], [2, 3, 7], [-1, 1, 0], [0, 1, 2]\} \subset \mathbb{R}^3.$$

First form a matrix using these vectors as columns:

$$\begin{bmatrix} 1 & 1 & 2 & -1 & 0 \\ 2 & 1 & 3 & 1 & 1 \\ 3 & 4 & 7 & 0 & 2 \end{bmatrix}$$

then row reduce to get

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & \frac{1}{2} \end{bmatrix}.$$

The leading 1's are in columns 1,2,and 4 of the row reduced echelon form, so vectors 1,2,and 4 form the basis. Thus the basis is $\{[1,2,3],[1,1,4],[-1,1,0]\}$. ♢

If you start with a linearly independent set L you can extend to a basis by using L for the starting columns and adding in a known basis for the last n columns. Then cut down to a basis. Since this approach always builds the linearly independent starting from the left, the basis will include the linearly independent set we started with.

Example: Extend $\{[1,2,4,7],[1,-2,1,-3]\}$ to a basis for $\mathbb{R}^4$.

We form the matrix

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 2 & -2 & 0 & 1 & 0 & 0 \\ 4 & 1 & 0 & 0 & 1 & 0 \\ 7 & -3 & 0 & 0 & 0 & 1 \end{bmatrix}$$

then row reduce to get

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \frac{3}{19} & \frac{1}{19} \\ 0 & 1 & 0 & 0 & \frac{7}{19} & -\frac{4}{19} \\ 0 & 0 & 1 & 0 & -\frac{10}{19} & \frac{3}{19} \\ 0 & 0 & 0 & 1 & \frac{8}{19} & -\frac{10}{19} \end{bmatrix},$$

so the basis is in the first four columns of the original matrix. Thus the basis is $\{[1,2,4,7],[1,-2,1,-3],[1,0,0,0],[0,1,0,0]\}$.♢

We can use these theorems to give a prescription for finding the basis for the sum of two subspaces: to find a basis for $\mathcal{U} + \mathcal{W}$ take the union of a basis for $\mathcal{U}$ and a basis for $\mathcal{W}$ and cut it down to a linearly independent set. Finding a basis for the intersection of two subspaces is more difficult, though the next theorem tells us how many basis vectors we need.

Theorem 8.3.6 *If $\mathcal{U}$ and $\mathcal{W}$ are subspaces of a finite dimensional space $\mathcal{V}$ then*

$$\dim(\mathcal{U} + \mathcal{W}) = \dim(\mathcal{U}) + \dim(\mathcal{W}) - \dim(\mathcal{U} \cap \mathcal{W}).$$

PROOF:

Start by finding a basis $\{\vec{b}_1, \ldots, \vec{b}_{\dim(\mathcal{U}\cap\mathcal{W})}\}$ for $\mathcal{U} \cap \mathcal{W}$. Next add vectors $\vec{c}_1, \ldots, \vec{c}_{\dim(\mathcal{U})-\dim(\mathcal{U}\cap\mathcal{W})}$ to get a basis for $\mathcal{U}$. Similarly we can add vectors $\vec{d}_1, \ldots, \vec{d}_{\dim(\mathcal{W})-\dim(\mathcal{U}\cap\mathcal{W})}$ to get a basis for $\mathcal{W}$. Notice that the set

$$B = \{\vec{b}_1, \ldots, \vec{b}_{\dim(\mathcal{U}\cap\mathcal{W})}, \vec{c}_1, \ldots, \vec{c}_{\dim(\mathcal{U})-\dim(\mathcal{U}\cap\mathcal{W})}, \vec{d}_1, \ldots \vec{d}_{\dim(\mathcal{W})-\dim(\mathcal{U}\cap\mathcal{W})}\}$$

has $\dim(\mathcal{U}) + \dim(\mathcal{W}) - \dim(\mathcal{U} \cap \mathcal{W})$ members. We need only show that it is a basis for $\mathcal{U} + \mathcal{W}$.

It is clear that B spans $\mathcal{U} + \mathcal{W}$: given $\vec{u} + \vec{w}$ we can write $\vec{u}$ as a linear combination of the $\vec{b}$'s and $\vec{c}$'s and $\vec{w}$ as a linear combination of the $\vec{b}$'s and $\vec{d}$'s. Adding these linear combinations gives $\vec{u} + \vec{w}$ as a linear combination of elements of B.

Independence of B is also clear: since

$$\{\vec{b}_1, \ldots, \vec{b}_{\dim(\mathcal{U}\cap\mathcal{W})}, \vec{c}_1, \ldots, \vec{c}_{\dim(\mathcal{U})-\dim(\mathcal{U}\cap\mathcal{W})}\}$$

is independent,

$$\{\vec{d}_1, \ldots \vec{d}_{\dim \mathcal{W}-\dim(\mathcal{U}\cap\mathcal{W})}\}$$

is independent, and none of the $\vec{d}$'s are in $\mathcal{U}$. ∎

Example: Finding a basis for the sum of two subspaces

Let $\mathcal{V} = \mathbb{R}[x]_4$, let $\mathcal{U}$ be the subspace of polynomials whose value at 3 is 0, and let $\mathcal{W}$ be the subspace of polynomials whose coefficients for even powers of x are all 0. The set $\{x^4 - 81, x^3 -$

$27, x^2-9, x-3\}$ is a basis for $\mathcal{U}$ and $\{x^3, x\}$ is a basis for $\mathcal{W}$. We obtain a spanning set for $\mathcal{U}+\mathcal{W}$ by taking the union of these bases. However, the set $\{x^4-81, x^3-27, x^2-9, x-3, x^3, x\}$ is dependent ($x = 1(x-3) + \frac{1}{9}(x^3-27) - \frac{1}{9}x^3$), so we do not yet have a basis. If we remove x we *do* get an independent set. So $\dim(\mathcal{U}+\mathcal{W}) = 5$.

Now from the bases we have found we can see that $\dim(\mathcal{U}) = 4$ and $\dim(\mathcal{W}) = 2$, so the theorem tells us that

$$\begin{aligned}\dim(\mathcal{U}\cap\mathcal{W}) &= \dim(\mathcal{U}) + \dim(\mathcal{W}) - \dim(\mathcal{U}+\mathcal{W}) \\ &= 4+2-5 \\ &= 1\end{aligned}$$

Thus to find a basis for $\mathcal{U}\cap\mathcal{W}$ we need only find one non-zero element. The polynomial x^3-9x will do. ◊

One more observation is useful. For finite dimensional vector spaces over a field the dimension really tells the whole story. The following theorem shows why.

Theorem 8.3.7 *If $\mathcal{V}$ and $\mathcal{W}$ are vector spaces of dimension n over a field F then $\mathcal{V}$ and $\mathcal{W}$ are isomorphic; that is, there is a linear transformation from $\mathcal{V}$ to $\mathcal{W}$ which has a linear inverse.*

PROOF:

Pick ordered bases for both spaces. Define a linear transformation by taking the i^{th} basis vector in $\mathcal{V}$ to the i^{th} basis vector in $\mathcal{W}$. Since any vector in $\mathcal{V}$ is a linear combination of basis vectors this defines the linear transformation uniquely. The inverse is the linear transformation which takes the i^{th} basis vector of $\mathcal{W}$ to the i^{th} basis vector of $\mathcal{V}$. ∎

In our discussion of bases we have limited ourselves to finite dimensional vector spaces for the most part. Indeed, there is an interesting question left hanging: does *every* vector space have a basis? The answer is yes, provided your foundations of mathematics allow use of the axiom of choice. (See Thomas J. Jech, *The Axiom of Choice*, North Holland, 1973, p. 12) This is the first place in most student's careers that the axiom of choice actually matters. The proof is developed in problem 22 below.

Exercises 8.3:

1. Extend $\{[3, -1]\}$ to a basis for $\mathbb{R}^2$.
2. Extend $\{[1, 2, -1]\}$ to a basis for $\mathbb{R}^3$.
3. Extend $\{x - 1\}$ to a basis for $\mathbb{R}[x]_2$.
4. Extend $\{x - 1, x^2 + 1\}$ to a basis for $\mathbb{R}[x]_3$.
5. Extend $\{[1, 1, -1, 1], [2, -1, 0, 1]\}$ to a basis for $\mathbb{R}^4$.
6. Cut the spanning set $\{[1, 2, 3], [4, 5, 6], [7, 8, 9], [10, 11, 12]\}$ down to a basis for $\mathbb{R}^3$.
7. Cut the spanning set $\{[1, -2, 3], [-4, 5, 6], [7, 8, -9], [10, -11, 12]\}$ down to a basis for $\mathbb{R}^3$.
8. Cut the spanning set
$$\{[1, 0, 4, 0], [-2, 1, -8, 0], [-5, 3, -20, 0], [-4, 0, 1, 2], [-10, 1, -6, 4], [-2, 0, 0, 1]\}$$
down to a basis for $\mathbb{R}^4$.
9. Cut the spanning set $\{1, x + 1, 2x + 3, x^2 + 1, x^3 + x^2, x^3 + 2\}$ down to a basis for $\mathbb{R}[x]_3$.
10. Cut the spanning set $\{x^3 + x^2 + x + 1, x^3 + 2x^2 + x, x - 1, x^2 + x + 1, x^3 + x + 1, 2x^3 + x^2 - 2\}$ down to a basis for $\mathbb{R}[x]_3$.
11. Give bases for $\mathcal{U} + \mathcal{W}$ and $\mathcal{U} \cap \mathcal{W}$: $\mathcal{V} = \mathbb{R}^3$,
$\mathcal{U} = \text{Span}(\{[1, 1, 1], [0, 1, 2]\})$,
$\mathcal{W} = \text{Span}(\{[1, 2, 3], [1, 0, 1]\})$
12. Give bases for $\mathcal{U} + \mathcal{W}$ and $\mathcal{U} \cap \mathcal{W}$: $\mathcal{V} = \mathbb{R}^5$,
$\mathcal{U} = \text{Span}(\{[1, 1, 1, 1, 1], [0, 1, 2, 0, 1], [1, 2, 3, 1, 2]\})$,
$$\mathcal{W} = \text{the row space of } \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

13. Give bases for $\mathcal{U}+\mathcal{W}$ and $\mathcal{U}\cap\mathcal{W}$: $\mathcal{V}=\mathbb{R}^5$,
$\mathcal{U}=\text{Span}(\{[1,1,1,1,1],[0,1,2,0,1],[1,2,3,1,2]\})$,
$\mathcal{W}=$ the solution space of

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

14. Prove that a basis is a minimal spanning set.

15. Prove that a basis is a maximal linearly independent set.

16. Prove: if $\mathcal{W}<\mathcal{V}$ and $\mathcal{V}$ is finite dimensional then $\dim(\mathcal{W})\leq\dim(\mathcal{V})$. If equality holds then $\mathcal{W}=\mathcal{V}$.

17. Give examples to show that if $\mathcal{W}_1$ and $\mathcal{W}_2$ are both subspaces of a vector space $\mathcal{V}$ of dimension n with $\dim(\mathcal{W}_1)=\dim(\mathcal{W}_2)$ then $\mathcal{W}_1$ need not equal $\mathcal{W}_2$.

18. Show that the row space of $\mathbf{M}$ is the same as the row space of the row-reduced echelon form of $\mathbf{M}$.

19. Show that the non-zero rows of the row-reduced echelon form of $\mathbf{M}$ form a basis for the row space.

20. Use the previous two problems to find bases for the row spaces of the following matrices:

    (a) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

    (b) $\begin{bmatrix} 3 & 1 & 4 \\ -1 & 1 & 1 \\ 4 & 4 & 10 \end{bmatrix}$

    (c) $\begin{bmatrix} 1 & 4 & 3 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 4 & 5 \end{bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \\ 6 & 2 & 3 \end{bmatrix}$

21. Use row reduction to find a basis for the space spanned by the set

$$S = \{[2,4,6,1],[3,1,1,0],[1,0,1,1],[0,-1,-6,-2],[1,-1,3,4]\}$$

22. (Project Problem) A proof that every vector space has a basis: Zorn's Lemma states that if S is a nonempty family of sets in which every chain $S_1 \subseteq S_2 \subseteq \ldots$ has an upper bound in S, then S has a maximal element. Zorn's Lemma is equivalent to the axiom of choice. Let S be the set of linearly independent subsets of $\mathcal{V}$.

    (a) Show that S is always nonempty, no matter what $\mathcal{V}$ is.

    (b) Show that any ascending chain $S_1 \subseteq S_2 \subseteq \ldots$ has an upper bound.

    (c) Apply Zorn's lemma. Why does this give you a basis for $\mathcal{V}$?

Chapter 9

Matrices for Linear Transformations

In Chapter 8 we saw that the use of a basis made it possible to represent all vectors in a vector space using only a few vectors. If we can show that a linear transformation is completely determined by its value on basis vectors we will have reduced considerably the amount of work needed to specify a linear transformation. We can record the information in a matrix. This chapter will show how to find the matrix for a linear transformation with respect to choices of ordered basis for the domain and codomain. We will see that operations on linear transformations carry over exactly to operations on matrices

9.1 Using Bases to Get Matrices

When you learned how to differentiate polynomials your first step was to learn that the derivative of x^n is nx^{n-1}. Then you used the linearity of the differentiation operator and this fact to find the derivative of finite linear combinations of x^n's. Now, $\{x^n | n \in \mathbb{Z}\}$ is a basis for the space of polynomials, so what you did was find out how the differentiation operator behaved on a basis and then use linearity to extend the operator to the whole space.

The example of differentiation uses an infinite dimensional vector space. For finite dimensional vector spaces this same property of linear transformations will allow us to represent linear transformations very concretely using matrices.

Theorem 9.1.1 *If $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation and if $\{\vec{b}_i | i \in I\}$ is a basis for $\mathcal{V}$, and if $L(\vec{b}_i)$ is known for each $\vec{b}_i$, then $L(\vec{v})$ can be calculated for any $\vec{v} \in \mathcal{V}$.*

This is often stated as "a linear transformation is completely determined by its action on a basis."

PROOF:

We need to find the value $L(\vec{v})$. To do so we use the fact that $\{\vec{b}_i | i \in I\}$ is a basis for $\mathcal{V}$ to write $\vec{v}$ as a linear combination of basis elements:

$$\vec{v} = \sum_{i=1}^{n} v_i \vec{b}_i.$$

Since L is linear

$$L(\sum_{i=1}^{n} v_i \vec{b}_i) = \sum_{i=1}^{n} L(v_i \vec{b}_i).$$

Now L preserves scalar multiplication so

$$L(\vec{v}) = \sum_{i=1}^{n} L(v_i \vec{b}_i) = \sum_{i=1}^{n} v_i L(\vec{b}_i).$$

Since we know the values of $L(\vec{b}_i)$ this gives us the value of $L(\vec{v})$.

∎

Example: A linear transformation on $\mathbb{R}^2$

Suppose $L : \mathbb{R}^2 \to \mathbb{R}^2$ is linear and $L[1,0] = [2,-1]$ and $L[0,1] = [3,6]$. To find $L[x,y]$ we note that $[x,y] = x[1,0] + y[0,1]$ so

$$\begin{aligned} L[x,y] &= xL[1,0] + yL[0,1] \\ &= x[2,-1] + y[3,6] \\ &= [2x+3y, -x+6y]. \end{aligned}$$

◇

In the particular case where both $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional we can use this theorem to obtain a particularly convenient representation for L. Start by choosing an ordered basis $B = (\vec{b}_1, \ldots, \vec{b}_n)$ for $\mathcal{V}$ (which thus has dimension n) and an ordered basis $C = (\vec{c}_1, \ldots, \vec{c}_m)$ for $\mathcal{W}$ (which thus has dimension m). We know that L is completely determined by the values $L(\vec{b}_i), i = 1 \ldots n$. Let us express these values in terms of the basis for $\mathcal{W}$.

$$\begin{aligned}
L(\vec{b}_1) &= l_{11}\vec{c}_1 + l_{21}\vec{c}_2 + \ldots + l_{m1}\vec{c}_m \\
L(\vec{b}_2) &= l_{12}\vec{c}_1 + l_{22}\vec{c}_2 + \ldots + l_{m2}\vec{c}_m \\
&\vdots \\
L(\vec{b}_n) &= l_{1n}\vec{c}_1 + l_{2n}\vec{c}_2 + \ldots + l_{mn}\vec{c}_m
\end{aligned}$$

This gives us a matrix $\mathbf{L} = [[l_{ij}]]$ which has as its j^{th} column the C-coordinates of the image of the j^{th} basis vector. The knowledge of which ordered bases we are using and the matrix $\mathbf{L}$ should allow us to calculate $L(\vec{v})$ for any $\vec{v}$.

Definition 9.1.1 *The matrix* $\mathbf{L}$ *for the linear transformation* $L : \mathcal{V} \to \mathcal{W}$ *with respect to ordered bases* $B = (\vec{b}_1, \ldots, \vec{b}_n)$ *for* $\mathcal{V}$ *and* $C = (\vec{c}_1, \ldots, \vec{c}_m)$ *for* $\mathcal{W}$ *has as its* j^{th} *column the* C*-coordinates of the image of the* j^{th} *basis vector in* B.

Example: A matrix using the standard bases

Suppose the linear map L takes $[x, y, z]$ to $[2x, y + z]$, then using the ordered bases $([1, 0, 0], [0, 1, 0], [0, 0, 1])$ for $\mathbb{R}^3$ and $([1, 0], [0, 1])$ for $\mathbb{R}^2$ (the standard bases), we need to find

$$\begin{aligned}
L([1, 0, 0]) = [2, 0] &= 2[1, 0] + 0[0, 1] \\
L([0, 1, 0]) = [0, 1] &= 0[1, 0] + 1[0, 1] \\
L([0, 0, 1]) = [0, 1] &= 0[1, 0] + 1[0, 1]
\end{aligned}$$

Thus the matrix $\mathbf{L}$ is

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Notice that the number of rows is the same as the dimension of the codomain $\mathbb{R}^2$ and the number of columns is the same as the dimension of the domain $\mathbb{R}^3$. Once we have the matrix for L we

should be able to use it to find the value of L at any vector in the domain. Suppose we want to find $L([2,3,4])$. The theorem says that we should write $[2,3,4]$ in terms of the ordered basis for $\mathbb{R}^3$: $[2,3,4] = 2[1,0,0]+3[0,1,0]+4[0,0,1]$. We then take the values of L on the basis vectors to get

$$\begin{aligned} L([2,3,4]) &= 2L([1,0,0]) + 3L([0,1,0]) + 4L([0,0,1]) \\ &= 2[2,0] + 3[0,1] + 4[0,1] \\ &= [4,7]. \end{aligned}$$

$\diamondsuit$

It is tedious to have to think this through each time so we define an operation, multiplication of an m by n matrix and an n by 1 column vector.

Definition 9.1.2 *The i^{th} entry of $\mathbf{M}\vec{x}^t$ (the ^{t} indicates transpose: use a column instead of a row) is defined to be*

$$\mathbf{M}\vec{x}^t = \sum_{j=1}^{n} m_{ij}x_j.$$

Using this definition we observe that

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

If we want $L([1,-3,2])$ we calculate

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

In the general case the matrix $\mathbf{L}$ for $L : \mathcal{V} \to \mathcal{W}$ with respect to ordered bases $B = (\vec{b}_1, \ldots, \vec{b}_n)$ for $\mathcal{V}$ and $C = (\vec{c}_1, \ldots, \vec{c}_m)$ for $\mathcal{W}$ is the $m \times n$ matrix whose j^{th} column is the C-coordinate representation of $L(\vec{b}_j)$, that is, the entry l_{ij} given by the coefficient of $\vec{c}_i$ in the representation of $L(\vec{b}_j)$ in terms of the basis C for $\mathcal{W}$. Thus

$$L(\vec{b}_j) = l_{1j}\vec{c}_1 + \ldots + l_{mj}\vec{c}_m.$$

To find the C-coordinate representation of $L(\vec{a})$ we represent $\vec{a}$ in terms of the basis B and calculate

$$\begin{aligned}
\vec{a} &= \sum_{i=1}^{n} a_i \vec{b}_i \\
L(\vec{a}) &= L(\sum_{i=1}^{n} a_i \vec{b}_i) \\
&= \sum_{i=1}^{n} L(a_i \vec{b}_i) \\
&= \sum_{i=1}^{n} a_i L(\vec{b}_i) \\
&= \sum_{i=1}^{n} a_i (\sum_{j=1}^{m} l_{ij} \vec{c}_j) \\
&= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} a_i l_{ij} \right) \vec{c}_j
\end{aligned}$$

The last step reverses the order of summation, noting that in adding a rectangular array of numbers one can either add across rows and then down or add down columns and then across. The result is precisely what we get if we calculate $L[a_1, \ldots, a_n]^t$ using the rule for multiplication of a matrix times a column vector. This is summarized in the following corollary.

Corollary 9.1.2 *If $L : \mathcal{V} \to \mathcal{W}$ has matrix $\mathbf{L}$ with respect to the ordered bases $B = (\vec{b}_i)_{i=1\ldots n}$ and $C = (\vec{c}_i)_{i=1\ldots m}$, then the C-coordinates of $L(\sum_{i=1}^{n} a_i \vec{b}_i)$ may be calculated by taking $\mathbf{L}[a_1, \ldots a_n]^t$.*

Example: A matrix for differentiation

Let $d : \mathbb{R}[x]_3 \to \mathbb{R}[x]_2$ be the differentiation operator for polynomials of degree at most 3. The natural ordered basis for $\mathbb{R}[x]_3$ is $(x^3, x^2, x, 1)$ and for $\mathbb{R}[x]_2$ we choose $(x^2, x, 1)$. The matrix for d will be a 3×4 matrix whose i, j entry is the coefficient of the i^{th} basis vector in the image of the j^{th} basis vector. As in the

previous example we calculate

$$\begin{aligned} d(x^3) &= 3x^2 + 0x + 0 \\ d(x^2) &= 0x^2 + 2x + 0 \\ d(x) &= 0x^2 + 0x + 1 \\ d(1) &= 0x^2 + 0x + 0 \end{aligned}$$

so that the matrix is

$$\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

To use this to calculate the derivative of $3x^3 + 2x^2 - x + 1$ we note that this polynomial is represented by the column vector $[3, 2, -1, 1]^t$ when we use the ordered basis $(x^3, x^2, x, 1)$ and then calculate

$$\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 9 \\ 4 \\ -1 \end{bmatrix}$$

to get $d(3x^3 + 2x^2 - x + 1) = 9x^2 + 4x - 1$. $\diamondsuit$

We adopted the artificial way of writing the sums in calculating the image of basis vectors so that the matrix representation would use multiplication on the left. This makes $L(\vec{a})$ represented by $\mathbf{L}[a_1, \ldots, a_n]^t$, a parallel notation. Since it is more common, we write our matrices on the left. Some texts use multiplication on the right in their representations. This transposes everything and makes matrix multiplications go in the same direction that we usually write our diagrams, thus it eliminates some of the artificiality. Analysts tend to use multiplication on the left (following the convention used for function composition in calculus); some algebraists (particularly those who work a lot with permutation groups) write the matrix on the right, so that matrix multiplication will follow their diagrams. Students should be careful when looking at other texts to be sure they know which convention is being used.

Example: Leslie matrices

Linear transformations are frequently used to model the growth of populations whose population dynamics vary with age. The population to be modeled is broken into age groups so that in one time period of the model each individual advances from one age group to the next. The population is then represented as a vector $[p_1(t), p_2(t), \ldots, p_n(t)]$ listing the population from the youngest age group to the oldest. The change in the population comes about through mortality and births. In each period some of the individuals do not survive to the next census, the proportion who do survive from age group i to age group $i+1$ is given as a survival rate s_i. Thus $p_{i+1}(t+1) = s_i p_i(t)$ for $i \geq 1$.

To get $p_1(t+1)$ we need to consider births. Since the birthrate also depends on age, we also get a rate b_i giving the number of live births per individual in the time period of an age class for individuals in age class i. Then

$$p_1(t+1) = \sum_{i=1}^{n} b_i p_i(t).$$

Since birth rates are appropriate only for females, models of this kind for human population usually model the female population. The dynamics of the population are then given by a linear transformation which has a matrix of the form

$$\begin{bmatrix} b_1 & b_2 & \ldots & b_{n-1} & b_n \\ s_1 & 0 & \ldots & 0 & 0 \\ 0 & s_2 & \ldots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & s_{n-1} & 0 \end{bmatrix}$$

with respect to the standard basis. As an example let us look at an age structured model of the female population of the U.S.

The U.S. Bureau of the Census provides massive amounts of data about the U.S. population in the *Statistical Abstract of the United States*, published every five years. Much of its data groups the population in five year age spreads. The data for females in 1992 for births and deaths comes from the 1995 edition; the other columns are calculated:

Age	deaths per 100,000	5 year survival rate	births per 1000	female birth rate for 5 years	1994 female population
0–4	889.0	.95555	0	0	9633
0-1	850				
1-4	39.0				
5–9	17.5	.999125	0	0	9201
10–14	17.5	.999125	1.4	.0035	9150
15–19	47.2	.99764	60.7	.15175	8580
20–24	47.2	.99764	114.6	.2865	9015
25–29	106.1	.994695	117.4	.2935	9558
30–34	106.1	.994695	80.2	.2005	11119
35–39	106.1	.994695	32.5	.08125	11040
40–44	106.1	.994695	5.9	.01475	9970
45–49	558.8	.97206	0.3	.00075	8498
From Tables	120 and 127		89		16

The survival rate was obtained by taking

$$\frac{100000 - 5(\text{deaths per } 100000)}{100000}$$

and the female birth rate for 5 years used .0025(births per 1000) to give female births per individual for a 5 year period. This makes the somewhat inaccurate assumption that half the births were female (the actual figures for 1992 were 16.9 male babies to 15.2 females).

This gives the following Leslie matrix $\mathbf{L}$ for the dynamics of the U.S. female population:

$$\begin{bmatrix} 0 & 0 & .0035 & .15175 & .2865 & .2935 & .2005 & .08125 & .01475 & .00075 \\ .95555 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .999125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .999125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .99764 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .99764 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .994695 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .994695 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .994695 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .994695 & 0 \end{bmatrix}$$

To predict the population in 1999 using these dynamics we form $\mathbf{L}\vec{P}$, where $\vec{P}$ is the column vector giving the 1994 population. This gives

$$\mathbf{L}\vec{P} = \begin{bmatrix} 10002 \\ 9205 \\ 9193 \\ 9142 \\ 8560 \\ 8994 \\ 9507 \\ 11060 \\ 10981 \\ 9917 \end{bmatrix}$$

$\diamondsuit$

Exercises 9.1:

For problems 1–10 you are given a linear transformation, bases for the domain and codomain, and a vector $\vec{v}$; give the matrix for each with respect to the given bases and then use it to find the value of the linear transformation at $\vec{v}$:

1. $f : \mathbb{R}^2 \to \mathbb{R}^2$ with $f[x, y] = [x+y, x-y]$ where the basis for the domain is $([1, 0], [0, 1])$ and for the codomain $([1, 0], [0, 1])$ and $\vec{v} = [4, 7]$.

2. $f : \mathbb{R}^2 \to \mathbb{R}^2$ where $f[x, y] = [x+y, x-y]$ basis for domain: $([1, 0], [1, 1])$ basis for codomain: $([1, 0], [0, 1])$ and $\vec{v} = [4, 7]$.

3. $f : \mathbb{R}^2 \to \mathbb{R}^2$ where $f[x, y] = [x+y, x-y]$ basis for domain: $([1, 0], [1, 1])$ basis for codomain: $([1, 0], [1, 1])$ and $\vec{v} = [4, 7]$.

4. $L : \mathbb{R}^4 \to \mathbb{R}^3$ where $L[x, y, z, w] = [x, y, z + 3w]$
basis for domain: $([1, 0, 0, 0], [0, 1, 0, 0], [0, 0, 1, 0], [0, 0, 0, 1])$
basis for codomain: $([1, 0, 0], [0, 1, 0], [0, 0, 1])$ and $\vec{v} = [2, 1, -3, 4]$

5. $L : \mathbb{R}^4 \to \mathbb{R}^3$ where $L[x, y, z, w] = [x, y, z + 3w]$
basis for domain: $([1, 0, 0, 0], [1, 1, 0, 0], [1, 2, 3, 0], [0, 0, 0, 1])$
basis for codomain: $([1, 0, 0], [1, 1, 0], [0, 1, 1])$ and $\vec{v} = [2, 1, -3, 4]$

6. $L : \mathbb{R}^4 \to \mathbb{R}^3$ where $L[x, y, z, w] = [x, y, z + 3w]$
basis for domain: $([1, 1, 1, 0], [1, 1, 0, 1], [1, 0, 1, 1], [0, 1, 1, 1])$
basis for codomain: $([1, 1, 0], [1, 0, 1], [0, 1, 1])$ and $\vec{v} = [2, 1, -3, 4]$

7. $L : \mathbb{R}^3 \to \mathbb{R}^4$ where $L[x, y, z] = [x + 2z, y - z, x + y + z, x - y - z]$
basis for domain: $([1, 0, 0], [0, 1, 0], [0, 0, 1])$
basis for codomain: $([1, 0, 0, 0], [0, 1, 0, 0], [0, 0, 1, 0], [0, 0, 0, 1])$ and $\vec{v} = [1, 2, 3]$

8. $L : \mathbb{R}^3 \to \mathbb{R}^4$ where $L[x, y, z] = [x + 2z, y - z, x + y + z, x - y - z]$
basis for domain: $([1, 1, 1], [1, 1, 0], [1, 0, 1])$
basis for codomain: $([1, 0, 0, 0], [0, 1, 0, 0], [0, 1, 1, 0], [0, 0, 0, 2])$ and $\vec{v} = [1, 2, 3]$

9. $L : \mathbb{R}^3 \to \mathbb{R}^4$ where $L[x, y, z] = [x + 2z, y - z, x + y + z, x - y - z]$
basis for domain: $([1, 2, 1], [1, 1, 2], [1, 2, 1])$
basis for codomain: $([1, 0, -1, 0], [0, 1, 0, -1], [0, -1, 1, 0], [-1, 0, 0, 2])$ and $\vec{v} = [1, 2, 3]$

10. $s : \mathbb{R}[x]_3 \to \mathbb{R}[x]_3$ with $s(p(x)) = p(x+3)$ basis for domain: $(x^3, x^2, x, 1)$
basis for codomain: $(x^3, x^2, x, 1)$ and $\vec{v} = x^3 + 3x^2 - 2x + 1$

For problems 11–13, suppose that T is a linear transformation:

11. $T : \mathbb{R}^3 \to \mathbb{R}^2$, $T[1, 0, 0] = [3, 2]$, $T[0, 1, 0] = [1, 4]$, $T[0, 0, 1] = [0, 0]$. What is $T[1, 2, 3]$?

12. $T : \mathbb{R} \to \mathbb{R}^3$, $T[1] = [6, -1, 4]$. What is $T[s]$?

13. $T : \mathbb{R}^4 \to \mathbb{R}^2$, $T([1, 1, 1, 1]) = [1, 2]$, $T[0, 1, 1, 1] = [2, 2]$, $T[0, 0, 1, 1] = [3, 6]$, $T[0, 0, 0, 1] = [0, -3]$; what is $T[1, 2, 3, 4]$?

14. The maps $L : \mathbb{R}^2 \to \mathbb{R}^2$ taking $[x, y]$ to $[3x, x - 2y]$ and $S : \mathbb{R}^2 \to \mathbb{R}^3$ taking $[x, y]$ to $[x, x + y, 2x + 2y]$ are linear.

    (a) Write down the matrices for L and S with respect to the ordered bases $([1, 0], [0, 1])$ for $\mathbb{R}^2$ and $([1, 0, 0], [0, 1, 0], [0, 0, 1])$ for $\mathbb{R}^3$.

    (b) $SL : \mathbb{R}^2 \to \mathbb{R}^3$ is a map found by first doing L and then doing S. Check that it is linear.

    (c) Write down the matrix for SL. Can you see how it is related to the matrices for S and L?

15. (Project Problem) Using the data in the example on Leslie matrices

    (a) What is the predicted age structured population of in 2004?

    (b) What is happening to the total population of women in this model over the years 1994–2014?

    (c) The bulge in the population in age groups 30–39 in 1994 is the the remnant of the post war baby boom. How does the distribution of ages change as this bulge works its way through the population?

    (d) How could you improve the accuracy of the predictions made by this model of the U.S. population ?

9.2 Operations on Matrices and Linear Transformations

In the last section we saw how choices of ordered bases for the domain and codomain of a linear transformation allowed us to represent it by a matrix. That representation depended on knowing that the action of a linear transformation on a basis is all that we need to know to calculate its action on any vector. We found that we could make the description more concrete by defining the multiplication of a matrix times a column vector, an operation which then became the prototype for linear transformations. In this section and the next we will look at other operations involving linear transformations and operations on matrices. Again the choice of ordered bases will give us the means to relate linear transformations and matrices. This makes it possible to prove theorems about matrices easily by proving theorems about linear transformations and makes it possible to manipulate linear transformations concretely in terms of matrices.

The first thing that we need to do is notice that we can add linear transformations and multiply by scalars and get new linear transformations.

Theorem 9.2.1 *For any vector spaces $\mathcal{V}$ and $\mathcal{W}$, the set of all linear transformations from $\mathcal{V}$ to $\mathcal{W}$, written $Hom(\mathcal{V}, \mathcal{W})$, is a vector space.*

PROOF:

First we note that the set of all functions from $\mathcal{V}$ to $\mathcal{W}$ is a vector space with $(F+G)(\vec{v}) = F(\vec{v}) + G(\vec{v})$ and $(kF)(\vec{v}) = k(F(\vec{v}))$

giving the operations and all of the axioms for a vector space following directly from the axioms for $\mathcal{W}$. We will show that the set of linear transformations forms a *subspace* of the space of all functions from $\mathcal{V}$ to $\mathcal{W}$. To do so we need only show that $\mathrm{Hom}(\mathcal{V}, \mathcal{W})$ is not empty and that it is closed under the operations addition and multiplication by a scalar (that is, that the sum of two linear transformations is a linear transformation and that a constant times a linear transformation is a linear transformation). That $\mathrm{Hom}(\mathcal{V}, \mathcal{W})$ is nonempty is easy: the constant map with value $\vec{0}$ is linear, so it is in $\mathrm{Hom}(\mathcal{V}, \mathcal{W})$. To see that the sum of linear maps is again linear let $F : \mathcal{V} \to \mathcal{W}$ and $G : \mathcal{V} \to \mathcal{W}$ be linear. Then

$$\begin{aligned}(F+G)(\vec{v}_1+\vec{v}_2) &= F(\vec{v}_1+\vec{v}_2)+G(\vec{v}_1+\vec{v}_2)\\ &= F(\vec{v}_1)+F(\vec{v}_2)+G(\vec{v}_1)+G(\vec{v}_2)\end{aligned}$$

by the linearity of F and G. But this in turn equals $F(\vec{v}_1) + G(\vec{v}_1) + F(\vec{v}_2) + G(\vec{v}_2)$ by commutativity. Thus

$$(F+G)(\vec{v}_1+\vec{v}_2) = (F+G)(\vec{v}_1)+(F+G)(\vec{v}_2).$$

Similarly

$$(F+G)(k\vec{v}) = F(k\vec{v})+G(k\vec{v}) = kF(\vec{v})+kG(\vec{v}) = k(F+G)(\vec{v}).$$

A similar pair of calculations shows that if r is a scalar and $F : \mathcal{V} \to \mathcal{W}$ is linear then so is rF :

$$(rF)(\vec{v}1+\vec{v}2) = r(F(\vec{v}1+\vec{v}2)) = r(F(\vec{v}1)+F(\vec{v}2)) = rF(\vec{v}1)+rF(\vec{v}2)$$

and

$$(rF)(k\vec{v}) = r(F(k\vec{v})) = r(kF(\vec{v})) = (rk)F(r) = k(rF)(\vec{v}).$$

∎

In the last section we saw how to establish a correspondence between matrices and linear transformations by making use of ordered bases. We now want to show that this correspondence respects the vector space operations. Recall that matrices form a vector space using addition and scalar multiplication componentwise.

Theorem 9.2.2 *If $F : \mathcal{V} \to \mathcal{W}$ and $G : \mathcal{V} \to \mathcal{W}$ are linear transformations and $(\vec{b}_1 \ldots \vec{b}_n)$ and $(\vec{c}_1 \ldots \vec{c}_m)$ are ordered bases for $\mathcal{V}$ and $\mathcal{W}$, respectively, then the matrix corresponding to $F + G$ with respect to the given bases is the sum of the matrix corresponding to F and the matrix corresponding to G.*

PROOF:

We get the j^{th} column of the matrix $\mathbf{A}$ corresponding to $F+G$ by writing

$$(F + G)(\vec{b}_j) = \sum_{i=1}^{m} a_{ij}\vec{c}_i.$$

Now

$$F(\vec{b}_j) = \sum_{i=1}^{m} f_{ij}\vec{c}_i \text{ and } G(\vec{b}_j) = \sum_{i=1}^{m} g_{ij}\vec{c}_i$$

tell us how to get the i^{th} column of the matrices corresponding to F and G. Adding gives

$$\begin{aligned}(F + G)(\vec{b}_j) &= \sum_{i=1}^{m} f_{ij}\vec{c}_i + \sum_{i=1}^{m} g_{ij}\vec{c}_i \\ &= \sum_{i=1}^{m}(f_{ij} + g_{ij})\vec{c}_i.\end{aligned}$$

Now a vector can be written only one way in terms of a given basis so the fact that

$$\sum_{i=1}^{m} a_{ij}\vec{c}_i = \sum_{i=1}^{m}(f_{ij} + g_{ij})\vec{c}_i$$

tells us that $a_{ij} = f_{ij} + g_{ij}$. Thus the matrix for a sum is the sum of the matrices. ∎

Theorem 9.2.3 *If $F : \mathcal{V} \to \mathcal{W}$ is linear and $(\vec{b}_1 \ldots \vec{b}_n)$ and $(\vec{c}_1 \ldots \vec{c}_m)$ are ordered bases for $\mathcal{V}$ and $\mathcal{W}$ then the matrix for rF with respect to the given bases is r times the matrix for F with respect to the given bases.*

PROOF:

Again recall that the i^{th} column of the matrix $\mathbf{A}$ corresponding to rF is obtained by writing

$$rF(\vec{b}_i) = \sum_{i=1}^{m} a_{ij}\vec{c}_i.$$

But

$$\begin{aligned}(rF)(\vec{b}_i) &= r(F(\vec{b}_i)) \\ &= r\sum_{i=1}^{m} f_{ij}\vec{c}_j \\ &= \sum_{i=1}^{m} rf_{ij}\vec{c}_j.\end{aligned}$$

Thus $a_{ij} = rf_{ij}$. ∎

Linear transformations have another way that they can be combined: composition. For example if $f : \mathbb{R}^2 \to \mathbb{R}^3$ takes $[x, y]$ to $[x, x + y, y]$ and $s : \mathbb{R}^3 \to \mathbb{R}$ takes $[x, y, z]$ to $x + y - 2z$ then

$$\begin{aligned}(s \circ f)([x, y]) &= s(f[x, y]) \\ &= s[x, x + y, y] \\ &= [x + x + y - 2y] \\ &= 2x - y\end{aligned}$$

defines a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}$. The general case is given in the next proposition:

Proposition 9.2.4 *If $f : \mathcal{V} \to \mathcal{W}$ and $g : \mathcal{W} \to \mathcal{U}$ are linear transformations then so is $g \circ f : \mathcal{V} \to \mathcal{U}$.*

PROOF:

We need to show that

$$(g \circ f)(\vec{v}_1 + \vec{v}_2) = (g \circ f)(\vec{v}_1) + (g \circ f)(\vec{v}_2)$$

and that

$$(g \circ f)(k\vec{v}) = k(g \circ f)(\vec{v}).$$

Both are easy:

$$\begin{aligned}
(g \circ f)(\vec{v}_1 + \vec{v}_2) &= g(f(\vec{v}_1 + \vec{v}_2)) \\
&= g(f(\vec{v}_1) + f(\vec{v}_2)) \\
&= g(f(\vec{v}_1)) + g(f(\vec{v}_2)) \\
&= (g \circ f)(\vec{v}_1) + (g \circ f))(\vec{v}_2)
\end{aligned}$$

and

$$\begin{aligned}
(g \circ f)(k\vec{v}) &= g(f(k\vec{v})) \\
&= g(kf(\vec{v})) \\
&= kg(f(\vec{v})) \\
&= k(g \circ f)(\vec{v}).
\end{aligned}$$

∎

To find the matrix analogue to composition for linear transformations we need to recall some of the details of how we got the matrix for a linear transformation. Suppose $f : \mathcal{V} \to \mathcal{W}$ and $g : \mathcal{W} \to \mathcal{U}$ are linear transformations and $(\vec{b}_1 \dots \vec{b}_n)$, $(\vec{c}_1 \dots \vec{c}_m)$, and $(\vec{d}_1 \dots \vec{d}_p)$ are ordered bases for $\mathcal{V}$, $\mathcal{W}$ and $\mathcal{U}$, respectively. Then the matrix for f with respect to the bases $(\vec{b}_1 \dots \vec{b}_n)$ and $(\vec{c}_1 \dots \vec{c}_m)$ has m rows and n columns with each column given by finding

$$f(\vec{b}_j) = \sum_{i=1}^{m} F_{ij}(\vec{c}_i).$$

To find $f(\vec{v})$ we write $\vec{v}$ as a column vector using its representation in terms of the basis $(\vec{b}_1 \dots \vec{b}_n)$ and then multiply on the left by the matrix $\mathbf{F}$. This gives a hint how to define the operation on matrices corresponding to composition: simply multiply each column of the matrix corresponding to f by the matrix corresponding to g, written on the left.

Definition 9.2.1 *The product* $\mathbf{GF}$ *of an* $m \times n$ *matrix* $\mathbf{F}$ *and a* $p \times m$ *matrix* $\mathbf{G}$ *is the* $p \times n$ *matrix with* ij *entry given by*

$$\sum_{k=1}^{m} g_{ik} f_{kj}.$$

In order for multiplication of two matrices to be defined the number of columns of the first must equal the number of rows of the second.

$$\begin{array}{cccc} \mathbf{G} & \mathbf{F} & = & \mathbf{GF} \\ p \times m & m \times n & & p \times n \end{array}$$

The answer has the same number of rows as the first and the same number of columns as the second.

Example:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 2 & 3 \\ -3 & 4 & 7 \\ -5 & 6 & 11 \end{bmatrix}$$

$\diamondsuit$

Example:

$$\begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 2 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 11 & 3 \end{bmatrix}$$

$\diamondsuit$

Example:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 3 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 3 & 7 \end{bmatrix}$$

Notice that these two results are not equal: matrix multiplication is not commutative. $\diamondsuit$

We have carefully defined matrix multiplication so that the next theorem will be true.

Theorem 9.2.5 *If $f : \mathcal{V} \to \mathcal{W}$ and $g : \mathcal{W} \to \mathcal{U}$ are linear transformations, $(\vec{b}_1 \ldots \vec{b}_m)$ an ordered basis for $\mathcal{V}$, $(\vec{c}_1 \ldots \vec{c}_n)$ an ordered basis for $\mathcal{W}$, and $(\vec{d}_1 \ldots \vec{d}_p)$ an ordered basis for $\mathcal{U}$, then the matrix for $g \circ f$ with respect to the bases $(\vec{b}_1 \ldots \vec{b}_m)$ and $(\vec{d}_1 \ldots \vec{d}_p)$ is the product of the matrix for g with respect to the bases $(\vec{c}_1 \ldots \vec{c}_n)$ and $(\vec{d}_1 \ldots \vec{d}_p)$ and the matrix for f with respect to the bases $(\vec{b}_1 \ldots \vec{b}_m)$ and $(\vec{c}_1 \ldots \vec{c}_n)$.*

PROOF:

We need to calculate the value of $(g \circ f)\vec{b}_j$. We know that

$$f(\vec{b}_j) = \sum_{i=1}^{n} f_{ij}\vec{c}_i$$

by the way the matrix $\mathbf{F} = (f_{ij})$ for f with respect to the bases $(\vec{b}_1 \ldots \vec{b}_m)$ and $(\vec{c}_1 \ldots \vec{c}_n)$ is defined. To find $(g \circ f)(\vec{b}_j)$ we calculate

$$\begin{aligned}
(g \circ f)(\vec{b}_j) &= g(f(\vec{b}_j)) \\
&= g(\sum_{i=1}^{n} f_{ij}\vec{c}_i) \\
&= \sum_{i=1}^{n} f_{ij} g(\vec{c}_i) \\
&= \sum_{i=1}^{n} f_{ij} \left(\sum_{k=1}^{p} g_{ki}\vec{d}_k \right)
\end{aligned}$$

using linearity and the definition of the matrix for g. Continuing we get

$$\begin{aligned}
g \circ f(\vec{b}_i) &= \sum_{i=1}^{n} \left(\sum_{k=1}^{p} f_{ij} g_{ki} \vec{d}_k \right) \\
&= \sum_{k=1}^{p} \left(\sum_{i=1}^{n} f_{ij} g_{ki} \right) \vec{d}_k
\end{aligned}$$

So

$$(g \circ f)_{kj} = \sum_{i=1}^{n} f_{ij} g_{ki} = \sum_{i=1}^{n} g_{ki} f_{ij}$$

is the kj entry in the product $\mathbf{GF}$. ∎

After a bit of practice matrix multiplication is not too difficult an operation to remember and calculate. The definition, however, is very awkward to work with for proving properties of matrix multiplication. For instance:

Proposition 9.2.6 *For $m \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, and $p \times m$ matrix $\mathbf{C}$ we get*

$$\mathbf{C}(\mathbf{A}+\mathbf{B}) = \mathbf{CA}+\mathbf{CB},$$

the distributive law for matrix multiplication over matrix addition.

PROOF:

We calculate both sides:

$$\begin{aligned}
(\mathbf{C}(\mathbf{A}+\mathbf{B}))_{ij} &= \sum_{k=1}^{m} c_{ik}(a_{kj}+b_{kj}) \\
&= \sum_{k=1}^{m}(c_{ik}a_{kj}+c_{ik}b_{kj}) \\
&= \sum_{k=1}^{m} c_{ik}a_{kj} + \sum_{k=1}^{m} c_{ik}b_{kj} \\
&= (\mathbf{CA}+\mathbf{CB})_{ij}
\end{aligned}$$

∎

This isn't too bad, provided you are very comfortable with the manipulation of sums. Associativity, however, is horrendous:

Proposition 9.2.7 *For matrices $\mathbf{A}$ of size $m \times n$, $\mathbf{B}$ of size $n \times p$, and $\mathbf{C}$ of size $p \times q$ we get $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.*

PROOF:

We need to show

$$\mathbf{A}(\mathbf{BC})_{ij} = (\mathbf{AB})\mathbf{C}_{ij}$$

that is, that

$$\sum_{k=1}^{n} a_{ik}(\mathbf{BC})_{kj} = \sum_{k=1}^{n} a_{ik} \sum_{h=1}^{p} b_{kh}c_{hj}$$

is equal to

$$\sum_{h=1}^{p} (\mathbf{AB})_{ih} c_{hj} = \sum_{h=1}^{p} (\sum_{k=1}^{n} a_{ik} b_{kh}) c_{hj}.$$

This is a matter of applying generalized distributivity in the reals to get:

$$\begin{aligned} \sum_{k=1}^{n} a_{ik} (\mathbf{BC})_{kj} &= \sum_{k=1}^{n} a_{ik} \sum_{h=1}^{p} b_{kh} c_{hj} \\ &= \sum_{k=1}^{n} \sum_{h=1}^{p} a_{ik} (b_{kh} c_{hj}) \end{aligned}$$

then apply associativity:

$$= \sum_{k=1}^{n} \sum_{h=1}^{p} (a_{ik} b_{kh}) c_{hj}.$$

If we then reverse the order of summation we get

$$= \sum_{h=1}^{p} \sum_{k=1}^{n} ((a_{ik} b_{kh}) c_{hj}).$$

Again we apply generalized distributivity for the reals to get

$$= \sum_{h=1}^{p} (\sum_{k=1}^{n} a_{ik} b_{kh}) c_{hj},$$

as needed. ∎

This proof is correct, but rather ugly and unenlightening. A much more elegant proof is based on the fact that the correspondence between linear transformations and matrices works for composition too. Associativity of composition is easy to prove:

Proposition 9.2.8 *Composition of linear maps is associative: if $f : \mathcal{V} \to \mathcal{W}$, $g : \mathcal{W} \to \mathcal{U}$ and $h : \mathcal{U} \to \mathcal{X}$ are all linear then the linear transformations $h \circ (g \circ f)$ and $(h \circ g) \circ f$ are equal.*

PROOF:

Function composition is always associative. All we need to do is check that both ways of composing always give the same result:

$$\begin{aligned}(h \circ (g \circ f))(\vec{v}) &= h(g(f(\vec{v}))) \\ &= (h \circ g)(f(\vec{v})) \\ &= ((h \circ g) \circ f)(\vec{v}).\end{aligned}$$

∎

Corollary 9.2.9 *Matrix multiplication is associative.*

PROOF:

Matrix multiplication represents composition of linear transformations. Since composition is associative, matrix multiplication must also be. ∎

These theorems contain the core of linear algebra: Operations on matrices can be understood in terms of operations on the linear transformations they represent, and operations on linear transformations can be made concrete and calculated using matrices. Choice of bases for domain and codomain determines an isomorphism of algebras between linear transformations and matrices: sums and scalar multiples are preserved and reflected, so the vector space of linear transformations is isomorphic to the vector space of matrices and matrix multiplication captures the composition operation exactly as well.

Exercises 9.2:

1. Let $F : \mathbb{R}^2 \to \mathbb{R}^3$ take $[x, y]$ to $[x + y, y - x, 3y]$ and G take $[x, y]$ to $[-2y, 3x + y, 6x]$.
   (a) write the matrices for F and G with respect to the standard bases
   (b) write expressions for $(F + G)([x, y])$ and $3G([x, y])$.
   (c) check that the matrix for the sum is the same as the sum of the matrices for the two linear transformations and that the matrix for the scalar product is the same as 3 times the matrix for G.

2. Repeat exercise 1 using the linear transformations $F : \mathbb{R}^5 \to \mathbb{R}^2$ and G given by $F([v, w, x, y, z]) = [x+y+z, v-w+2x]$ and $G([v, w, x, y, z]) = [2w + y, 3x]$.

3. If
$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 2 \\ 0 & 1 & 1 & 4 \\ 4 & 1 & 6 & 7 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 1 \\ 0 & 4 \end{bmatrix} \quad \text{and } \mathbf{C} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 2 & 1 \\ 5 & 3 \end{bmatrix}$$
find $\mathbf{AB}$, $\mathbf{AC}$, $\mathbf{B} + \mathbf{C}$, $\mathbf{A}(\mathbf{B} + \mathbf{C})$ and $\mathbf{AB} + \mathbf{AC}$.

4. If
$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 4 & 2 \end{bmatrix} \quad \text{and } \mathbf{C} = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 1 & 0 & 1 & 2 \end{bmatrix}$$
find $\mathbf{AB}$, $\mathbf{BC}$, $\mathbf{A}(\mathbf{BC})$, and $(\mathbf{AB})\mathbf{C}$.

5. Quadratic forms in the variables x and y can be written using a matrix by taking
$$[x, y]M \begin{bmatrix} x \\ y \end{bmatrix}$$
See what you get using the matrices

   (a) $\mathbf{M} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

   (b) $\mathbf{M} = \begin{bmatrix} 1 & -1 \\ 0 & -1 \end{bmatrix}$

   (c) $\mathbf{M} = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}$

   A symmetric matrix is one there $a_{ij} = a_{ji}$ for all i, j. Find a symmetric matrix to represent the forms

   (d) $2x^2 - 4xy + y^2$

   (e) $2x^2 + 3y^2 + 4z^2 + 2xy - 5yz + 16xz$

6. The $n \times n$ square matrix $\mathbf{I}_n$ with 1's in the positions with both indices the same and 0's elsewhere is called an identity matrix. Prove that for any $n \times m$ matrix $\mathbf{A}$, $\mathbf{IA} = \mathbf{A}$ and that for any $m \times n$ matrix $\mathbf{B}$, $\mathbf{BI} = \mathbf{B}$.

7. Give examples to show that $\mathbf{AB}$ can be defined when $\mathbf{BA}$ is not and that even if both are defined they need not be of the same shape.

8. Find a matrix $\mathbf{A}$ which is not composed entirely of zeros but which has $\mathbf{A}^2$ all zeros.

9. Prove that composition distributes over sums for linear transformations: that is, if f is a linear transformation from $\mathcal{V}$ to $\mathcal{W}$ and g and h are linear transformations from $\mathcal{U}$ to $\mathcal{V}$ then $f \circ (g+h) = (f \circ g) + (f \circ h)$.

10. Use the previous exercise to give a conceptual proof of the distributive law for matrix multiplication over matrix addition.

Chapter 10

Inverses and Rank

10.1 Inverses of Linear Transformations and Matrices

Since a linear transformation $L : \mathcal{V} \to \mathcal{W}$ is a function, it makes sense to ask when it has an inverse under composition and what properties that inverse might have.

Definition 10.1.1 *A linear transformation $L : \mathcal{V} \to \mathcal{W}$ is called invertable if there is another linear transformation $L^{-1} : \mathcal{W} \to \mathcal{V}$ such that $L \circ L^{-1} = id_{\mathcal{W}}$ and $L^{-1} \circ L = id_{\mathcal{V}}$. In such a case L^{-1} is called the inverse of L.*

We know from our study of inverse functions in precalculus and calculus that a *function* f has an inverse if and only if it is one-to-one (whenever $f(x) = f(x')$ we have $x = x'$) and onto (the range of f is the same as the codomain of f). Our next theorem shows that the existence of an inverse as a function is sufficient to prove the existence of an inverse as a linear transformation:

Theorem 10.1.1 *If a linear transformation $L : \mathcal{V} \to \mathcal{W}$ has an inverse as a function, then that inverse is a linear transformation.*

PROOF:

We need to show that L^{-1} preserves scalar product and sums. Now $L^{-1}(r\vec{w})$ is the unique member of $\mathcal{V}$ which maps to $r\vec{w}$ under

L. If we let $\vec{v} = L^{-1}(\vec{w})$, then

$$\begin{aligned} L(r\vec{v}) &= L(rL^{-1}(\vec{w})) \\ &= rL(L^{-1}\vec{w}) \\ &= r\vec{w}. \end{aligned}$$

But this shows that $L^{-1}(r\vec{w})$ must have been $rL^{-1}(\vec{w})$. A similar argument works for sums and is left as an exercise. ∎

As usual, we want to see how the matrix with respect to given ordered bases for a linear transformation is related to the matrix with respect to the same ordered bases for its inverse. Not surprisingly the matrix for the inverse is the inverse (with respect to matrix multiplication) of the matrix for the linear transformation. Once we have an algorithm for finding the inverse of a matrix this will tell us how to find inverses for linear transformations. We need a definition and a theorem relating these concepts:

Definition 10.1.2 *The inverse of a square matrix* $\mathbf{M}$ *is a matrix* $\mathbf{M}^{-1}$ *such that* $\mathbf{M}\mathbf{M}^{-1} = \mathbf{M}^{-1}\mathbf{M} = I$*, an identity matrix. If an inverse for* $\mathbf{M}$ *exists we say* $\mathbf{M}$ *is invertable.*

Theorem 10.1.2 *If* $F : \mathcal{V} \to \mathcal{W}$ *has matrix* $\mathbf{F}$ *with respect to the ordered bases* $(\vec{b}_1, \ldots, \vec{b}_n)$ *for* $\mathcal{V}$ *and* $(\vec{c}_1, \ldots, \vec{c}_m)$ *for* $\mathcal{W}$*, then the matrix for* F^{-1} *with respect to the same ordered bases is* $\mathbf{F}^{-1}$*.*

PROOF:

First note that the matrix for the identity linear transformation $\mathrm{id}_{\mathcal{V}}$ with respect to the ordered basis $(\vec{b}_1, \ldots, \vec{b}_n)$ for both the domain and the codomain is the identity matrix, and similarly for $id_{\mathcal{W}}$. Since $F\ F^{-1} = \mathrm{id}_{\mathcal{W}}$ we get $\mathbf{F}\ \mathbf{G} = \mathbf{I}$ where $\mathbf{G}$ is the matrix for f^{-1}. Since $F^{-1}F = \mathrm{id}_{\mathcal{V}}$ we get $\mathbf{G}\mathbf{F} = \mathbf{I}$. Thus $\mathbf{G} = \mathbf{F}^{-1}$. ∎

A linear transformation $L : \mathcal{V} \to \mathcal{W}$ will be onto if the image is all of $\mathcal{W}$. The image of a spanning set under L will span $\mathrm{Im}(L)$, so if $\{\vec{b}_i | i \in I\}$ is a basis for $\mathcal{V}$, then $\{L(\vec{b}_i) | i \in I\}$ spans $\mathrm{Im}(L)$.

A linear transformation is one-to-one if and only if $\mathrm{Ker}(L) = \{\vec{0}\}$. If L is one-to-one it will preserve linear independence since if

$$\sum a_k L(\vec{v}_k) = \vec{0}$$

then

$$L(\sum a_k \vec{v}_k) = \vec{0}$$

so

$$\sum a_k \vec{v}_k \in \mathrm{Ker}(L)$$

and thus

$$(\sum a_k \vec{v}_k) = \vec{0}.$$

If the vectors $\vec{v}_i$ were independent this tells us that all of the $a_k = 0$.

Summarizing, we have shown

Proposition 10.1.3 *Any invertable linear transformation $L : \mathcal{V} \to \mathcal{W}$ takes a basis for $\mathcal{V}$ to a basis for $\mathcal{W}$. Thus if $\mathcal{V}$ is finite dimensional, so must $\mathcal{W}$ be and the dimensions must be the same.*

Corollary 10.1.4 *In order for a matrix to have an inverse it must be square.*

PROOF:

> An invertable matrix represents an invertable linear transformation with respect to ordered bases. The number of rows in the matrix is the same as to the dimension of the codomain of the linear transformation and the number of columns agrees with the dimension of the domain. In order for a linear transformation to have an inverse these two dimensions must agree. ∎

Corollary 10.1.5 *For an $n \times n$ matrix $\mathbf{M}$ to have an inverse it must have column space of dimension n and the only solution to $\mathbf{M}\vec{x}^t = \vec{0}^t$ is $\vec{0}^t$.*

PROOF:

> Let $\mathbf{M}$ represent a linear transformation L with respect to some choice of ordered bases. The first assertion corresponds to the linear transformation being onto. The second says that $\mathrm{Ker}(L) = \{\vec{0}\}$., so that L is one-to-one. ∎

Now then, how do we go about finding the inverse of a matrix? Since linear transformations take spanning sets to spanning sets, the dimension of the domain and codomain of a linear transformation must be equal in order for it to be possible for there to be an inverse. Thus a matrix must be square

before it makes sense to ask whether or not it has an inverse. Not all square matrices have inverses. If $\mathbf{M}$ has a row consisting entirely of zeros, so will $\mathbf{MM}'$ for any matrix $\mathbf{M}'$. This makes it easy to tell when a matrix in row reduced echelon form has an inverse: a square matrix in row reduced echelon form either has a row of zeros or it is an identity matrix; if it has a row of zeros, then it has no inverse; if it is an identity matrix, then it has an inverse.

The next proposition tells us that the product of invertable matrices is also invertable. This will let us find the inverse of a matrix as the product of matrices known to have inverses.

Proposition 10.1.6 *If $\mathbf{A}$ and $\mathbf{B}$ are invertable matrices then*

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

PROOF:

This is a simple calculation:

$$\begin{aligned}(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) &= \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1}\\ &= \mathbf{AIA}^{-1}\\ &= \mathbf{AA}^{-1}\\ &= \mathbf{I}.\end{aligned}$$

And the calculation of $(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB})$ is similar. The order of the factors is reversed in the inverse because multiplication of matrices is not commutative. ∎

A straightforward induction argument will extend this result to any finite number of factors. To find the inverse of the product of invertable matrices, take the product of the inverses in reversed order.

Matrices for elementary row operations The row reduction process we used in Chapter 7 made use of three kinds of operations on the rows of a matrix: interchanging two rows, multiplying a row by a non-zero number, and adding a multiple of one row to another. Each of these can be undone by an operation of the same type. If you want to undo interchange of rows 5 and 23, all you need to do is interchange them again. If you want to undo multiplication by a nonzero number, divide by it. If you want to undo addition of 6 times row 2 to row 7, subtract 6 times row 2 from row 7.

Proposition 10.1.7 *Each of the elementary row operations can be accomplished by left multiplication by a matrix which has an inverse.*

PROOF:

The matrix which accomplishes the elementary row operation is the one you get by applying the row operation to the identity matrix. It has an inverse because the operation can be undone by applying an appropriate row operation. ∎

Example: Interchange

The matrix which accomplishes interchange of rows 2 and 3 in 3 by 3 matrices is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0. \end{bmatrix}$$

The matrix which adds 3 times row 1 to row 2 is

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which has inverse

$$\begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1. \end{bmatrix}$$

Try them on some 3 by 3 matrices to check that they perform as claimed. ♢

Finding inverses by Row Reduction The algorithm for row reduction tells us how to multiply by a sequence of matrices which we know have inverses (those corresponding to elementary row operations) to get a matrix in row reduced echelon form. This will lead to an algorithm for finding the inverse of a matrix based on the algorithm for row reduction. (Gaussian elimination keeps coming up in different forms in this subject!)

To find the inverse of an $n \times n$ matrix $\mathbf{M}$ we do the following steps:

1. Augment the matrix to an $n \times 2n$ matrix by adjoining an $n \times n$ identity matrix on the right:
$$[\mathbf{M}|\mathbf{I}]$$

2. Reduce $\mathbf{M}$ to row reduced echelon form, performing all of the row operations on the identity matrix as you go:
$$[\mathbf{M}|\mathbf{I}] \rightsquigarrow [\mathbf{E}|\mathbf{B}]$$

3. If the row reduced echelon form matrix $\mathbf{E}$ is an identity matrix, then the product of the matrices which did the row reduction, now stored as $\mathbf{B}$, is $\mathbf{M}^{-1}$;

4. If the row reduction did not end in an identity matrix, then $\mathbf{M}$ does not have an inverse.

This works because each row operation can be done by multiplying on the right by an invertable matrix. Applying these operations to an identity matrix keeps track of the product of the matrices which do the row operations.

Example: Finding an inverse

To find the inverse of the matrix
$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ -1 & 3 & 3 \end{bmatrix}$$
we augment it with an identity
$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 7 & 0 & 1 & 0 \\ -1 & 3 & 3 & 0 & 0 & 1 \end{bmatrix}$$
and find the row reduced echelon form:
$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 7 & 0 & 1 & 0 \\ -1 & 3 & 3 & 0 & 0 & 1 \end{bmatrix} \quad \begin{matrix} R_2 - 2R_1 \\ R_3 + R_1 \\ \rightsquigarrow \end{matrix} \quad \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 5 & 6 & 1 & 0 & 1 \end{bmatrix}$$

$$\begin{array}{c} R_3 - 5R_2 \\ \rightsquigarrow \end{array} \quad \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & 11 & -5 & 1 \end{bmatrix}$$

$$\begin{array}{c} R_2 - R_3 \\ R_1 - 3R_3 \\ \rightsquigarrow \end{array} \quad \begin{bmatrix} 1 & 2 & 0 & -32 & 15 & -3 \\ 0 & 1 & 0 & -13 & 6 & -1 \\ 0 & 0 & 1 & 11 & -5 & 1 \end{bmatrix}$$

$$\begin{array}{c} R_1 - 2R_2 \\ \rightsquigarrow \end{array} \quad \begin{bmatrix} 1 & 0 & 0 & -6 & 3 & -1 \\ 0 & 1 & 0 & -13 & 6 & -1 \\ 0 & 0 & 1 & 11 & -5 & 1 \end{bmatrix}$$

So the inverse of

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ -1 & 3 & 3 \end{bmatrix}$$

is

$$\begin{bmatrix} -6 & 3 & -1 \\ -13 & 6 & -1 \\ 11 & -5 & 1 \end{bmatrix}.$$

$\diamondsuit$

Example: A matrix with no inverse

To show that the matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

does not have an inverse we can try the same algorithm and see where we get stuck:

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & -6 & -12 & -7 & 0 & 1 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & 4/3 & -1/3 & 0 \\ 0 & -6 & -12 & -7 & 0 & 1 \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & 4/3 & -1/3 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

Since we cannot get an identity on the left, there is no inverse. $\diamondsuit$

Exercises 10.1:

1. Write down the 5 by 5 matrix you would multiply on the left by to interchange rows 3 and 5? What happens if you multiply on the right by the matrix you have given?

2. Write down the 5 by 5 matrix you would multiply on the left by to add 3 times row 2 to row 4? What happens if you multiply on the right by the matrix you have given?

3. Write down the 5 by 5 matrix you would multiply on the left by to multiply row 4 by -2? What happens if you multiply on the right by the matrix you have given?

4. Write down the 5 by 5 matrix you would multiply on the left by to add -4 times row 1 to row 3? What happens if you multiply on the right by the matrix you have given?

For problems 5–10, find the inverse, if any:

5. $\begin{bmatrix} 1 & 4 & 6 \\ 2 & 3 & 1 \\ 1 & -1 & -1 \end{bmatrix}$

6. $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & -9 \end{bmatrix}$

7. $\begin{bmatrix} 0 & 2 & 4 \\ 1 & 0 & 3 \\ 2 & 1 & 3 \end{bmatrix}$

8. $\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{bmatrix}$

9. $\begin{bmatrix} 3 & 6 & 9 \\ 2 & 3 & -2 \\ 1 & 0 & 1 \end{bmatrix}$

10. $\begin{bmatrix} 1 & 0 & 1 & 0 & 2 \\ 2 & 1 & 0 & 1 & 1 \\ 0 & 1 & 3 & 2 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 \end{bmatrix}$

11. Prove that if L preserves sums and has an inverse as a function, then that inverse also preserves sums.

12. Assume that $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional. Show that if $L : \mathcal{V} \to \mathcal{W}$ is onto then there is a linear function M with $L \circ M = \mathrm{Id}_{\mathcal{W}}$. Such an M is called a *right inverse* for L.

13. Assume that $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional. Show that if $L : \mathcal{V} \to \mathcal{W}$ is one-to-one then there is a linear function N with $N \circ L = \mathrm{Id}_{\mathcal{V}}$. Such an N is called a *left inverse* for L.

14. Prove that if L is invertable and $L \circ M = \mathrm{Id}_{\mathcal{W}}$ then $M = L^{-1}$.

15. (Project Problem) Operation counts tell us when it is advantageous to use one algorithm over another.

    (a) Count the number of multiplications necessary to find the inverse of an $n \times n$ matrix using row reduction.

    (b) Count the number of multiplications needed to multiply an $n \times n$ matrix and an n-element column vector.

    (c) If you have m systems of linear equations with the same $n \times n$ coefficient matrix but different constant vectors, when (if ever) is it advantageous to find the inverse and then multiply the constant vector by it to find the solutions rather than doing Gaussian elimination with backsolving m times?

10.2 Rank

We noted above that the dimensions of the kernel and the image of a linear transformation could be useful in deciding whether or not it has an inverse.

Since further information can also be derived from these notions and the related rank of a matrix, which can be easily computed using row reduction.

Definition 10.2.1 *The* **rank** *of a linear transformation $L : \mathcal{V} \to \mathcal{W}$ is the dimension of ImL.*

Definition 10.2.2 *The* **nullity** *of a linear transformation $L : \mathcal{V} \to \mathcal{W}$ is* $\dim(Ker(L))$*.*

The rank of a linear transformation is made more useful by the next theorem which relates it to the dimension of the kernel.

Theorem 10.2.1 (Rank-Nullity Theorem) *If $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation and $\mathcal{V}$ is finite dimensional then* $\dim(\mathcal{V}) = \dim(Ker(L)) + rank(L)$*.*

PROOF:

We know that $Ker(L)$ is a subspace of $\mathcal{V}$, so we can find a basis for it, say $\vec{k}_1, \ldots, \vec{k}_n$. Since $\mathcal{V}$ is finite dimensional we know how to extend this to a basis for $\mathcal{V}$ by adding more vectors $\vec{b}_1, \ldots, \vec{b}_m$. We claim that the vectors $L(\vec{b}_1), \ldots, L(\vec{b}_m)$ form a basis for $\mathrm{Im}(L)$. It is clear that the images of all of the basis vectors in $\mathcal{V}$ form a spanning set for the image of L. Since $L(\vec{k}_i) = \vec{0}$ for all i, this tells us that the set of $L(\vec{b}_j)$ is a spanning set. Now suppose that

$$a_1 L(\vec{b}_1) + \ldots + a_m L(\vec{b}_m) = \vec{0}.$$

Then

$$L(a_1\vec{b}_1 + \ldots + a_m\vec{b}_m) = \vec{0}.$$

So $a_1\vec{b}_1 + \ldots + a_m\vec{b}_m$ is in $Ker(L)$. But this contradicts the independence of $\vec{k}_1, \ldots, \vec{k}_n, \vec{b}_1, \ldots, \vec{b}_m$ unless all of the a_i are 0. Thus the set $\{L(\vec{b}_1), \ldots, L(\vec{b}_m)\}$ is a basis for $Im(L)$. This shows that

$$\dim(\mathcal{V}) = \dim(Ker(L)) + \mathrm{rank}(L)$$

. ∎

The theorem is useful for obtaining information about a linear transformation from its rank. For instance, a linear transformation is one to one if and only if its kernel is just the zero vector. By the rank nullity theorem this can be shown by showing that $dim(Im(L)) = dim(\text{domain of } L)$. A linear transformation will have an inverse if it is one to one and onto; this in turn can be determined by looking at the dimensions of its domain and codomain and its rank, all of which must be equal.

As usual, in order to find the rank of a linear transformation we will want to do computations on the matrix representing it with respect to ordered bases.

Definition 10.2.3 *The row space of an n by m matrix $\mathbf{A}$ is the subspace of $\mathbb{R}^m$ spanned by the rows $\mathbf{A}_{i\cdot} = [a_{i1}, a_{i2}, \ldots, a_{im}]$. The column space is the subspace of $\mathbb{R}^n$ spanned by the columns $\mathbf{A}_{\cdot j} = [a_{1j}, a_{2j}, \ldots, a_{nj}]^t$.*

Definition 10.2.4 *The row rank of a matrix $\mathbf{A}$ is the dimension of its row space. The column rank is the dimension of its column space.*

All of our work with matrices so far has been linked to notions for linear transformations. If we recall that the columns of the matrix associated to a linear transformation are obtained by taking the image of basis vectors, the following proposition becomes clear:

Proposition 10.2.2 *The rank of a linear transformation is the same as the column rank of its matrix with respect to any choice of ordered bases.*

Given an arbitrary matrix it would appear to be a lot of work to find either of the ranks, though if the form is nice one of them may be easier to calculate than the other. For matrices in row reduced echelon form the row rank is particularly easy to identify since the nonzero rows in an echelon matrix are easily seen to be independent. Thus the row rank of an echelon matrix will be equal to the number of nonzero rows. To see this we need some propositions:

Proposition 10.2.3 *Elementary row operations leave the row space unchanged.*

PROOF:

It is clear that interchanging the order of the rows does not change the space which is spanned. It is also clear that multiplying a row by a non-zero scalar also does not change the space spanned. Adding a multiple of row i to row j replaces row j with a linear combination of rows i and j and does not change the space spanned by the rows. ∎

Proposition 10.2.4 *Elementary row operations leave the dimension of the column space unchanged.*

Proof:

It will suffice to show that a linear combination of the columns after an elementary row operation is applied gives $\vec{0}$ if and only if the same linear combination of the columns before the row operation is applied gives $\vec{0}$. Then if a basis for the column space of the original matrix uses, say, columns 1,2,3,5, and 8, then those same columns of the matrix after the row operation form a basis of its column space. Let $\mathbf{C}$ be the matrix with the columns in question from $\mathbf{M}$. Then linear independence of the columns of C says that the only solution to the system

$$\mathbf{C}\vec{x} = \vec{0}$$

is $\vec{x} = \vec{0}$. Now let $\mathbf{R}$ be the matrix which does the elementary row operation. The matrix $\mathbf{RC}$ is the matrix of columns after the row operation. Linear independence of the columns then asks for a unique answer to

$$\mathbf{RC}\vec{x} = \vec{0}.$$

Now any answer for this system gives an answer for the system

$$\mathbf{R}^{-1}\mathbf{RC}\vec{x} = \mathbf{R}^{-1}\vec{0}$$

which is the same as

$$\mathbf{C}\vec{x} = \mathbf{R}^{-1}\vec{0} = \vec{0}.$$

Similarly any answer to

$$\mathbf{C}\vec{x} = \vec{0}$$

gives an answer to

$$\mathbf{RC}\vec{x} = \mathbf{R}\vec{0} = \vec{0}.$$

Thus these two systems have the same number of solutions. If the solution to one is unique, so is the solution to the other. ∎

Combining these two propositions and the algorithm for reducing matrices to row reduced echelon form we get

Theorem 10.2.5 *The row rank of a matrix always equals its column rank.*

PROOF:

The process of reduction to row reduced echelon form does not change either the row space or the dimension of the column space, so it suffices to consider only matrices in row reduced echelon form. In the row reduced echelon form the first nonzero entry in each row is a one and all of the other entries in its column are zeros. This guarantees that the set of rows which are not all zero is a linearly independent set. It also spans the row space. Furthermore, any column in the matrix can be written as a linear combination of the columns in which the first non-zero entries in the rows of the (row reduced) matrix appear, so those columns form a spanning set for the column space. This set of columns is easily seen to be independent, since each such column has only one non zero entry and the non-zero entries occur in different rows. Thus we have a basis for the column space which has exactly the same number of members as the number of non-zero rows. Since both of these are bases, we have shown that the dimension of the row space is equal to the dimension of the column space. This says that the row rank equals the column rank. ∎

Example: Rank of a matrix

Matrix	rank
$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$	3
$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$	2
$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix}$	2

$\diamondsuit$

Once we know how to find the rank of a matrix we can use the Rank-Nullity Theorem to determine how many basic solutions there are to various kinds of linear problems. If we have a homogeneous system of linear equations $\mathbf{M}\vec{v}^t = \vec{0}$, then the dimension of the domain equals the number of variables, the rank of $\mathbf{M}$ can be computed easily, and the rank-nullity theorem tells us that the dimension of the space of solutions = number of variables - rank $\mathbf{M}$.

We can also see how the Rank-Nullity Theorem relates to the structure of the row reduced echelon form for a matrix. Recall that the leading 1's in the rows of the row reduced echelon form give the columns in the original matrix which formed a basis for the column space. Thus the number of non-zero rows gives the rank of the matrix. In solving a homogeneous system of linear equations we had a free choice for each variable corresponding to a column which did not have a leading 1 in the row reduced echelon form. Thus the number of columns not having leading 1's gives the dimension of Ker(L), the nullity. Since this accounts for all of the columns of the matrix for L, we get

$$\begin{aligned} \dim(\mathcal{V}) &= \text{number of columns} \\ &= \text{number of columns with leading 1's } + \text{number without} \\ &= \text{rank}(L) + \text{nullity}(L) \end{aligned}$$

Exercises 10.2:

1. Find the rank of $\begin{bmatrix} -1 & 2 & -3 \\ 4 & -5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

2. Find the rank of $\begin{bmatrix} 1 & 4 & 6 & 2 \\ 4 & -5 & 1 & 6 \\ 2 & 8 & -1 & -9 \\ 1 & 7 & 2 & -5 \end{bmatrix}$

3. Find the rank of $\begin{bmatrix} 1 & 2 & 6 \\ 2 & 4 & 7 \\ 3 & 3 & 9 \\ 1 & 1 & 1 \\ 2 & 5 & 10 \end{bmatrix}$

4. Find the rank of $\begin{bmatrix} 1 & 3 & 7 & 2 \\ 2 & 4 & 8 & 3 \\ 3 & 9 & 5 & 1 \\ 2 & 1 & 1 & 4 \\ 3 & 1 & 1 & 6 \end{bmatrix}$

5. Find the rank of $\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 3 & 4 & 5 & 6 & 7 \\ 2 & 0 & 4 & 1 & 1 \\ 2 & 1 & 1 & 1 & 1 \end{bmatrix}$

6. Find the rank of $F : \mathbb{R}^3 \to \mathbb{R}^3$ taking $[x, y, z]$ to $[2x, y - z, 0]$. What is the dimension of $\mathrm{Ker}(F)$?

7. Find the rank of $L : \mathbb{R}^4 \to \mathbb{R}^3$ taking $[x, y, z, w]$ to $[x + y + z, x + y + w, x + y + z + w]$. What is the dimension of $\mathrm{Ker}(L)$?

8. Find the rank of $G : \mathbb{R}^3 \to \mathbb{R}^4$ taking $[x, y, z]$ to $[y+z, x+z, 2z, x+y+z]$. What is the dimension of $\mathrm{Ker}(G)$?

9. Find the rank of $H : \mathbb{R}^3 \to \mathbb{R}^5$ taking $[x, y, z]$ to $[x+y+z, x+z, 2z, 2x+y+4z, z-x-y]$. What is the dimension of $\mathrm{Ker}(G)$?

10. Find the rank of $K : \mathbb{R}^6 \to \mathbb{R}^4$ taking $[u, v, w, x, y, z]$ to $[u - y + z, v + x + z, 3u + w - 2z, -2u + 2v - w + 2x - y + 5z]$. What is the dimension of $\text{Ker}(G)$?

11. Show that if $r \neq 0$ then the nullity of rL is the same as the nullity of L.

12. Show that if $r \neq 0$ then the rank of rL is the same as the rank of L.

13. Show that if $L : \mathcal{V} \to \mathcal{W}$ and $M : \mathcal{U} \to \mathcal{V}$ then the nullity$(L \circ M) \geq$ nullity(M).

14. Show that if $L : \mathcal{V} \to \mathcal{W}$ and $M : \mathcal{U} \to \mathcal{V}$ then the rank$(L \circ M) \leq$ rank(L).

15. Give examples of 2×2 matrices $\mathbf{A}$ and $\mathbf{B}$ such that

    (a) $\text{rank}(\mathbf{A}) = 2, \text{rank}(\mathbf{B}) = 2$ and $\text{rank}(\mathbf{A} + \mathbf{B}) = 1$

    (b) $\text{rank}(\mathbf{A}) = 1, \text{rank}(\mathbf{B}) = 1$ and $\text{rank}(\mathbf{A} + \mathbf{B}) = 2$

16. Prove that in general $\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \geq \text{rank}(\mathbf{A} + \mathbf{B})$.

10.3 LU Decomposition

Modern matrix algebra makes use of a number of factorization theorems. One class of these involves factorization of a matrix into the product of a lower triangular matrix and an upper triangular matrix. Different choices of what the diagonal entries should be lead to different factorizations with different properties more properly discussed in a numerical analysis course. In this section we will note that Gaussian elimination can be seen as a means of obtaining such a factorization.

In the Gaussian elimination with backsolving algorithm we divided the work of reducing a matrix to row reduced echelon form into two parts: first we got the ones as first non-zero entries and the zeros below those ones and then we went back to get the zeros above the ones. Between the two phases of that algorithm we had a matrix with a nice form worth giving a name.

Definition 10.3.1 *A matrix which has $a_{ij} = 0$ whenever $i < j$ is called* **lower triangular.** *A matrix with $a_{ij} = 0$ whenever $i > j$ is called* **upper**

triangular. *We add the adjective* strictly *if all of the diagonal entries are 0.*

Lemma 10.3.1 *A product of lower triangular matrices is lower triangular.*

PROOF:

Suppose we have two lower triangular matrices $\mathbf{L}$ and $\mathbf{L}'$. Since these are lower triangular, we know that $l_{ik} = 0$ if $i < k$ and $l'_{kj} = 0$ if $k < j$ The ij entry of $\mathbf{LL}'$ is the sum of terms of the form $l_{ik}l'_{kj}$. If $i < j$ then either $i < k$ or $k < j$, so one of the two terms in the product will be 0. ∎

Proposition 10.3.2 *Any matrix* $\mathbf{M}$ *which can be reduced to echelon form without row interchanges can be written as the product of a lower triangular matrix and an upper triangular matrix.*

PROOF:

First observe that the row operations "multiply a row by a constant" and "add a multiple of a row to a row below it" are lower triangular. If no interchanges are needed in reduction to echelon form then organizing the work so that we get the ones as first nonzero entries in each row and then get the zeros below those entries gives us an upper triangular matrix $\mathbf{U}$. Keeping track of the row operations in the same way that we did in the algorithm for finding the inverse of a matrix gives:

$$[\mathbf{M}|\mathbf{I}] \underset{\leadsto}{} [\mathbf{U}|\mathbf{R}],$$

where $\mathbf{R} = \mathbf{R}_k \ldots \mathbf{R}_1$ is the lower triangular matrix which is obtained as the product of the row operation matrices and $\mathbf{U}$ is the upper triangular matrix we find at the halfway point of the Gaussian elimination algorithm. The matrices $\mathbf{R}_i$ are associated with elementary row operations of the type "multiply a row by a constant" or "add a multiple of a row to a row below it." Thus each has an inverse which is also lower triangular: we undo these operations by either "dividing a row by a constant" or "subtracting a multiple of a row from a row below it." Thus we get

$$\mathbf{R}^{-1} = \mathbf{R}_1^{-1} \ldots \mathbf{R}_k^{-1}.$$

Since each of the $\mathbf{R}_i^{-1}$ is lower triangular, this tells us that $\mathbf{R}^{-1}$ is.

From our row reduction of $\mathbf{M}$ we saw that

$$\mathbf{RM} = \mathbf{U},$$

so multiplying both sides by $\mathbf{R}^{-1}$ gives

$$\mathbf{R}^{-1}\mathbf{U} = \mathbf{M}.$$

This gives the desired factorization. ∎

Example: LU factorization

To factor the matrix

$$\mathbf{M} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 0 & 12 \\ 1 & 1 & 3 \end{bmatrix}$$

we first apply elementary row operations until we have an upper triangular matrix, keeping track of what row operations we have done in the same fashion as we did when we were finding the inverse of $\mathbf{M}$:

$$\left[\begin{array}{cccccc} 1 & 2 & 4 & 1 & 0 & 0 \\ 2 & 0 & 12 & 0 & 1 & 0 \\ 1 & 1 & 3 & 0 & 0 & 1 \end{array}\right] \quad \begin{array}{c} R_2 - 2R_1 \\ R_3 - R_1 \\ \rightsquigarrow \end{array} \quad \left[\begin{array}{cccccc} 1 & 2 & 4 & 1 & 0 & 0 \\ 0 & -4 & 4 & -2 & 1 & 0 \\ 0 & -1 & -1 & -1 & 0 & 1 \end{array}\right]$$

$$\begin{array}{c} R_3 - \frac{1}{4}R_2 \\ \rightsquigarrow \end{array} \quad \left[\begin{array}{cccccc} 1 & 2 & 4 & 1 & 0 & 0 \\ 0 & -4 & 4 & -2 & 1 & 0 \\ 0 & 0 & -2 & \frac{1}{2} & -\frac{1}{4} & 1 \end{array}\right]$$

This tells us that we can use

$$\mathbf{U} = \begin{bmatrix} 1 & 2 & 4 \\ 0 & -4 & 4 \\ 0 & 0 & -2 \end{bmatrix} \text{ and } \mathbf{L}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ \frac{1}{2} & -\frac{1}{4} & 1 \end{bmatrix}.$$

In order to complete the factorization we will need to find $\mathbf{L}$ by taking the inverse of $\mathbf{L}^{-1}$:

$$\left[\begin{array}{cccccc} 1 & 0 & 0 & 1 & 0 & 0 \\ -2 & 1 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & -\frac{1}{4} & 1 & 0 & 0 & 1 \end{array}\right] \quad \begin{array}{c} R_2 + 2R_1 \\ R_3 - \frac{1}{2}R_1 \\ \rightsquigarrow \end{array} \quad \left[\begin{array}{cccccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & -\frac{1}{4} & 1 & -\frac{1}{2} & 0 & 1 \end{array}\right]$$

$$\begin{array}{c} R_3 + \frac{1}{4}R_2 \\ \rightsquigarrow \end{array} \quad \left[\begin{array}{cccccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 1 & \frac{1}{4} & 1 \end{array}\right]$$

Now notice that

$$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 0 & 12 \\ 1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & \frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & -4 & 4 \\ 0 & 0 & -2 \end{bmatrix}$$

as desired. $\diamondsuit$

Note that if $\mathbf{A}$ cannot be reduced to echelon form without interchanges (say $a_{11} = 0$, for instance) it still can be written in the form $\mathbf{XU}$ where $\mathbf{X}$ is the product of matrices corresponding to elementary row operations and $\mathbf{U}$ is upper triangular. It is also possible to group the interchanges together so that the factorization is into a lower triangular matrix, a permutation matrix, and an upper triangular matrix.

Exercises 10.3:

Write the following matrices as the product of a lower triangular matrix and an upper triangular matrix:

1. $\begin{bmatrix} 2 & 8 \\ -4 & 3 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 7 \\ -1 & 1 & 5 \end{bmatrix}$

3. $\begin{bmatrix} 2 & 4 & 12 \\ 1 & 6 & -14 \\ 3 & 8 & 9 \end{bmatrix}$

4. To show that a matrix which uses interchange of rows in the first phase of Gaussian elimination can be written in the form **LXU** where **X** is a permutation matrix, we need to be able to collect the row interchanges into one matrix. To show this can be done, prove that
   (a) The sequence of row operations consisting of interchanging two rows and then adding a multiple of one row to another can be done by adding a multiple of one row to another and then interchanging rows, though the rows referred to may change.
   (b) The sequence of row operations consisting of interchanging two rows and then multiplying a row by a constant can be done by multiplying a row by a constant and then interchanging rows, though the rows referred to may change.
   (c) The sequence of row operations consisting of interchanging two rows and then interchanging two rows can be done by interchanging two rows and then interchanging rows, though the rows referred to may change.

5. (Project Problem) The LU decomposition can be useful when solving many systems of equations with the same coefficient matrix. To see why
   (a) Find how many multiplications are required to find $\mathbf{L}^{-1} = \mathbf{R}$.
   (b) Find how many multiplications are needed to multiply $\mathbf{L}^{-1}$ times a column vector of constants.
   (c) We showed earlier that it takes
   $$\frac{n^3}{3} + n^2 - \frac{n}{3}$$
   multiplications to solve a system of n linear equations in n unknowns using Gaussian elimination with backsolving and that the backsolving itself took
   $$\frac{n^2 - n}{2}$$
   multiplications. Suppose you have m systems all using the same coefficient matrix. We can then use one of the following three strategies:

i. Use Gaussian elimination with backsolving for each system, ignoring previous work.
ii. Find $\mathbf{L}^{-1}$ for the matrix of coefficients, then solve each system by multiplying the column of constants by $\mathbf{L}^{-1}$ and then backsolving.
iii. Finding the inverse of the matrix of coefficients and then solving each system by multiplying the column of constants by that inverse.

Under what conditions on m will each of these strategies involve less work that the others?

Chapter 11

Change of Basis and Similar Matrices

In the exercises after section 1 of the Chapter 8 you were asked to find the matrices for a linear transformation with respect to several different pairs of ordered bases. You did this by calculating the effect of the linear transformation on each basis vector in the domain and then finding ways to represent the answer in terms of the basis on the codomain. The resulting matrices were different, even though they represented the same linear transformation, because they used different bases. In this section we will see how to use the isomorphism of matrix algebra and the algebra of transformations to change bases without having to recalculate everything.

Our approach is to think of change of basis as composition with an identity transformation which is then represented using the old and the new ordered bases so that the basis we want is on the outside. Change of basis on the domain is then multiplying on the right by an appropriate matrix and change of basis on the codomain is multiplication on the left. Because our matrices will come from compositions and our multiplication of matrices corresponds to composition written from right to left, we will write our diagrams with the arrows going from right to left in this chapter.

11.1 Matrices for Changing Basis

Suppose we have a linear transformation $M : \mathcal{W} \longleftarrow \mathcal{V}$, (notice that $\mathcal{V}$ is the domain and $\mathcal{W}$ is the codomain) which has matrix $\mathbf{M}$ with respect to

the ordered bases $(\vec{b}_1, \ldots, \vec{b}_p)$ for $\mathcal{V}$ and $(\vec{c}_1, \ldots, \vec{c}_q)$ for $\mathcal{W}$. Let us see how to change these bases one at a time:

11.1.1 How to change the basis on the domain

Suppose we want to use the basis $(\vec{a}_1, \ldots \vec{a}_p)$ for the domain instead of the basis $(\vec{b}_1, \ldots \vec{b}_p)$. The trick here is to multiply on the right by a matrix $\mathbf{Q}$ obtained by looking at the identity transformation on $\mathcal{V}$ going *from* our new basis *to* the original basis.

$$\begin{array}{rccccc} \text{vector spaces:} & \mathcal{W} & \stackrel{M}{\longleftarrow} & \mathcal{V} & \stackrel{\text{Id}}{\longleftarrow} & \mathcal{V} \\ \text{ordered bases:} & (\vec{c}_1, \ldots \vec{c}_q) & & (\vec{b}_1, \ldots \vec{b}_p) & & (\vec{a}_1, \ldots \vec{a}_p) \\ \text{matrices :} & & \mathbf{M} & & \mathbf{Q} & \\ & & & & = \mathbf{N} & \end{array}$$

For example, if $\mathcal{V} = \mathbb{R}^3$, the original basis was the standard basis, and the new basis is $([1,2,3],[1,0,4],[2,1,0])$ then the matrix $\mathbf{Q}$ is found by calculating:

$$\begin{aligned} [1,2,3] &= 1[1,0,0] + 2[0,1,0] + 3[0,0,1] \\ [1,0,4] &= 1[1,0,0] + 0[0,1,0] + 4[0,0,1] \\ [2,1,0] &= 2[1,0,0] + 1[0,1,0] + 0[0,0,1] \end{aligned}$$

This gives the matrix

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & 1 \\ 3 & 4 & 0 \end{bmatrix}.$$

The *columns* are given by the coefficients in the equations.

11.1.2 How to change the basis on the codomain

Suppose we want to use the basis $(\vec{d}_1, \ldots \vec{d}_q)$ for the codomain instead of the basis $(\vec{c}_1, \ldots \vec{c}_q)$. The trick here is to multiply on the left by a matrix $\mathbf{P}^{-1}$ obtained by looking at the identity transformation on $\mathcal{W}$ going *from* our original basis *to* the new basis. The inverse in this matrix is there to remind us that this is the opposite direction from the placement of the bases when we change bases on the domain.

$$\begin{array}{rccccc} \text{vector spaces:} & \mathcal{W} & \xleftarrow{\text{Id}} & \mathcal{W} & \xleftarrow{M} & \mathcal{V} \\ \text{ordered bases:} & (\vec{d}_1, \ldots \vec{d}_q) & & (\vec{c}_1, \ldots, \vec{c}_q) & & (\vec{b}_1, \ldots \vec{b}_p) \\ \text{matrices :} & & \mathbf{P}^{-1} & & \mathbf{M} & \\ & & & & = \mathbf{N} & \end{array}$$

Again an example may help: let $\mathcal{W} = \mathbb{R}^2$, $(\vec{c}_1, \ldots, \vec{c}_q) = ([1,0],[0,1])$, and $(\vec{d}_1, \ldots \vec{d}_q) = ([2,4],[1,3])$. We need to find our original bases in terms of the new bases. This involves solving two systems of equations:

$$\begin{array}{rcl} 2x + y & = & 1 \\ 4x + 3y & = & 0 \end{array} \text{ and } \begin{array}{rcl} 2x + y & = & 0 \\ 4x + 3y & = & 1 \end{array}$$

These are exactly the systems we solve as we find the inverse of the matrix

$$\mathbf{P} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}.$$

The solution is obtained by row reduction :

$$\begin{array}{ccc} \begin{bmatrix} 2 & 1 & 1 & 0 \\ 4 & 3 & 0 & 1 \end{bmatrix} & \begin{array}{c} R_2 - 2R_1 \\ \rightsquigarrow \end{array} & \begin{bmatrix} 2 & 1 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix} \\ & \begin{array}{c} R_1 - R_2 \\ \rightsquigarrow \end{array} & \begin{bmatrix} 2 & 0 & 3 & -1 \\ 0 & 1 & -2 & 1 \end{bmatrix} \\ & \begin{array}{c} \frac{1}{2}R_1 \\ \rightsquigarrow \end{array} & \begin{bmatrix} 1 & 0 & \frac{3}{2} & -\frac{1}{2} \\ 0 & 1 & -2 & 1 \end{bmatrix} \end{array}$$

Thus the matrix $\mathbf{P}^{-1}$ which changes the basis on the codomain is

$$\begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -2 & 1 \end{bmatrix}.$$

11.1.3 General change of basis

It should be clear that to change the basis on both the domain and the codomain we need both change of basis matrices. The diagram then looks like this:

$$\begin{array}{lccccccc} \text{Vector spaces:} & \mathcal{W} & \xleftarrow{\text{Id}} & \mathcal{W} & \xleftarrow{M} & \mathcal{V} & \xleftarrow{\text{Id}} & \mathcal{V} \\ \text{ordered bases:} & (\vec{d}_1, \ldots \vec{d}_q) & & (\vec{c}_1, \ldots, \vec{c}_q) & & (\vec{b}_1, \ldots \vec{b}_p) & & (\vec{a}_1, \ldots \vec{a}_p) \\ \text{matrices :} & & \mathbf{P}^{-1} & & \mathbf{M} & & \mathbf{Q} & \\ & & & & = \mathbf{N} & & & \end{array}$$

Since composition of linear transformations is represented by multiplication of matrices $\mathbf{P}^{-1}\mathbf{M}\mathbf{Q}$ represents the composition $\mathrm{Id}M\mathrm{Id} = M$ with respect to the outside bases. But this is exactly what the matrix $\mathbf{N}$ does.

Example: Using change of basis matrices

Let us start by looking at the example of the linear transformation M from $\mathbb{R}^2$ to $\mathbb{R}^3$ taking $[x, y]$ to $[x + y, x - y, 3y]$. If we use the standard bases on the domain and codomain, then we get the matrix for M by calculating

$$M[1, 0] = [1, 1, 0] = 1[1, 0, 0] + 1[0, 1, 0] + 0[0, 0, 1]$$

and

$$M[0, 1] = [1, -1, 3] = 1[1, 0, 0] + -1[0, 1, 0] + 3[0, 0, 1],$$

getting the matrix

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 3 \end{bmatrix}.$$

If we use the ordered bases $([1, 1], [1, -1])$ and $([1, 0, 0], [1, 1, 0], [1, 1, 1])$ instead the calculation gives

$$M[1, 1] = [2, 0, 3] = 2[1, 0, 0] + -3[1, 1, 0] + 3[1, 1, 1]$$

and

$$M[1, -1] = [0, 2, -3] = -2[1, 0, 0] + 5[1, 1, 0] + -3[1, 1, 1]$$

so the matrix is

$$\begin{bmatrix} 2 & -2 \\ -3 & 5 \\ 3 & -3 \end{bmatrix}$$

with respect to these bases.

Next let us find this same matrix representation using change of basis matrices. We know how to find the matrix for M with respect to the standard bases. In order to represent it with respect to the alternate bases we compose with the identity on both

sides and find matrices which represent the identity transformation with respect to the bases in question. We then get the matrix for L by multiplying these matrices.

To change basis from $([1,1],[1,-1])$ to the standard basis $([1,0],[0,1])$ we calculate

$$\mathrm{Id}([1,1]) = [1,1] = 1[1,0] + 1[0,1]$$

$$\mathrm{Id}([1,-1]) = [1,-1] = 1[1,0] + -1[0,1],$$

so the matrix for the identity $\mathbb{R}^2$ to itself using the new basis on the domain and the standard basis on the codomain is

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

To get the matrix for changing the basis on $\mathbb{R}^3$ from the standard basis to the basis $([1,0,0],[1,1,0],[1,1,1])$ we calculate

$$[1,0,0] = 1[1,0,0] + 0[1,1,0] + 0[1,1,1]$$

$$[0,1,0] = -1[1,0,0] + 1[1,1,0] + 0[1,1,1]$$

and

$$[0,0,1] = 0[1,0,0] + -1[1,1,0] + 1[1,1,1],$$

so the matrix for the identity on $\mathbb{R}^3$ with the standard basis on the domain and the alternative basis on the codomain is

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Using the fact that matrix multiplication represents composition of linear transformations, we can calculate the matrix for M with respect to the alternative bases by first changing from the basis $([1,1],[1,-1])$ to the standard basis, then using the matrix for M with respect to the standard bases, and then using the matrix for the change of basis from the standard basis to $([1,0,0],[1,1,0],[1,1,1])$. Since composition is written from right to left, so is the multiplication of these matrices:

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ -3 & 5 \\ 3 & -3 \end{bmatrix}$$

A diagram may help you understand this. The top level gives the linear transformations, the middle level gives the bases, and the bottom level gives the matrices:

$$
\begin{array}{lcccccc}
\text{Vector Spaces:} & \mathbb{R}^3 & \xleftarrow{\text{Id}} & \mathbb{R}^3 & \xleftarrow{M} & \mathbb{R}^2 \xleftarrow{\text{Id}} & \mathbb{R}^2 \\
\text{Bases:} & \begin{array}{c}([1,0,0],\\ [1,1,0],\\ [1,1,1])\end{array} & & \begin{array}{c}([1,0,0],\\ [0,1,0],\\ [0,0,1])\end{array} & & \begin{array}{c}([1,0],\\ [0,1])\end{array} & \begin{array}{c}([1,1],\\ [1,-1])\end{array} \\
\text{Matrices:} & & \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} & & \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 3 \end{bmatrix} & & \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\
\text{Final Matrix:} & & & \begin{bmatrix} 2 & -2 \\ -3 & 5 \\ 3 & -3 \end{bmatrix} & & &
\end{array}
$$

Since composition of linear transformations is represented by multiplication of matrices

$$
\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}
$$

represents the composition $\text{Id}\ M\ \text{Id} = M$ with respect to the outside bases. But this is exactly what the matrix

$$
\begin{bmatrix} 2 & -2 \\ -3 & 5 \\ 3 & -3 \end{bmatrix}
$$

does. $\diamondsuit$

This gives the following theorem:

Theorem 11.1.1 *Two matrices* $\mathbf{M}$ *and* $\mathbf{N}$ *represent the same linear transformation with respect to different bases if and only if there are matrices* $\mathbf{P}$ *and* $\mathbf{Q}$ *which have inverses and* $\mathbf{N} = \mathbf{P}^{-1}\mathbf{M}\mathbf{Q}$.

The usual practice is to restrict our attention to square matrices. We then think of them as representing a linear transformation from a vector space to itself with a particular choice of basis used for both the domain

and codomain. If we want to change the basis, the matrices $\mathbf{P}$ and $\mathbf{Q}$ in the theorem will be the same. Thus two square matrices $\mathbf{A}$ and $\mathbf{B}$ represent the same linear transformation if and only if there is a matrix $\mathbf{P}$ such that $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}$.

This motivates the following definition:

Definition 11.1.1 : *Two square matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *are said to be similar,* $\mathbf{A} \sim \mathbf{B}$, *if and only if there is a matrix* $\mathbf{P}$ *such that* $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}$.

It is clear that similarity is an equivalence relation on square matrices, since two matrices are similar if and only if they represent the same linear transformation.

Exercises 11.1:

For problems 1–5, write down the matrix for the identity transformation from $\mathcal{V}$ to $\mathcal{V}$ using the given bases for the domain and codomain:

1. $\mathcal{V} = \mathbb{R}^2$ Basis for domain: $([1,0],[0,1])$
   Basis for codomain: $([2,3],[-1,5])$

2. $\mathcal{V} = \mathbb{R}^3$ Basis for domain: $([1,1,0],[1,0,1],[1,1,1])$
   Basis for codomain: $([1,0,0],[0,1,0],[0,0,1])$

3. $\mathcal{V} = \mathbb{Z}_2^3$ Basis for domain:$([1,1,1],[1,0,1],[0,1,1])$
   Basis for codomain: $([1,0,0],[0,1,0],[0,0,1])$

4. $\mathcal{V} = \mathbb{Z}_2^3$ Basis for domain:$([1,0,0],[0,1,0],[0,0,1])$
   Basis for codomain: $([1,1,1],[1,0,1],[0,1,1])$

5. $\mathcal{V} = \mathbb{R}[x]_2$ Basis for domain: $(1, x, x^2)$
   Basis for codomain: $(1+x, 1-x, x^2+x+1)$

For problems 6–10 the matrix $\mathbf{M}$ for a linear transformation from $\mathbb{R}^n$ to itself using the standard basis for both domain and codomain is given. Find the matrix for the same linear transformation with respect to the given basis.

6. $\mathbf{M} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ Basis: $([1,2,3],[1,0,1],[0,1,2])$

7. $\mathbf{M} = \begin{bmatrix} 8 & 1 & -2 \\ 1 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$ Basis: $([1,1,1],[-1,1,1],[1,1,-1])$

8. $\mathbf{M} = \begin{bmatrix} -1 & -2 & -3 \\ 0 & 1 & 1 \\ 2 & 7 & -2 \end{bmatrix}$ Basis: $([0,1,2],[1,-2,3],[2,3,1])$

9. $\mathbf{M} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & -1 & 1 & -1 \\ 2 & 3 & 3 & 2 \\ 0 & -1 & 0 & -2 \end{bmatrix}$ Basis: $([1,0,0,0],[1,1,0,0],[1,1,1,0],[1,1,1,1])$

10. $\mathbf{M} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 3 & 1 & 4 \\ 1 & 2 & 3 & 4 & 5 \\ 0 & 1 & 1 & 0 & 0 \\ 2 & 3 & 4 & 3 & 2 \end{bmatrix}$
 Basis: $([1,0,0,0,0],[1,1,0,0,0],[1,1,1,0,0],[0,1,1,1,1],[0,0,0,0,1])$

For problems 11–15 the matrix $\mathbf{M}$ for a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ with respect to the standard basis is given. Give the matrices for the change of basis on both domain and codomain and the matrix for the linear transformation with respect to the new bases given in the problems.

11. $\mathbf{M} = \begin{bmatrix} 1 & 3 & -2 & 1 \\ 0 & 2 & -1 & 6 \\ 1 & 1 & 1 & 1 \end{bmatrix}$
 Basis for domain:
 $$([1,1,0,0],[1,0,-1,0],[1,0,0,2],[0,1,1,0])$$
 Basis for codomain:
 $$([1,1,0],[0,1,1],[1,0,1])$$

12. $\mathbf{M} = \begin{bmatrix} 2 & -4 & 1 \\ 0 & 2 & 6 \\ 1 & 1 & 10 \\ 1 & 2 & 3 \end{bmatrix}$
 Basis for domain:
 $$([1,1,0],[0,1,1],[1,0,1])$$

Basis for codomain:

$$([1,1,0,0],[1,0,-1,0],[1,0,0,1],[0,1,2,0])$$

13. $\mathbf{M} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 5 & -6 \\ 1 & 0 & 1 & 0 \end{bmatrix}$

    Basis for domain:

    $$([1,1,1,0],[1,1,0,1],[1,0,1,1],[0,1,1,1])$$

    Basis for codomain:

    $$([1,-1,0],[0,1,-1],[1,0,1])$$

14. $\mathbf{M} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 2 & 2 \\ 5 & -6 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

    Basis for domain:

    $$([1,-1,0],[0,1,-1],[1,0,-1])$$

    Basis for codomain:

    $$([1,1,1,0],[1,1,0,1],[1,0,1,1],[0,1,1,1])$$

15. $\mathbf{M} = \begin{bmatrix} 1 & 3 & -2 & 1 \\ 2 & 2 & 5 & -6 \\ 0 & 2 & -1 & 6 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

    Basis for domain:

    $$([1,1,0,0],[1,0,-1,0],[1,0,0,2],[0,1,1,0])$$

    Basis for codomain:

    $$([1,1,0,2],[2,0,1,1],[1,0,2,1],[1,1,1,1])$$

16. Prove that similarity of matrices is an equivalence relation (i.e., that it is transitive, symmetric, and reflexive) using properties of matrix multiplication.

11.2 Similar Matrices and Canonical Forms

One of the important problems of linear algebra is to find matrices similar to a given matrix which have particularly nice forms. Given the definition of similarity and Theorem 11.1.1, this is the same as finding nice bases to use in the representation of a linear transformation. In this section we will explore some of the ways to choose a basis so that the matrices for particular kinds of linear transformations have desirable forms.

Similar matrices from a factorization One way to get a matrix similar to a given matrix is to find a factorization of that matrix. If

$$\mathbf{M} = \mathbf{AB}$$

and $\mathbf{B}$ is invertable, then

$$\mathbf{M} = (\mathbf{B}^{-1}\mathbf{B})\mathbf{AB} = \mathbf{B}^{-1}(\mathbf{BA})\mathbf{B}$$

so $\mathbf{M}$ is similar to $\mathbf{BA}$. This is often used in numerical linear algebra to find matrices similar to a given matrix with small off-diagonal entries. Such similarity transformations are widely used in iterative numerical techniques for finding eigenvalues.

We have seen one standard factorization: the LU decomposition which encapsulates the work getting to an upper triangular form in the Gaussian elimination with backsolving algorithm. While this is not one of the factorizations used in finding eigenvalues we can use it to find similar matrices:

Example: Similar matrices from an LU decomposition

> We observed earlier that if a matrix can be reduced to row reduced echelon form without use of exchanges, then the Gaussian elimination algorithm gives a factorization into a lower triangular matrix and an upper triangular matrix. We keep track of the row operations used to get our matrix to upper triangular form and then apply their inverses in the opposite order to an identity matrix to get $\mathbf{L}$. If we apply that process to the matrix
>
> $$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 0 & 12 \\ 1 & 1 & 3 \end{bmatrix},$$

we get

$$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 0 & 12 \\ 1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & \frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & -4 & 4 \\ 0 & 0 & -2 \end{bmatrix}.$$

Thus

$$\begin{bmatrix} 1 & 2 & 4 \\ 2 & 0 & 12 \\ 1 & 1 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 4 \\ 0 & -4 & 4 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & \frac{1}{4} & 1 \end{bmatrix} = \begin{bmatrix} 9 & 3 & 4 \\ 4 & -3 & 4 \\ -2 & \frac{-1}{2} & -2 \end{bmatrix}.$$

$\diamondsuit$

Companion Matrices and L-cyclic Subspaces Let us fix a linear transformation $L : \mathcal{V} \to \mathcal{V}$. Assume that $\mathcal{V}$ is finite dimensional. If we start with a vector $\vec{v}$ and consider the ordered set $(\vec{v}, L(\vec{v}), L^2(\vec{v}), L^3(\vec{v}), \ldots)$ we will eventually reach an n such that $L^n(\vec{v})$ is a linear combination of the previous powers. Just for concreteness let us assume that n is the first power for which this happens. Then $(\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v}))$ is linearly independent and forms a basis for a subspace of $\mathcal{V}$. This subspace is important enough to have a name:

Definition 11.2.1 *The L-cyclic subspace of $\mathcal{V}$ generated by $\vec{v}$ is*

$$Span(\{\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v})\}) = C^L_{\vec{v}}.$$

If we are working with subspaces and a linear transformation from $\mathcal{V}$ to itself, then we will frequently want the linear transformation to map the subspaces into themselves.

Definition 11.2.2 *A subspace $\mathcal{U} \leq \mathcal{V}$ is called L-invariant if whenever $\vec{u} \in \mathcal{U}$ then $L(\vec{u}) \in \mathcal{U}$.*

Notice that it is clear from the definitions that $C^L_{\vec{v}}$ is L-invariant.

If we write the matrix for $L|_{C^L_{\vec{v}}} : C^L_{\vec{v}} \to C^L_{\vec{v}}$ with respect to the basis $(\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v}))$, we will get a particularly nice form, called a *companion matrix*. Each basis vector is taken to the next basis vector by L until we reach the last. Then we notice that

$$L^n(\vec{v}) = a_0\vec{v} + a_1 L(\vec{v}) + \ldots + a_{n-1}L^{n-1}(\vec{v})$$

so that the matrix for L with respect to this basis is

$$\begin{bmatrix} 0 & 0 & \dots & 0 & a_0 \\ 1 & 0 & \dots & 0 & a_1 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ddots & 0 & a_{n-2} \\ 0 & 0 & \dots & 1 & a_{n-1} \end{bmatrix}.$$

Example: Finding a companion matrix

If we let $L : \mathbb{R}^3 \to \mathbb{R}^3$ have $L([x, y, z]) = [x + y, 2y + 2z, x + z]$ then the L-cyclic subspace generated by $[1, 0, 0]$ is all of $\mathbb{R}^3$. The basis is $([1, 0, 0], [1, 0, 1], [1, 2, 2])$ and

$$L^3([1, 0, 0]) = [3, 8, 3] = 4[1, 0, 0] - 5[1, 0, 1] + 4[1, 2, 2].$$

The matrix for L with respect to the basis $([1, 0, 0], [1, 0, 1], [1, 2, 2])$ is

$$\begin{bmatrix} 0 & 0 & 4 \\ 1 & 0 & -5 \\ 0 & 1 & 4 \end{bmatrix}.$$

$\diamondsuit$

It is too much to ask for the whole space to be a L-cyclic subspace in every case, but if we can decompose the space into a direct sum of L-cyclic subspaces then the matrix for L with respect to bases built out of the generators will be built up of little companion matrices. Indeed, if we can write a vector space as a direct sum of L-invariant subspaces we can reduce the problem of finding a nice basis for $\mathcal{V}$ to many simpler problems of finding nice bases for the invariant subspaces. This is because of the next theorem:

Theorem 11.2.1 *Let $L : \mathcal{V} \to \mathcal{V}$ be a linear transformation and $\mathcal{V}$ have a decomposition of $\mathcal{V} = \mathcal{U}_1 \oplus \mathcal{U}_2$ where $\mathcal{U}_1$ and $\mathcal{U}_2$ are L-invariant. If $(\vec{b}_1 \dots \vec{b}_r)$ is a basis for $\mathcal{U}_1$, $(\vec{c}_1 \dots \vec{c}_s)$ is a basis for $\mathcal{U}_2$, the matrix for the restriction of L to $\mathcal{U}_1$ with respect to its basis is $\mathbf{M_1}$, and the matrix for the restriction of L to $\mathcal{U}_2$ with respect to its basis is $\mathbf{M_2}$, then the matrix for L with respect to the basis $(\vec{b}_1 \dots \vec{b}_r, \vec{c}_1 \dots \vec{c}_s)$ is the block matrix*

$$\begin{bmatrix} \mathbf{M}_1 & 0 \\ 0 & \mathbf{M}_2 \end{bmatrix}.$$

PROOF:

We obtain the matrix for L with respect to the basis

$$(\vec{b}_1 \ldots \vec{b}_r, \vec{c}_1 \ldots \vec{c}_s)$$

by finding the image of each basis vector. If we use one of the basis vectors for the basis for $\mathcal{U}_1$ the image will be in $\mathcal{U}_1$ because $\mathcal{U}_1$ is L-invariant. This means that the coefficient of $\vec{c}_i$ will be 0 for all i. Similarly if we take the image of any of the $\vec{c}_i$ it will have coefficient for each $\vec{b}_j$ equal to 0 because $\mathcal{U}_2$ is L-invariant. Thus we have a block diagonal matrix.

To see that the blocks on the diagonal are the matrices for the restrictions to the subspaces note that writing $L(\vec{b}_i)$ in terms of the basis $(\vec{b}_1, \ldots, \vec{b}_r)$ in $\mathcal{U}$ gives the same coefficients as writing it in terms of the basis for $\mathcal{V}$. ∎

In Chapter 15 we will use a decomposition into invariant subspaces to get a block diagonal form with the blocks all companion matrices. That decomposition will use inner products, but using more advanced ideas it is possible to find the *rational canonical form*, which is a block diagonal form with all blocks companion matrices, for any matrix.

Canonical forms for nilpotent transformations In some ways matrices behave quite differently than numbers. Matrix multiplication is not commutative and there are zero divisors. Indeed, it is possible to find matrices such that $\mathbf{M} \neq \mathbf{0}$ but for which there is an n with $\mathbf{M}^n = \mathbf{0}$. These matrices correspond to *nilpotent* linear transformations:

Definition 11.2.3 *A linear transformation $L : \mathcal{V} \to \mathcal{V}$ is said to be nilpotent with index of nilpotence n if L^n is always 0, but lower powers of L are not.*

Example: A nilpotent linear transformation

Consider the linear transformation $L : \mathbb{R}^3 \to \mathbb{R}^3$ given by $L([x, y, z]) = [0, x, y]$. It is not difficult to see that L^3 is identically $[0, 0, 0]$, but $L^2([1, 1, 1]) = [0, 0, 1]$. Thus L is nilpotent with index 3. ♢

Now suppose L is nilpotent with index of nilpotence n. Then there must be a vector in $\mathcal{V}$ such that $L^{n-1}(\vec{v}) \neq \vec{0}$. We get a particularly nice form if we use this $\vec{v}$ in an attempt to form an L-cyclic subspace.

Lemma 11.2.2 *If $L : \mathcal{V} \to \mathcal{V}$ is nilpotent with index of nilpotence n and if $L^{n-1}(\vec{v}) \neq \vec{0}$ then $\{\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v})\}$ is linearly independent.*

PROOF:

We will use a proof by induction on the index of nilpotence. Certainly if $n = 1$, then the set $\{\vec{v}\}$ is independent since $\vec{v} \neq \vec{0}$.

Now suppose that we have the theorem for all index of nilpotence k and we want it for index of nilpotence $k+1$. Note that if L is nilpotent with index of nilpotence $k+1$ as a linear transformation from $\mathcal{V}$ to itself then it is nilpotent with index of nilpotence k on $\mathrm{Im}(L)$. If $\vec{v}$ had $L^k(\vec{v}) \neq 0$ then $\vec{w} = L(\vec{v})$ is in $\mathrm{Im}(L)$ and has $L^{k-1}(\vec{w}) \neq 0$. Thus by the induction hypothesis the set

$$\{\vec{w}, L(\vec{w}), \ldots, L^{k-1}(\vec{w})\} = \{L(\vec{v}), \ldots, L^k(\vec{v})\}$$

is independent.

Now suppose that

$$a_0\vec{v} + a_1 L(\vec{v}) + \ldots + a_k L^k(\vec{v}) = \vec{0},$$

then

$$a_0 L(\vec{v}) + a_1 L(\vec{v}) + \ldots + a_k L^{k+1}(\vec{v}) = \vec{0}$$

as well since L is linear. Now $L^{k+1}(\vec{v}) = 0$ so this tells us that

$$a_0 L(\vec{v}) + a_1 L(\vec{v}) + \ldots + a_k L^k(\vec{v}) = \vec{0}.$$

Now since the set $\{L(\vec{v}), \ldots, L^k(\vec{v})\}$ is independent, this tells us that each a_i is 0 for $1 = 0, \ldots, k-1$. Thus our linear combination giving $\vec{0}$ reduces to

$$a_k L^k(\vec{v}) = \vec{0}.$$

Since we assumed that $L^k(\vec{v}) \neq 0$, we can conclude that $a_k = 0$ as well.

Thus, by induction, for any n, if $L : \mathcal{V} \to \mathcal{V}$ is nilpotent with index of nilpotence n and if $L^{n-1}(\vec{v}) \neq \vec{0}$ then $\{\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v})\}$ is linearly independent. ∎

Notice that the matrix for L on the subspace with basis $\{\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v})\}$ has the form

$$\begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

The next theorem tells us that a nilpotent linear transformation can be written as a block diagonal matrix with all blocks of this special form.

Theorem 11.2.3 *If $L : \mathcal{V} \to \mathcal{V}$ is nilpotent with index of nilpotence n and if $L^{n-1}(\vec{v}) \neq \vec{0}$ then $\mathcal{V} = C^L_{\vec{v}} \oplus \mathcal{V}_1$, where $\mathcal{V}_1$ is L-invariant.*

PROOF:

Direct sums are most easily obtained by taking a basis, breaking it into two pieces and looking at the subspaces spanned by the pieces. Since we already have a basis for $C^L_{\vec{v}}$, namely $\{\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v})\}$, we will extend it to a basis for $\mathcal{V}$ in such a way that the subspace spanned by the new basis vectors is L-invariant.

Because L is nilpotent with index of nilpotence n on $\mathcal{V}$, we know that $\mathcal{V}$ can be thought of as the nested family of subspaces

$$\mathcal{V} = \mathrm{Ker}(L^n) \supseteq \mathrm{Ker}(L^{n-1}) \supseteq \mathrm{Ker}(L^{n-2}) \supseteq \ldots \supseteq \mathrm{Ker}(L)$$

with L taking each subspace into the next. Now note that $L^{n-1}(\vec{v}) \in \mathrm{Ker}(L)$ and extend to get a basis $\{L^{n-1}(\vec{v}) = \vec{b}_1, \vec{b}_2, \ldots, \vec{b}_k\}$ for $\mathrm{Ker}(L)$. The set $\{\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v}), \vec{b}_2, \ldots, \vec{b}_k\}$ is linearly independent since if

$$\vec{0} = \sum_{i=2}^{k} a_i \vec{b}_i + \sum_{j=0}^{n-1} c_j L^j(\vec{v})$$

then

$$\vec{0} = L(\vec{0}) = \sum_{i=2}^{k} a_i L(\vec{b}_i) + \sum_{j=0}^{n-1} c_j L^{j+1}(\vec{v}) = \vec{0} + \sum_{j=0}^{n-2} c_j L^{j+1}(\vec{v})$$

so that all of the coefficients $c_j = 0$ for $j = 0 \ldots n-2$ by the independence of $\{\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v})\}$. Independence of the set

$\{L^{n-1}(\vec{v}), \vec{b}_2, \ldots, \vec{b}_k\}$ tells us that the rest of the coefficients must be 0.

Now continue the process by extending

$$\{L^{n-2}(\vec{v}), L^{n-1}(\vec{v}), \vec{b}_2, \ldots, \vec{b}_k\}$$

to a basis for $\mathrm{Ker}(L^2)$. We can always do this by adjoining vectors with $L(\vec{b})$ not involving $L^{n-1}(\vec{v})$. If $L(\vec{b}) = \sum_{i=1}^{k} c_i \vec{b}_i$ then $\vec{b} - c_1 L^{n-2}(\vec{v})$ can replace $\vec{b}$ in our basis without losing either independence or spanning, but gaining the L-invariance of the subspace spanned by the new basis vectors.

Working our way backwards in this fashion, we get bases for $\mathrm{Ker}(L^m)$ whose new vectors all lie in a L-invariant subspace. Eventually we will have a basis $\{\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v}), \vec{b}_2, \ldots, \vec{b}_z\}$ for $\mathcal{V}$. Now $\mathrm{Span}(\{\vec{b}_2, \ldots, \vec{b}_z\})$ is L-invariant, so we can use it as $\mathcal{V}_1$. This shows that $\mathcal{V} = C_{\vec{v}}^{L} \oplus \mathcal{V}_1$, as needed. ∎

Example: Canonical form for a nilpotent linear transformation

The linear transformation $L : \mathbb{R}^5 \to \mathbb{R}^5$ with

$$L([u, v, w, x, y]) = [0, u, v, 0, x]$$

is nilpotent with index of nilpotence 3. If we start with $\vec{v} = [1, 2, 3, 4, 5]$ we get $L(\vec{v}) = [0, 1, 2, 0, 4]$ and $L^2(\vec{v} = [0, 0, 1, 0, 0]$. We can extend

$$\{[0, 0, 1, 0, 0]\}$$

to a basis for $\mathrm{Ker}(L)$ by adjoining the vector $[0, 0, 0, 0, 1]$. We then extend

$$\{[0, 1, 2, 0, 4], [0, 0, 1, 0, 0], [0, 0, 0, 0, 1]\}$$

to a basis for $\mathrm{Ker}(L^2)$ by adjoining $[0, 1, 0, 1, 0]$. This was not the best choice since

$$L([0, 1, 0, 1, 0]) = [0, 0, 1, 0, 1] = 1[0, 0, 1, 0, 0] + 1[0, 0, 0, 0, 1],$$

so we use $[0, 1, 0, 1, 0] - [0, 1, 2, 0, 4] = [0, 0, -2, 1, -4]$ instead. This gives the basis

$$\{[0, 1, 2, 0, 4], [0, 0, 1, 0, 0], [0, 0, 0, 0, 1], [0, 0, -2, 1, -4]\}$$

for $\mathrm{Ker}(L^2)$.

We then note that

$$\{[1,2,3,4,5],[0,1,2,0,4],[0,0,1,0,0],[0,0,-2,1,-4],[0,0,0,0,1]\}$$

is a basis for $\mathbb{R}^5$. The matrix for L with respect to this basis is

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$\diamondsuit$

Exercises 11.2:

For problems 1–4, factor the matrix $\mathbf{M}$ into a lower triangular matrix followed by an upper triangular matrix and then reverse the order to find a matrix similar to $\mathbf{M}$ for each of the following matrices:

1. $\begin{bmatrix} 1 & 4 \\ 3 & 6 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 4 & 1 \\ 3 & 6 & 5 \\ 1 & 1 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 3 & 12 & 12 \\ -1 & 0 & 0 \\ 1 & 6 & 9 \end{bmatrix}$

4. $\begin{bmatrix} 4 & 8 & 12 \\ 1 & 5 & 12 \\ -1 & 0 & 9 \end{bmatrix}$

For problems 5–8 find a basis for the m-cyclic subspace generated by $\vec{v}$, if you get $C_{\vec{v}}^m = \mathcal{V}$, then give the matrix for m with respect to your basis:

5. $m : \mathbb{R}^3 \to \mathbb{R}^3$ has $m([x, y, z] = [x, x + y, x + y + z]$ and $\vec{v} = [1, 0, 0]$.
6. $m : \mathbb{R}^3 \to \mathbb{R}^3$ has $m([x, y, z] = [x, 2y, x + 3z]$ and $\vec{v} = [1, 1, 1]$.
7. $m : \mathbb{R}^3 \to \mathbb{R}^3$ has $m([x, y, z] = [x, x + y, x + 2y]$ and $\vec{v} = [1, 0, 0]$.
8. $m : \mathbb{R}^3 \to \mathbb{R}^3$ has $m([x, y, z] = [x + y, x - y, x + y + z]$ and $\vec{v} = [1, 1, 0]$.

The following linear transformations are nilpotent. Find a basis decomposing the domain into cyclic subspaces so that the matrix for L has a block diagonal form like that given in the example.

9. $L : \mathbb{R}^4 \to \mathbb{R}^4$ with $L([x, y, z, w]) = [0, x, x + w, y]$
10. $L : \mathbb{R}^4 \to \mathbb{R}^4$ with $L([x, y, z, w]) = [x - w, y + z, -y - z, x - w]$
11. $L : \mathbb{R}^5 \to \mathbb{R}^5$ with $L([x, y, z, s, t]) = [0, s, x, z, 2y + z]$
12. $L : \mathbb{R}^6 \to \mathbb{R}^6$ with $L([u, v, w, x, y, z]) = [u-v+w, u-v-w, y-z, w, x, x]$

Chapter 12

Determinants

The determinant of a square matrix is a number associated with the matrix which can be used to determine whether or not the matrix has an inverse. Several possible methods can be used to define the function det which takes the set of square matrices to the reals. One approach is to define it for 2 by 2 matrices using

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

and then find a way (called reduction by minors) to reduce all larger matrices to the 2 by 2 case. Occasionally this gives a quick way to find determinants, but for general matrices it maximizes computational effort. In any case it leads to a definition which is very hard to prove things about. Our approach is to list the properties that we want a determinant function to have, show that they provide a means to calculate the value of the determinant of a matrix efficiently, and then use the list of properties as a definition of the determinant.

12.1 Properties and Efficient Calculation

We take the following as the definition of a determinant function

$$\det : n \times n\text{-square matrices} \rightarrow \mathbb{R}.$$

Until we show that these properties allow us to compute the determinant we cannot be sure that these properties define *the* determinant rather than *a* determinant, but we will postpone that problem until later.

Definition 12.1.1 *A determinant function*

$$\det : n \times n\text{-}matrices \ \rightarrow \mathbb{R}$$

is a function with the following properties:

1. $\det$ *is multiplicative :* $\det(\mathbf{AB}) = \det \mathbf{A} \det \mathbf{B}$
2. $\det \mathbf{A}^t = \det \mathbf{A}$
3. $\det$ *is multilinear in the rows, i.e., if the matrices* $\mathbf{A}, \mathbf{B}$*, and* $\mathbf{C}$ *are identical except in row* i *and* $\mathbf{A}_{ij} = k\mathbf{B}_{ij} + h\mathbf{C}_{ij}$ *then* $\det \mathbf{A} = k \det \mathbf{B} + h \det \mathbf{C}$.
4. $\det$ *is not identically* 0.

Proposition 12.1.1 *The determinant of an identity matrix is 1.*

PROOF:

Since det is not identically 0, there is some $\mathbf{A}$ with $\det \mathbf{A} \neq 0$. We know that $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$, so by axiom 1, $\det \mathbf{A} = \det \mathbf{AI} = \det \mathbf{A} \det \mathbf{I}$. Since $\det \mathbf{A} \neq 0$ we can divide both sides by $\det \mathbf{A}$ to get $\det \mathbf{I} = 1$. ∎

Proposition 12.1.2 *If* $\mathbf{A}$ *has an inverse then* $\det \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}$ *, thus* $\det \mathbf{A}$ *cannot be 0.*

PROOF:

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

∎

Proposition 12.1.3 *The determinant of a diagonal matrix (one with 0's in all of the off diagonal positions) is the product of its diagonal entries.*

PROOF:

We use induction on the number of diagonal entries not equal to 1. If there are none then we have an identity matrix and the result is given by Proposition 12.1.1. So let us assume that the proposition holds for k non-one diagonal entries and show that it must hold for $k+1$ non-one diagonal entries. If $\mathbf{A}$ is a diagonal matrix with $k+1$ diagonal entries not equal to 1 then pick a row i in which $\mathbf{A}$ has a diagonal entry which is not 1. Let $\mathbf{A}'$ be the matrix which is identical to $\mathbf{A}$ except that its i, i-entry is a 1. Then $\det \mathbf{A} = \mathbf{a}_{ii} \det \mathbf{A}'$ by axiom 3, multilinearity. The matrix $\mathbf{A}'$ has only k non-one diagonal entries so its determinant is the product of its diagonal entries. This proves that $\det \mathbf{A}$ is the product of the diagonal entries in $\mathbf{A}$, and we are finished by induction. ∎

Next we note some cases in which we can tell by inspection that the determinant is 0.

Proposition 12.1.4 *If* $\mathbf{A}$ *has a row which is all zeros then* $\det \mathbf{A} = 0$.

PROOF:

Let $\mathbf{B}$ be identical to $\mathbf{A}$ except in the row which is all zeros. Then $\det \mathbf{B} = \det \mathbf{A} + \det \mathbf{B}$ by multilinearity in the rows. Thus $\det \mathbf{A} = 0$. ∎

Proposition 12.1.5 *If* $\mathbf{A}$ *has two rows which are the same then* $\det \mathbf{A} = 0$.

PROOF:

Suppose rows i and j of $\mathbf{A}$ are identical. Then multiplying $\mathbf{A}$ by the matrix corresponding to the elementary row operation "add -1 times row i to row j," which has an inverse and thus has nonzero determinant, gives a matrix with a row of zeros. Thus $\det \mathbf{A} = 0$. ∎

Proposition 12.1.6 *If* $\mathbf{A}$ *is the result of interchanging two rows in an identity matrix then* $det \mathbf{A} = -1$.

Example:

The essential features of the argument are found in the example

$$\det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -1.$$

This is demonstrated by observing that

$$\det \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + \det \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

by multilinearity. Continuing in the same manner we get

$$\det \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \det \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + \det \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} + \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

$$0 = 1 + 0 + 0 + \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

This tells us that

$$\det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = -1$$

using Propositions 12.1.4 and 12.1.5. The general case just adds a pair of indices to keep track of which two rows are being interchanged. ♢

This gets us some of the easy cases. To see how to progress further we need to see what elementary row operations do to determinants. If we recall that elementary row operations can be done by multiplying on the left by appropriate matrices we can reduce our work to calculating the determinants of those matrices.

Proposition 12.1.7 *The elementary row operations have the following effect on the determinant of* $\mathbf{A}$*:*

1. *multiplying row* i *by* r *multiplies the determinant by* r
2. *interchanging two rows multiplies the determinant by -1*
3. *adding a multiple of one row to another does not change the determinant.*

PROOF:

1) The elementary row operation of multiplying row i by r is effected by left multiplication by a matrix with r in the ii position and 1 elsewhere on the diagonal and all other entries 0. By Proposition 12.1.3 this has determinant r. Thus multiplying by it multiplies the determinant of $\mathbf{A}$ by r.

2) Interchanging rows i and j is accomplished by multiplying on the left by a matrix obtained by interchanging rows i and j of an identity matrix. Such a matrix has determinant -1. Thus interchanging two rows changes the sign of the determinant.

3) Adding a multiple of one row to another is accomplished by left multiplication by the matrix obtained by performing the same elementary row operation on an identity matrix. For example, adding m times row 1 to row 3 can be achieved by multiplying by the matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m & 0 & 1 \end{bmatrix}$$

Now by linearity in row 3 we get

$$\begin{aligned} \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m & 0 & 1 \end{bmatrix} &= m \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} + \det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= m0 + 1 = 1. \end{aligned}$$

So this operation leaves the determinant unchanged. ∎

Corollary 12.1.8 *A square matrix is invertable if and only if its determinant is nonzero.*

PROOF:

None of the row operations changes the determinant from zero to non-zero or from non-zero to zero. Thus a matrix has a non-zero determinant if and only if its row reduced echelon form has a non-zero determinant. From our algorithm for finding inverses we know that the $\mathbf{M}^{-1}$ exists if and only if the row reduced echelon form for $\mathbf{M}$ is the identity matrix, which has determinant equal to 1, hence non-zero. ∎

The work so far lets us calculate determinants for all cases. For example we calculate

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

as follows:

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \det \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix} = -2\det \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = -2\det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = -2.$$

Another example is

$$\det \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 3 \end{bmatrix} = \det \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = 0.$$

Yet another

$$\det \begin{bmatrix} 1 & 0 & 2 & 1 \\ 2 & 1 & 4 & 6 \\ 3 & 0 & 1 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & -5 & 5 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

$$= \det \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & -5 & 5 \\ 0 & 0 & -1 & -4 \end{bmatrix} = -5\det \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & -4 \end{bmatrix}$$

$$= -5\det \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -5 \end{bmatrix} = 25\det \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= 25\det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = 25.$$

In each case we calculate the determinant by keeping track of the effect of the row reduction operations as we reduce the matrix to row reduced echelon form. This always leads to a case we can solve by inspection because the row reduced echelon form of a square matrix either has a row of zeros or is an identity matrix. Actually we can stop before we get that far using the next proposition.

Proposition 12.1.9 *If* $\mathbf{A}$ *is an upper triangular matrix then* $\det \mathbf{A}$ *is the product of the diagonal entries.*

PROOF:

If $\mathbf{A}$ has no diagonal entries equal to 0 then we can use elementary row operations of the form "add a multiple of row i to row j" to eliminate all of the off diagonal entries. Since that form of elementary row operation does not change the determinant this tells us $\det \mathbf{A} = \det \mathbf{A_{diag}}$ where $\mathbf{A_{diag}}$ has the same diagonal entries as $\mathbf{A}$ and zeros off the diagonal. If $a_{ii} = 0$ for some i then consider the last row with $a_{ii} = 0$. All of the entries a_{ij} with $j < i$ are 0 and all of the entries a_{jj} with $j > i$ are nonzero. Since a_{jj} is nonzero when $j > i$ we can use the elementary row operation "Add $-a_{ij}/a_{jj}$ times row j to row i" to eliminate the a_{ij} entry. Thus applying row operations which do not change the determinant will result in a matrix with a row of zeros. Thus $\det \mathbf{A} = 0$.
∎

Examples:

$$\det \begin{bmatrix} 2 & 1 & 3 \\ 0 & 4 & 1 \\ 0 & 0 & 6 \end{bmatrix} = 48$$

$$\det \begin{bmatrix} 2 & 1 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 6 \end{bmatrix} = 0$$

Matrices can sometimes be thought of as being built of blocks. For instance we can think of the matrix

$$\mathbf{M} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 4 & 5 & 6 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

as being

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \\ \mathbf{B} &= \begin{bmatrix} 3 & 4 & 5 \\ 1 & 1 & 1 \end{bmatrix} \\ \mathbf{C} &= \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 1 & 1 & 1 \end{bmatrix} \\ \mathbf{0} &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

This particular example is block upper triangular because of the block of 0's in the lower left corner.

Theorem 12.1.10 *If the matrix* $\mathbf{M}$ *can be partitioned into four blocks*

$$\mathbf{M} = \mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{bmatrix}$$

where $\mathbf{A}$ *and* $\mathbf{C}$ *are square and* $\mathbf{0}$ *has all entries 0, then* $\det(\mathbf{M}) = \det(\mathbf{A})\det(\mathbf{C})$.

PROOF:

To calculate the determinant of $\mathbf{M}$ we do row operations to reduce it to upper triangular form. In this process we first do row operations to get the rows in blocks $\mathbf{A}$ and $\mathbf{B}$ to be upper triangular, then we deal with the rows in $\mathbf{C}$. This gives a block upper triangular matrix

$$\mathbf{M}' = \mathbf{M} = \begin{bmatrix} \mathbf{A}' & \mathbf{B}' \\ \mathbf{0} & \mathbf{C}' \end{bmatrix}$$

with both $\mathbf{A}'$ and $\mathbf{C}'$ upper triangular. The determinant $\det(\mathbf{M})$ is then given by the product of the determinants for the elementary row operations times the product of the diagonal elements of $\mathbf{M}'$. We can break this into the product of the determinants to get the first rows upper triangular times the diagonal elements of $\mathbf{A}'$ times the product of the determinants of the elementary matrices to get the last rows in upper triangular form times the diagonal elements of $\mathbf{C}'$. But this is just $\det(\mathbf{A})\det(\mathbf{C})$, as claimed. ∎

Exercises 12.1:
For problems 1–11 find the determinants using row reduction or theorems from this section:

1. $\det \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 4 \\ 7 & 6 & 5 \end{bmatrix}$

2. $\det \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 5 \end{bmatrix}$

3. $\det \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$

4. $\det \begin{bmatrix} 1 & 2 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

5. $\det \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 4 & 2 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 8 \end{bmatrix}$

6. $\det \begin{bmatrix} 1 & 2 & 3 & 0 \\ 3 & 1 & 4 & 0 \\ 6 & 0 & 5 & 0 \\ 0 & 0 & 0 & 17 \end{bmatrix}$

7. $\det \begin{bmatrix} 1 & 0 & 0 & 2 \\ 3 & 1 & 4 & 2 \\ 6 & 0 & 5 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

8. $\det \begin{bmatrix} -1 & 0 & -2 & 3 \\ 2 & 1 & 4 & 5 \\ 1 & 7 & 3 & 0 \\ 1 & -1 & -1 & -1 \end{bmatrix}$

9. $\det \begin{bmatrix} 0 & 0 & 0 & a \\ 1 & 0 & 0 & b \\ 0 & 1 & 0 & c \\ 0 & 0 & 1 & d \end{bmatrix}$

10. $\det \begin{bmatrix} 0 & 0 & 0 & d \\ a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \end{bmatrix}$

11. $\det \begin{bmatrix} 0 & 0 & 0 & a \\ 0 & 0 & b & 0 \\ 0 & c & 0 & 0 \\ d & 0 & 0 & 0 \end{bmatrix}$

12. (Project Problem)

    (a) Prove that if $\mathbf{M}$ is $n \times n$ then

    $$\det \mathbf{M} = \sum_{k=1}^{n} \det \hat{\mathbf{M}}_k$$

    where $\hat{\mathbf{M}}_k$ is identical to $\mathbf{M}$ except that all entries other than the k^{th} in row 1 are 0.

    (b) Use an induction argument together with part (a) to show that the $\det \mathbf{M}$ is the sum of all the determinants of matrices which can be made from $\mathbf{M}$ by changing all entries to 0 except for one entry in each row and column.

    (c) Show that if $\mathbf{M}$ has only one nonzero entry in each row and column, say e_i in column i, then $\det \mathbf{M}$ is $(-1)^p \prod e_i$ where p is then number of columns in which the nonzero entry is not the main diagonal entry. If we write $e_i = m_{\sigma(i)i}$ for a permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ then $(-1)^p$ is called the parity of σ, written $\operatorname{sgn}(\sigma)$.

    (d) Using the earlier parts of this problem prove that

    $$\det M = \sum_{\sigma} (\operatorname{sgn}(\sigma) \prod_{i=1}^{n} m_{\sigma(i)i})$$

12.2 Expansion by Minors

In the previous section we found determinants using row reduction. This is the computationally efficient way to solve the problem of computing a determinant. In this section we give another, iterative, approach which is widely used for small matrices, particularly when determinants are used outside mathematics. We start with the definition $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$. We then reduce the calculation of an $n \times n$ determinant to n calculations of $(n-1) \times (n-1)$ determinants by the following procedure. In order to define det for $n \times n$ matrices, we will reduce the problem to $n \times n$ matrices of a very special form: those for which the only non zero entry in the first row or column is in the 1,1 position.

$$\begin{bmatrix} a & 0 \; 0 \; \cdots \; 0 \\ 0 & \\ 0 & \\ \vdots & \mathbf{A_1} \\ 0 & \end{bmatrix} = \mathbf{A}$$

For such a matrix, $\det \mathbf{A} = a \det \mathbf{A_1}$. To see this note that the determinant of $\mathbf{A_1}$ can be found by applying elementary row operations until it is upper triangular and then multiplying the diagonal elements. Row i of $\mathbf{A_1}$ appears with a 0 appended in the first column as row $i+1$ of $\mathbf{A}$, so if we add 1 to all of the row references in the process of reducing $\mathbf{A_1}$ to upper triangular form we get the operations needed to reduce $\mathbf{A}$ to upper triangular form. Since a is the diagonal element this tells us that $\det \mathbf{A} = a \det \mathbf{A_1}$. Now suppose we are given an $n \times n$ matrix $\mathbf{M} = [[m_{ij}]]$. For each pair (i, j) let $\mathbf{M_{ij}}$ be the $(n-1) \times (n-1)$ matrix obtained by crossing out the i^{th} row and the j^{th} column of $\mathbf{M}$. This is called the ij minor. We will show that

$$\det \mathbf{M} = \sum_{j=1}^{n} (-1)^{j+1} m_{1j} \det \mathbf{M_{1j}}.$$

Consider the matrices $\mathbf{A_j}$ which are identical with $\mathbf{M}$ except that all of the entries in the first row except the j^{th} have been replaced by zeros. By

multilinearity, we must have

$$\det \mathbf{M} = \sum_{j=1}^{n} \det \mathbf{A_j}.$$

Since we know that adding a multiple of one row to another does not change determinants, each the $\mathbf{A_j}$ may be replaced by the matrix $\mathbf{B_j}$ which has all entries in the j^{th} column except the first changed to zero as well. These matrices $\mathbf{B_j}$ would look like the special type of matrices for which we know the determinant, if only we were to exchange columns until the non zero entry in the first row were also in the first column. This involves $j-1$ interchanges. If we call the resulting matrix $\mathbf{C_j}$ then

$$\mathbf{C_j} = \begin{bmatrix} m_{i,j} & 0 & 0 & \cdots & 0 \\ 0 & & & & \\ 0 & & & & \\ \vdots & & \mathbf{M_{1,j}} & & \\ 0 & & & & \end{bmatrix}$$

and

$$\det \mathbf{C_j} = (-1)^{j-1} \det \mathbf{B_j}.$$

This tells us that $\det \mathbf{C_j} = m_{1j} \det \mathbf{M_{1j}}$. In summary we have to have the following calculations:

$$\det \mathbf{M} = \sum_{j=1}^{n} \det \mathbf{A_j} = \sum_{j=1}^{n} \det \mathbf{B_j} = \sum_{j=1}^{n} (-1)^{j-1} \det \mathbf{C_j} = \sum_{j=1}^{n} (-1)^{j-1} m_{1j} \det \mathbf{M_{1j}}.$$

In general we can use row i instead of row 1 by interchanging rows $i-1$ times to get it to the row 1 position (multiplying by $(-1)^{i-1}$) and then expanding by minors:

$$\det \mathbf{A} = \sum_{j=1}^{n} (-1)^{i-1}(-1)^{j-1} a_{ij} \det \mathbf{A_{ij}} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det \mathbf{A_{ij}}.$$

We can also use the fact that $\det \mathbf{A} = \det \mathbf{A}^t$ to get expansion using a column instead of a row:

$$\det \mathbf{A} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det \mathbf{A_{ij}}.$$

Example: Calculating determinants

$$\begin{aligned}\det\begin{bmatrix}1&2&0\\0&0&1\\0&0&3\end{bmatrix} &= 1\det\begin{bmatrix}0&1\\0&3\end{bmatrix}\\ &= 1\cdot 0 = 0\end{aligned}$$

$\diamondsuit$

Example: A 4 by 4 example

$$\det\begin{bmatrix}1&0&2&1\\2&1&4&6\\3&0&1&8\\1&1&1&1\end{bmatrix}$$

$$= 1\det\begin{bmatrix}1&4&6\\0&1&8\\1&1&1\end{bmatrix} - 0\det\begin{bmatrix}2&4&6\\3&1&8\\1&1&1\end{bmatrix} + 2\det\begin{bmatrix}2&1&6\\3&0&8\\1&1&1\end{bmatrix} - 1\det\begin{bmatrix}2&1&4\\3&0&1\\1&1&1\end{bmatrix}$$

$$= 1(1\det\begin{bmatrix}1&8\\1&1\end{bmatrix} - 4\det\begin{bmatrix}0&8\\1&1\end{bmatrix} + 6\det\begin{bmatrix}0&1\\1&1\end{bmatrix}) - 0$$

$$+2(2\det\begin{bmatrix}0&8\\1&1\end{bmatrix} - 1\det\begin{bmatrix}3&8\\1&1\end{bmatrix} + 6\det\begin{bmatrix}3&0\\1&1\end{bmatrix})$$

$$-1(2\det\begin{bmatrix}0&1\\1&1\end{bmatrix} - 1\det\begin{bmatrix}3&1\\1&1\end{bmatrix} + 4\det\begin{bmatrix}3&0\\1&1\end{bmatrix})$$

$$= 1(1(1-8) - 4(0-8) + 6(0-1)) - 0$$

$$+2(2(0-8)-1(3-8)+6(3-0))-1(2(0-1)-1(3-1)+4(3-0)) = 25.$$

$\diamondsuit$

This same example was also done in the previous section using elementary row operations. Compare the amount of work involved.

Exercises 12.2:
Rework exercises 1-11 of the previous section using expansion minors. Use any convenient row or column.

12. A tridiagonal matrix has nonzero entries only on the main diagonal and in the diagonals directly above and below the main diagonal. Count how many multiplications are needed to find the determinant of an $n \times n$ tridiagonal matrix using row reduction to an upper triangular matrix. Then count the number of multiplications needed using expansion by minors. Which method is preferable for these matrices?

Chapter 13

Eigenvalues and Eigenvectors

In this chapter we will turn our attention to a very important problem involving linear transformations, that of finding eigenvalues and eigenvectors. We will see that eigenvalues tell us what the long term behavior of iterations of a linear transformation are. We will also see how they enable us to find bases giving particularly nice forms for the matrix of a linear transformation.

13.1 Eigenvalues and Characteristic Polynomials

The simplest example of a linear transformation is one obtained from a scalar λ by taking each vector $\vec{v}$ to $\lambda\vec{v}$. The matrix of this linear transformation with respect to any basis (using the same basis for both domain and codomain) is $\lambda\mathbf{I}$. Certainly we would not ask for every linear transformation to look just like this simplest case, but it does turn out to be useful to ask which vectors have scalars λ for which the linear transformation looks like multiplication by λ. If we can find a basis of such vectors we will be able to represent the linear transformation using a diagonal matrix. We need some terminology for this situation.

Definition 13.1.1 *An* **eigenvalue** *for a linear transformation $T : \mathcal{V} \to \mathcal{V}$ is a scalar λ such that there is a nonzero $\vec{v} \in \mathcal{V}$ with $T(\vec{v}) = \lambda\vec{v}$. The vector $\vec{v}$ is called an* **eigenvector** *for the eigenvalue λ.*

Definition 13.1.2 *An eigenvalue for a matrix $\mathbf{T}$ is a scalar λ so that there is a nonzero column vector $\vec{v}$ with $\mathbf{T}\vec{v} = \lambda\vec{v}$.*

It is clear that the eigenvalues of a linear transformation and the eigenvalues of the matrix representing it with respect to a basis are the same. This means that we can find eigenvalues for linear transformations by finding eigenvalues for matrices. It also means that if two different matrices represent the same linear transformation with respect to different bases, then they will have the same eigenvalues, since the eigenvalues are properties of the transformation and not artifacts of how we represent it. Some of the utility of eigenvalues arises from the nice forms that they let us find for matrices–upper triangular forms and for very nice transformations, diagonal forms.

Example:

The linear transformation $L : \mathbb{R}^2 \to \mathbb{R}^2$ taking $[x, y]$ to $[2x, 3y]$ has eigenvalues 2 and 3. The vector $[1, 0]$ is an eigenvector for the eigenvalue 2, and $[0, 1]$ is an eigenvector for 3. These eigenvectors form a basis. If we use the basis of eigenvectors, the matrix for L is the diagonal matrix

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

♢

It is not always so easy to tell what the eigenvalues of a linear transformation are. It is, however, easy to check that a particular vector is an eigenvector for a given eigenvalue.

Example:

The scalars 2 and 1 are eigenvalues for the linear transformation with matrix

$$\begin{bmatrix} .8 & .3 \\ -.8 & 2.2 \end{bmatrix}$$

though this is far from obvious from looking at the matrix. If, however, we are told that $[1, 4]$ is an eigenvector for 2 and $[3, 2]$ is an eigenvector for 1, we can check:

$$\begin{bmatrix} .8 & .3 \\ -.8 & 2.2 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$$

and

$$\begin{bmatrix} .8 & .3 \\ -.8 & 2.2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

In fact, all we really needed to be told was that 2 and 1 were eigenvalues. We can find eigenvectors by looking at systems of equations:

$$\begin{array}{rcl} .8x + .3y & = & 2x \\ -.8x + 2.2y & = & 2y \end{array}$$

which says that

$$\begin{array}{rcl} -1.2x + .3y & = & 0 \\ -.8x + .2y & = & 0. \end{array}$$

We can solve this system

$$\begin{bmatrix} -1.2 & .3 & 0 \\ -.8 & .2 & 0 \end{bmatrix} \overset{\frac{1}{-1.2}R_1}{\leadsto} \begin{bmatrix} 1 & -.25 & 0 \\ -.8 & .2 & 0 \end{bmatrix} \overset{R_2 - .8R_1}{\leadsto} \begin{bmatrix} 1 & -.25 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

so we have nontrivial solutions. One such solution can be given by taking $y = 4$, from which we get $x = 1$. A similar computation would produce an eigenvector for the eigenvalue 1. $\diamondsuit$

This tells us that we can reduce the problem of finding an eigenvalue and an eigenvector to that of finding the eigenvalue first and then using it to find the eigenvector.

If we can find the eigenvalue λ for a matrix then we can find the eigenvector by looking at a system of equations arising from $\mathbf{M}\vec{v} = \lambda\vec{v}$. We do this by subtracting $\lambda\vec{v}$ from both sides to get a homogeneous system of linear equations. This suggests a way to find eigenvalues: see when the system $\mathbf{M}\vec{v} = \lambda\vec{v}$ has non-trivial solutions. This equation has $\vec{v}$ on both sides so we subtract $\lambda\vec{v}$ to get $\mathbf{M}\vec{v} - \lambda\vec{v} = \vec{0}$. Writing this system in matrix form gives $(\mathbf{M} - \lambda\mathbf{I})\vec{v} = \vec{0}$. This system of equations will have nontrivial solutions if and only if $(\mathbf{M} - \lambda\mathbf{I})$ does not have an inverse; that is, when its determinant is zero.

Definition 13.1.3 *The* **characteristic polynomial** *of the matrix* $\mathbf{M}$ *is* $p_{\mathbf{M}}(\lambda) = \det(\mathbf{M} - \lambda\mathbf{I})$.

Example:

The characteristic polynomial of the matrix

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \text{ is } \det \begin{bmatrix} 2-\lambda & 0 \\ 0 & 3-\lambda \end{bmatrix}$$

which is $(2-\lambda)(3-\lambda) = 6-5\lambda+\lambda^2$. The roots of the characteristic polynomial are 2 and 3, the eigenvalues of the matrix. $\diamondsuit$

Example:

The characteristic polynomial of the matrix

$$\begin{bmatrix} .8 & .3 \\ -.8 & 2.2 \end{bmatrix} \text{ is } \det \begin{bmatrix} .8-\lambda & .3 \\ -.8 & 2.2-\lambda \end{bmatrix}$$

which is $\lambda^2 - 3\lambda + 2$. The roots of the characteristic equation are the eigenvalues 2 and 1. $\diamondsuit$

The examples illustrate the method used to find eigenvalues, at least for 2 by 2 and 3 by 3 matrices. In theory the method would work well for larger matrices as well, but in practice it can be very difficult to find the roots of the characteristic polynomial. It is also easy to find examples of matrices which do not have real eigenvalues, since there are lots of polynomials with no real roots.

Example: Complex eigenvalues

As an example, consider the rotation by 45 degrees, which has matrix with respect to the standard basis

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

and characteristic polynomial

$$\lambda^2 - 2\lambda + 1$$

which has no real roots. It does, however have two complex roots. Thus if we consider the matrix as representing a linear transformation from $\mathbb{C}^2$ to $\mathbb{C}^2$ instead of from $\mathbb{R}^2$ to $\mathbb{R}^2$ there will be eigenvalues. $\diamondsuit$

Indeed the fundamental theorem of algebra (which says that any polynomial over the complex numbers factors into linear pieces) tells us that we can always find n eigenvalues for an n by n matrix over the complex numbers. The usual proofs of the fundamental theorem of algebra are, however, both beyond the scope of this course and nonconstructive. Still, when working problems involving eigenvalues we usually work over the complex numbers.

Example: Finding complex eigenvalues

Find the eigenvalues (in the complex numbers) for the matrix

$$\begin{bmatrix} 2 & 4 \\ 6 & -8 \end{bmatrix}.$$

We find the characteristic polynomial

$$\begin{aligned} \det \begin{bmatrix} 2-\lambda & 4 \\ 6 & -8-\lambda \end{bmatrix} &= (2-\lambda)(-8-\lambda) - 24 \\ &= -16 + 6\lambda + \lambda^2 - 24 \\ &= \lambda^2 + 6\lambda - 40. \end{aligned}$$

This has roots

$$\lambda = \frac{-6 \pm \sqrt{36+160}}{2}$$

or $\lambda = -10$ or 4. $\diamondsuit$

Example:

Find the eigenvalues of

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

by taking

$$\det \begin{bmatrix} 1-\lambda & -1 \\ 1 & 1-\lambda \end{bmatrix}$$

to get $(1-\lambda)(1-\lambda) + 1 = 2 - 2\lambda + \lambda^2$, which has roots

$$\lambda = \frac{2 \pm \sqrt{4-8}}{2} = 1 \pm i.$$

$\diamondsuit$

Exercises 13.1:

1. For the following matrices $\mathbf{M}$ show that the vector $\vec{v}$ is an eigenvector:

   (a) $\mathbf{M} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

   (b) $\mathbf{M} = \begin{bmatrix} 5 & -3 & -9 \\ -3 & 5 & 9 \\ 3 & -3 & -7 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

2. Given that λ is an eigenvalue for the matrix $\mathbf{M}$ find an eigenvector with eigenvalue λ:

   (a) $\lambda = 3$ $\quad\mathbf{M} = \begin{bmatrix} 3 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$

   (b) $\lambda = 3$ $\quad\mathbf{M} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -6 \\ 0 & 1 & 5 \end{bmatrix}$

   (c) $\lambda = 2$ $\quad\mathbf{M} = \begin{bmatrix} -1 & -9 & 0 \\ 1 & 5 & 0 \\ -4 & -12 & 1 \end{bmatrix}$

3. Find the characteristic polynomials of the following matrices:

   (a) $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{bmatrix}$

   (b) $\begin{bmatrix} 0 & 0 & 3 \\ 1 & 0 & 1 \\ 0 & 1 & -3 \end{bmatrix}$

   (c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 6 \\ 0 & 8 & 9 \end{bmatrix}$

   (d) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

4. Find the eigenvalues of the following matrices and give an eigenvector for each eigenvalue you find:

   (a) $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{bmatrix}$

   (b) $\begin{bmatrix} 0 & 0 & 3 \\ 1 & 0 & 1 \\ 0 & 1 & -3 \end{bmatrix}$

   (c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 6 \\ 0 & 8 & 9 \end{bmatrix}$

5. Prove that if λ is an eigenvalue for $\mathbf{M}$ with eigenvector $\vec{V}$ then $\lambda - r$ is an eigenvalue for $\mathbf{M} - r\mathbf{I}$ with eigenvector $\vec{v}$.

6. Suppose you know the eigenvalues of $\mathbf{M}$; how do you get the eigenvalues for $\mathbf{M}^n$?

7. How are the eigenvalues of $\mathbf{M}$ related to the eigenvalues of $\mathbf{M}^{-1}$?

8. Show that if 0 is an eigenvalue of L then L is not invertable. Does the converse also hold?

9. Show that if L is nilpotent (i.e., $L^n = 0$ for some n) then 0 is the only eigenvalue for L.

10. Show that applying elementary row operations to a matrix will, in general, change its eigenvalues.

11. Show that similar matrices have the same eigenvalues.

13.2 Eigenvalues and Special Forms of Matrices

Because of the properties of determinants there are certain forms of matrices for which the eigenvalues can be determined by inspection. It is not hard to see that a diagonal matrix has its diagonal entries as eigenvalues: the

eigenvectors are the corresponding standard basis elements. It turns out that the eigenvalues of an upper triangular matrix are also given by the diagonal elements.

Theorem 13.2.1 *If* $\mathbf{A}$ *is upper triangular then the eigenvalues of* $\mathbf{A}$ *are the diagonal entries* a_{ii}.

PROOF:

The characteristic polynomial is $\det(\mathbf{A} - \lambda\mathbf{I})$. The matrix $\mathbf{A} - \lambda\mathbf{I}$ is upper triangular, so its determinant is the product of its diagonal entries. This tells us that the characteristic polynomial of $\mathbf{A}$ is the product of linear factors of the form $(a_{ii} - \lambda)$. Thus the roots of the characteristic polynomial, the eigenvalues, are the diagonal entries. ∎

Example: An upper triangular matrix

The eigenvalues of the matrix

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & 7 & 195 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

are 1,-4, and -1, with 1 repeated. ♢

It is easiest to find eigenvalues of 2 by 2 matrices since we can always use the quadratic formula to solve the characteristic equation. The algebra involved in finding eigenvalues of larger matrices increases precipitously as the size of the matrix increases. One approach to larger matrices is to partition them into smaller matrices. If there is a block of zeros in the lower left hand corner of the matrix we get the following result:

Theorem 13.2.2 *If the matrix* $\mathbf{M}$ *can be partitioned into four blocks*

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{bmatrix}$$

where $\mathbf{A}$ *and* $\mathbf{C}$ *are square matrices, then the characteristic polynomial of* $\mathbf{M}$ *is the product of the characteristic polynomials of* $\mathbf{A}$ *and* $\mathbf{C}$.

PROOF:

The matrix $\mathbf{M}-\lambda\mathbf{I}$ also has a block structure with $\mathbf{A}-\lambda\mathbf{I}$ and $\mathbf{C}-\lambda\mathbf{I}$ in the diagonal positions. To calculate the determinant of $\mathbf{M}-\lambda\mathbf{I}$ we apply row operations to get it to upper triangular form. The row operations applied to $\mathbf{A}-\lambda\mathbf{I}$ to make it upper triangular are the same as the ones used to make the first block of columns of $\mathbf{M}-\lambda\mathbf{I}$ look upper triangular and they do not affect the lower right hand block. The row operations used to complete the upper triangularization of $\mathbf{M}-\lambda\mathbf{I}$ make $\mathbf{C}-\lambda\mathbf{I}$ upper triangular and do not change the work already done. This means that the calculation of $\det(\mathbf{M}-\lambda\mathbf{I})$ proceeds by making $\mathbf{A}-\lambda\mathbf{I}$ and $\mathbf{C}-\lambda\mathbf{I}$ upper triangular and then multiplying their diagonal entries. That means that $\det(\mathbf{M}-\lambda\mathbf{I}) = \det(\mathbf{A}-\lambda\mathbf{I})\det(\mathbf{C}-\lambda\mathbf{I})$. This says that the characteristic polynomial of $\mathbf{M}$ is the product of the characteristic polynomials of $\mathbf{A}$ and $\mathbf{C}$. ∎

Corollary 13.2.3 *Under the same hypotheses as the previous theorem, the eigenvalues of* $\mathbf{M}$ *are those of* $\mathbf{A}$ *together with those of* $\mathbf{C}$.

PROOF:

The roots of $\det(\mathbf{A}-\lambda\mathbf{I})\det(\mathbf{C}-\lambda\mathbf{I})$ will be the roots of either factor. Thus the eigenvalues of $\mathbf{M}$ are those of $\mathbf{A}$ and those of $\mathbf{C}$. ∎

Example: A block diagonal matrix

The eigenvalues of

$$\begin{bmatrix} 2 & 1 & 3 & 4 \\ 1 & 2 & -5 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 2 & 2 \end{bmatrix}$$

are those of

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \text{ and } \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}$$

They are the roots of $\lambda^2-4\lambda+3$ and those of $\lambda^2-5\lambda+4$. These give eigenvalues 1 and 3 and 1 and 4, respectively. ♢

At the end of Chapter 9 we considered companion matrices for m on m-cyclic subspaces. These had 1's below the diagonal and no other non-zero entries except in the last column. We can read the characteristic polynomial of such a matrix out of the entries in the last column:

Theorem 13.2.4 *If* $\mathbf{M}$ *is a companion matrix with final column*

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix}$$

then

$$\chi_{\mathbf{M}}(\lambda) = (-1)^{n+1}(a_0 + a_1\lambda + a_2\lambda^2 + \ldots + a_{n-1}\lambda^{n-1} - \lambda^n).$$

PROOF:

This is actually just a calculation of $\det(\mathbf{M}-\lambda\mathbf{I})$; the question is how best to take advantage of the special form. What we do is to make judicious use of row operations and then expand by minors using the first row. We start with the determinant

$$\det \begin{bmatrix} -\lambda & 0 & 0 & 0 & \ldots & a_0 \\ 1 & -\lambda & 0 & 0 & \ldots & a_1 \\ 0 & 1 & -\lambda & 0 & \ldots & a_2 \\ \vdots & & & & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & -\lambda & a_{n-2} \\ 0 & 0 & 0 & \ldots & 1 & (a_{n-1}-\lambda) \end{bmatrix}$$

and then systematically use the 1 below the diagonal to get rid of the $-\lambda$ on the diagonal. These row operations do not change the value of the determinant, so

$$\det(\mathbf{M}-\lambda\mathbf{I}) = \det \begin{bmatrix} -\lambda & 0 & 0 & 0 & \ldots & a_0 \\ 1 & -\lambda & 0 & 0 & \ldots & a_1 \\ 0 & 1 & -\lambda & 0 & \ldots & a_2 \\ \vdots & & & & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & a_{n-2} + a_{n-1}\lambda - \lambda^2 \\ 0 & 0 & 0 & \ldots & 1 & (a_{n-1}-\lambda) \end{bmatrix}$$

$$= \det \begin{bmatrix} -\lambda & 0 & 0 & 0 & \dots & & a_0 \\ 1 & -\lambda & 0 & 0 & \dots & & a_1 \\ \vdots & & & & \ddots & & \vdots \\ 0 & \dots & 1 & 0 & & & a_{n-3} + a_{n-2}\lambda + a_{n-1}\lambda^2 - \lambda^3 \\ 0 & 0 & \dots & 1 & 0 & & a_{n-2} + a_{n-1}\lambda - \lambda^2 \\ 0 & 0 & 0 & \dots & 1 & & (a_{n-1} - \lambda) \end{bmatrix}$$

$$= \det \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & a_0 + a_1\lambda + a_2\lambda^2 + \dots + a_{n-1}\lambda^{n-1} - \lambda^n \\ 1 & 0 & 0 & 0 & \dots & a_1 + a_2\lambda + \dots + a_{n-1}\lambda^{n-2} - \lambda^{n-1} \\ \vdots & & & & \ddots & \vdots \\ 0 & \dots & 1 & 0 & & a_{n-3} + a_{n-2}\lambda + a_{n-1}\lambda^2 - \lambda^3 \\ 0 & 0 & \dots & 1 & 0 & a_{n-2} + a_{n-1}\lambda - \lambda^2 \\ 0 & 0 & 0 & \dots & 1 & (a_{n-1} - \lambda) \end{bmatrix}$$

This last determinant can be expanded by minors using the first row to get

$$\det(\mathbf{M} - \lambda\mathbf{I}) = (-1)^{n+1}(a_0 + a_1\lambda + a_2\lambda^2 + \dots + a_{n-1}\lambda^{n-1} - \lambda^n).$$

∎

Eigenvalues can be easy to find for matrices in special forms; similarly, special forms for matrices for a linear transformation can be found by using eigenvectors in a basis. In particular, if there is a basis consisting entirely of eigenvectors then we can represent the linear transformation by a diagonal matrix. It is therefore of possible utility to study linear independence for eigenvectors.

Theorem 13.2.5 *If the vectors $\vec{v}_1, \dots, \vec{v}_m$ are eigenvectors for the distinct eigenvalues $\lambda_1, \dots, \lambda_m$ for the linear transformation L, then the set $\{\vec{v}_1, \dots, \vec{v}_m\}$ is linearly independent.*

PROOF:

Suppose that the set of $\vec{v}$'s is dependent. Then there is a first $\vec{v}_n$ which can be written as a linear combination of the previous $\vec{v}$'s. By taking the first we guarantee that the set $\{\vec{v}_1, \dots, \vec{v}_{n-1}\}$ is independent, so there is only one way to write $\vec{v}_n$ as a linear combination of the others, say

$$\vec{v}_n = \sum_{i=1}^{n-1} a_i \vec{v}_i.$$

Applying the transformation L we get

$$L(\vec{v}_n) = \sum_{i=1}^{n-1} a_i L(\vec{v}_i).$$

Using the fact that the $\vec{v}$'s are eigenvectors this gives

$$\lambda_n \vec{v}_n = \sum_{i=1}^{n-1} a_i \lambda_i \vec{v}_i.$$

If $\lambda_n = 0$ this shows that the set $\{\vec{v}_1, \ldots \vec{v}_n\}$ is not independent, so we may assume that $\lambda_n \neq 0$. Dividing by λ_n we get

$$\vec{v}_n = \sum_{i=1}^{n-1} \frac{a_i \lambda_i}{\lambda_n} \vec{v}_i$$

which is a different way of writing $\vec{v}_n$ as a linear combination of the previous $\vec{v}$'s since $\lambda_i \neq \lambda_n$ for all i. This contradicts the uniqueness of the representation, so our original set of vectors must have been linearly independent. ▮

Corollary 13.2.6 *If a linear transformation from an n-dimensional space to itself has n distinct eigenvalues then it can be represented by a diagonal matrix.*

Proof:

Use a set consisting of one eigenvector for each eigenvalue for the basis. ▮

We cannot always find a basis so that the matrix for a linear transformation is diagonal–if eigenvalues are repeated there may or may not be a basis of eigenvectors. We can, however, always get an upper triangular matrix by proper choice of basis.

Theorem 13.2.7 *For any $n \times n$ complex matrix $\mathbf{M}$ there is an upper triangular matrix $\mathbf{U}$ which is similar to $\mathbf{M}$.*

Proof:

[Following Wesson, *Lessons in Linear Algebra*, Merrill,1974] We argue by induction on n.

If $n = 1$, then $\mathbf{M}$ is already upper triangular.

Now suppose that we have the theorem for all matrices which are $k \times k$ and we wish to show that the theorem holds for $(k+1) \times (k+1)$ matrices. Let $\mathbf{A}$ be a $(k+1) \times (k+1)$ matrix. By the fundamental theorem of algebra, the characteristic polynomial for $\mathbf{A}$ has a root in $\mathbb{C}$, say λ_1. This eigenvalue will have an eigenvector, say $\vec{v}_1$. We can extend to a basis $(\vec{v}_1, \vec{b}_1, \ldots, \vec{b}_k)$. With respect to this basis the matrix for the linear transformation has the form

$$\mathbf{B} = \begin{bmatrix} \lambda_1 & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}$$

where $\mathbf{A}_{22}$ is a $k \times k$ matrix and $\mathbf{0}$ is a column of zeros. Since matrices for the same linear transformation with respect to two different bases are similar, this gives us a matrix similar to $\mathbf{A}$.

Now $\mathbf{A}_{22}$, as a $k \times k$ matrix is similar to an upper triangular matrix, say $\mathbf{U}_{22}$. So there is a matrix $\mathbf{P}_{22}$ such that $\mathbf{P}_{22}^{-1}\mathbf{A}_{22}\mathbf{P}_{22} = \mathbf{U}_{22}$. Now let $\mathbf{P}$ be the matrix

$$\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{22} \end{bmatrix}$$

and observe that

$$\mathbf{P}^{-1} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{22}^{-1} \end{bmatrix}$$

so

$$\mathbf{P}^{-1}\mathbf{B}\mathbf{P} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{22} \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{A}_{12}\mathbf{P}_{22} \\ \mathbf{0} & \mathbf{U}_{22} \end{bmatrix}$$

which is upper triangular. Similarity is transitive, so this shows that $\mathbf{A}$ is similar to an upper triangular matrix. ∎

Example: Triangularization

Let us find an upper triangular matrix similar to

$$\mathbf{M} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

which has characteristic polynomial $p(\lambda) = \lambda^3 + \lambda^2 - \lambda - 1 = (\lambda + 1)^2(\lambda - 1)$. A little work gives us $[-1, 0, 1]$ as an eigenvector for the eigenvalue $\lambda = -1$. Extending gives the basis $([-1, 0, 1], [1, 0, 0], [0, 1, 0])$. If the linear transformation M has matrix $\mathbf{M}$ with respect to the standard basis then the matrix for the linear transformation M with respect to this new basis is

$$\begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

which is thus similar to $\mathbf{M}$.

Let us now focus on the 2×2 matrix

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

in the lower right corner. It has eigenvalues $\lambda = 1$ and $\lambda = -1$, so it is similar to the diagonal matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

obtained by using a basis of eigenvectors. One possible such basis is $([1, 1], [1, -1])$. The matrix $\mathbf{P}_{22}$ which does the change of basis is then

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

which has inverse

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

so that

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Now we know that

$$\mathbf{M} \sim \begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

and

$$\begin{bmatrix} -1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix},$$

so

$$\mathbf{M} \sim \begin{bmatrix} -1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

$\diamondsuit$

Exercises 13.2:

For problems 1–10 find the eigenvalues of the following matrices:

1. $\begin{bmatrix} 4 & 1 & -1 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{bmatrix}$

2. $\begin{bmatrix} 4 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & -5 & 3 \end{bmatrix}$

3. $\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}$

4. $\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix}$

5. $\begin{bmatrix} 1 & 2 & 0 \\ 2 & 2 & 1 \\ 0 & 1 & 4 \end{bmatrix}$

6. $\begin{bmatrix} 0 & 0 & 6 \\ 1 & 0 & 1 \\ 0 & 1 & -6 \end{bmatrix}$

7. $\begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & -3 \\ 0 & 1 & -3 \end{bmatrix}$

8. $\begin{bmatrix} 1 & 4 & 2 & 3 \\ 2 & 3 & 1 & 1 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 2 & 1 \end{bmatrix}$

9. $\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$

10. $\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix}$

11. Show that the matrix $\begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix}$ is similar to a diagonal matrix. Give the change of basis matrices as well.

12. Show that the matrix $\begin{bmatrix} 0 & 0 & -4 \\ 1 & 0 & 4 \\ 0 & 1 & 1 \end{bmatrix}$ is similar to a diagonal matrix. Give the change of basis matrices as well.

13. Show that the matrix $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -3 \\ 0 & 1 & 3 \end{bmatrix}$ is similar to a upper triangular matrix. Give the change of basis matrices as well.

14. A matrix is symmetric if $m_{ij} = m_{ji}$ for all i and j. Show that a symmetric 2×2 matrix must have real eigenvalues.

13.3 The Power Method

Using the characteristic equation to find eigenvalues can be a very cumbersome approach. In many applications (notably those of predicting the long term behavior of population models) what is needed is the principle eigenvalue, that is, the one with the largest absolute value. The power method, described in this section, provides an iterative approach to approximation of the largest eigenvalue of a reasonably large class of matrices.

To see how the method works, suppose that $\mathbf{M}$ has distinct real eigenvalues $\lambda_1, \ldots, \lambda_n$. Then there is a basis of eigenvectors $(\vec{b}_1, \ldots, \vec{b}_n)$ with each $\vec{b}_i$ an eigenvector for λ_i. If we take any vector $\vec{v}$, we can write it as

$$\vec{v} = \sum_{i=1}^{n} a_i \vec{b}_i.$$

Multiplying by $\mathbf{M}$ we get

$$\mathbf{M}\vec{v} = \sum_{i=1}^{n} a_i \lambda_i \vec{b}_i.$$

If we then divide by the eigenvalue with largest absolute value we will get a new vector with the same component for the principle eigenvector but with all of the other components reduced. Continuing this process will eventually lead to all of the components other than the component in the direction of the principle eigenvector vanishing. Thus in some sense the principle eigenvalue tells us what the long term behavior is.

As a procedure for finding the principle eigenvalue what we have just done is no help, since we needed to know what the principle eigenvalue was so that we could divide by it. So we devise a method which lets us estimate the eigenvalue while we are converging (we hope) to an eigenvector. We start with a vector $\vec{v}$ which has its largest component equal to 1 when it is represented using the standard basis. We then multiply by $\mathbf{M}$ and then divide by the entry with largest absolute value. If all goes well this will converge to an eigenvector. Eventually multiplying by $\mathbf{M}$ has the effect of multiplying

each component by the same (or approximately the same) number. The process we have described will then stabilize.

Example: Power method

Let us look at an example. The matrix

$$\mathbf{M} = \begin{bmatrix} 4 & 0 \\ 2 & 1 \end{bmatrix}$$

has eigenvalues 4 and 1. If we start with the vector $[1, 1]$ and apply the power method we first multiply by $\mathbf{M}$ to get $[4, 3]$, then divide by the largest entry (4) to get the next approximation to an eigenvector: $[1, \frac{3}{4}]$. Repeating this process leads to the following results:

largest entry	next iterate
4	[1,.6875]
4	[1,.671875]
4	[1,.66796875]
4	[1,.666992187]
4	[1,.666748046]

It is not hard to see that the eigenvalue is 4. The approximate eigenvectors are approaching the eigenvector $[1, \frac{2}{3}]$. ♢

Let us try an example where it is not so obvious what the eigenvalues are.

Example:

Let $\mathbf{M}$ be the matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -1 & 4 & 0 \end{bmatrix}$$

and start with the vector $[1, 1, 1]$. The following results are obtained:

largest entry of $\mathbf{M}\vec{v}$	next approximate eigenvector
12	[.5,1,.25]
6.5	[.5,1,.538461538]
8.23076923	[.5,1,.425233645]
7.55140187	[.5,1,.463490099]
7.7809406	[.5,1,.449817083]
7.6989025	[.5,1,.454610251]
7.72766151	[.5,1,.452918389]
after 20 iterations	
7.72015326	[.5,1,.453358876]

at which point no further changes occur to within eight decimal places. ♢

How can the algorithm fail? Two possibilities exist: if we start with a vector which completely misses the subspace of $\mathcal{V}$ spanned by the eigenvectors for the largest eigenvector the algorithm may converge to a smaller eigenvalue and eigenvector for it. It can also happen that there are two eigenvalues with the same absolute value.

Example:

An example of this is the matrix

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

which has eigenvalues 1 and -1. If we start with the vector [1,1] we will get the eigenvalue 1 and the eigenvector [1,1] immediately (lucky guess for the starting point!). If on the other hand we start with [1,0], we will get an alternation between [1,0] and [0,1]. ♢

Periodic behavior in the approximate eigenvalue and approximate eigenvectors suggests multiple principle eigenvalues. If we have complex eigenvalues we also get periodic behavior. Since complex eigenvalues for real matrices occur in conjugate pairs, we always have two complex eigenvalues with the same absolute value.

Example:

As an example consider the matrix

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

which has complex eigenvalues. If we apply the power method starting with [1,1] we get the following pattern:

largest entry of $\mathbf{M}\vec{v}$	next approximate eigenvector
2	[0,1]
-1	[1,-1]
-2	[1,0]
1	[1,1]
2	[0,1]
-1	[1,-1]
-2	[1,0]
1	[1,1]

Clearly we are caught in a loop. $\diamondsuit$

Exercises 13.3:

1. Use the power method to find the principle eigenvalue and an eigenvector for it for the following matrices:

   (a) $\begin{bmatrix} 1 & 2 \\ 2 & -3 \end{bmatrix}$

   (b) $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

   (c) $\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & -1 \\ 2 & -1 & 3 \end{bmatrix}$

2. Suppose that the n by n matrix $\mathbf{M}$ has n distinct eigenvalues $\lambda_1 \dots \lambda_n$. Show that $\mathbf{M} - \lambda_1 \mathbf{I}$ has n distinct eigenvalues $0, \lambda_2 - \lambda_1, \dots \lambda_n - \lambda_1$. This allows us to use the power method to find smaller eigenvalues too.

Chapter 14

Inner Products and Approximation

So far we have not put the ideas related to the dot product of two vectors in the plane in a more general setting. Our vectors have no notion of magnitude or angle between them yet. In some cases such notions would be artificial. (What would the angle between two polynomials mean, particularly if they do not have intersecting graphs? We will see that there is a notion of orthogonality for polynomials which is important, though it has nothing to do with the angle formed at the intersection of the graphs.) When we add the analogue of a dot product, called an inner product, we are adding additional structure. This additional structure turns out to be very useful.

14.1 Definition and Examples

What additional structure should we add to get the analogue of a dot product? Let us consider the properties of a dot product: the dot product of two vectors is a scalar; it doesn't matter what order we write the product in; scalar multiples and sums behave nicely with respect to dot product. There were other properties, too; look back at the end of Chapter 4 . The next definition captures what is needed. (In this book we only consider real valued inner products. Complex valued inner products are also important, but a bit more complicated.)

Definition 14.1.1 *An* **inner product** *on a vector space $\mathcal{V}$ over the reals is a function from $\mathcal{V} \times \mathcal{V}$ to $\mathbb{R}$ whose value at $(\vec{a}, \vec{b})$ is denoted $\langle \vec{a} \mid \vec{b} \rangle$, satisfying*

the following axioms:

Symmetry:	$\langle \vec{a} \mid \vec{b} \rangle = \langle \vec{b} \mid \vec{a} \rangle$
Positive definiteness:	$\langle \vec{a} \mid \vec{a} \rangle \geq 0$ *with equality if and only if* $\vec{a} = \vec{0}$
Linearity:	$\langle \vec{a} \mid k\vec{b} \rangle = k\langle \vec{a} \mid \vec{b} \rangle$
	$\langle \vec{a} \mid \vec{b} + \vec{c} \rangle = \langle \vec{a} \mid \vec{b} \rangle + \langle \vec{a} \mid \vec{c} \rangle$

Note that symmetry tells us that what we have said for the linearity in the second variable also holds for the first variable.

Definition 14.1.2 *If $\langle \ \mid \ \rangle$ is an inner product on $\mathcal{V}$ and $\vec{a}$ is a vector in $\mathcal{V}$, then the norm of $\vec{a}$, denoted $\|\vec{a}\|$ is $\sqrt{\langle \vec{a} \mid \vec{a} \rangle}$.*

The square root in question always exists because of positive definiteness. The norm gives us a notion of length for vectors in an inner product space. We also get a notion of orthogonality:

Definition 14.1.3 *Two vectors $\vec{v}$ and $\vec{w}$ are said to be* **orthogonal** *if $\langle \vec{v} | \vec{w} \rangle = 0$.*

We will call a vector space over the reals which is equipped with a (specified) inner product an **inner product space**. The examples will show that it is quite possible for a vector space to have several different inner products.

14.1.1 Examples in $\mathbb{R}^n$

Dot product: The dot product of vectors in $\mathbb{R}^2$ is a function from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$ satisfying the axioms for an inner product space. That is, after all, where we got the idea for the axioms.

A similar dot product may be defined on $\mathbb{R}^n$ by

$$[a_1, \ldots, a_n] \cdot [b_1, \ldots, b_n] = \sum_{i=1}^{n} a_i b_i.$$

The symmetry of the dot product will follow from n applications of commutativity of the reals. Positive definiteness follows from the fact that the dot product of a vector in $\mathbb{R}^n$ with itself is the sum of the squares of the scalar components and so is positive unless all of the components are zero. Linearity follows from the associative and distributive laws in $\mathbb{R}$.

Example:

The dot product of the vectors $[1, 2, 3, 4]$ and $[1, -2, -1, 3]$ is $1 \times 1 + 2 \times -2 + 3 \times -1 + 4 \times 3 = 1 - 4 - 3 + 12 = 6$.

The norm $\|[1, 2, 3, 4]\|$ is $\sqrt{1^2 + 2^2 + 3^2 + 4^2} = \sqrt{30}$. ♢

Weighted dot product If $\vec{w}$ is a vector in $\mathbb{R}^n$ all of whose components are positive, then $\vec{w}$ can be used as a set of weights in defining an inner product:

$$\langle \vec{a} \mid \vec{b} \rangle = \sum_{i=1}^{n} a_i b_i w_i.$$

It is clear that this inner product satisfies axioms symmetry and linearity. The reason that we need all the components of $\vec{w}$ to be positive is so that $\langle \vec{a} \mid \vec{a} \rangle$ is certain to be a sum with no negative terms and with zeros only when $\vec{a}$ has a zero component.

Example:

The weighted dot product of the vectors $[1, 2, 3, 4]$ and $[1, -2, -1, 3]$ using the weights $\vec{w} = [1, 2, 2, 1]$ is $1 \times 1 \times 1 + 2 \times 2 \times -2 + 2 \times 3 \times -1 + 1 \times 4 \times 3 = 1 - 8 - 6 + 12 = -1$.

Using this inner product

$$\|[1, 2, 3, 4]\| = \sqrt{1 \times 1 \times 1 + 2 \times 2 \times 2 + 2 \times 3 \times 3 + 1 \times 4 \times 4} = \sqrt{43}.$$

♢

14.1.2 Function space examples

We next consider a family of inner products defined using definite integrals. We have chosen to call the inner products by the names of certain families of polynomials associated with them.

Legendre inner product: Let $\mathcal{V}$ be the vector space of continuous functions from the interval $[-1, 1]$ to $\mathbb{R}$. Since continuous functions are integrable and the product of continuous functions is continuous we can define the inner product $\langle f \mid g \rangle$ by

$$\langle f \mid g \rangle = \int_{-1}^{1} f(x) g(x) dx.$$

Symmetry will hold because $f(x)g(x) = g(x)f(x)$. Linearity follows from simple properties of integrals. The only sticky point is showing that positive definiteness holds. Now $\langle f \mid f\rangle$ is the integral of f^2 on $[-1, 1]$, and f^2 is always bigger than or equal to 0. We need to show that the integral will be zero only if f is identically 0. Now f^2 is continuous, so if it is strictly positive at a point a it will be bigger than $f^2(a)/2$ on some interval $(a - \delta, a + \delta)$ for small enough δ. This is enough to guarantee that the integral of f^2 is at least $\delta f^2(a)/2$, which is strictly positive.

Example: Legendre inner product

Let $f(x) = x$ and $g(x) = x^2 - 1$, then

$$\begin{aligned} \langle f \mid g\rangle &= \int_{-1}^{1} x(x^2 - 1)\ dx = 0 \\ \|f\| &= \sqrt{\int_{-1}^{1} x^2\ dx} = \sqrt{\frac{2}{3}} \end{aligned}$$

$\diamondsuit$

The concept of magnitude which results from this kind of inner product is important in engineering and physics. If instead of the interval $[-1, 1]$ you use one period for a periodic function (say $[-\pi, \pi]$ for the sine) and then divide by the magnitude of the constant function 1, you get the root mean square magnitude of the function. This turns out to be exactly the right approach for measuring, for instance, the power in a signal.

One of the major uses of inner products on spaces of functions is the measurement of the error in an approximation: use the inner product to calculate the magnitude of the difference between what you want and what you have. Often it makes a difference where the error occurs. For instance, if you are approximating the mortality function in calculating life insurance rates you will be less concerned about errors in the approximation for people over 90 than you would be for people between 30 and 40, so you would put more weight on the interval between 30 and 40 than you would on the interval 90 to 100. For such situations we use a weighting function $w(x)$ in the integral in the same way we used a weighting vector $\vec{w}$ in the earlier example. For the result to be an inner product it suffices for $w(x)$ to be strictly positive on the interval used in the integral. One can actually allow $w(x)$ to be 0 in isolated points, though it is rare that such weights are desired.

The most important examples of weighted inner products use improper integrals.

Hermite inner product: One possibility, the Hermite inner product, is to use the bell shaped curve of a normal distribution as a weighting function:

$$\langle f \mid g \rangle = \int_{-\infty}^{\infty} f(x)g(x)e^{-x^2}dx.$$

The integration technique involved in using this inner product is a bit tricky, so we give a table for some selected powers of x:

n	$\int_{-\infty}^{\infty} x^n e^{-x^2}\ dx$
0	$\sqrt{\pi}$
1	0
2	$\frac{1}{2}\sqrt{\pi}$
3	0
4	$\frac{3}{4}\sqrt{\pi}$
5	0
6	$\frac{15}{8}\sqrt{\pi}$
7	0
8	$\frac{105}{16}\sqrt{\pi}$
9	0
10	$\frac{945}{32}\sqrt{\pi}$

Example:

Let $f(x) = x$ and $g(x) = x^2 - 1$, then

$$\begin{aligned} \langle f \mid g \rangle &= \int_{-\infty}^{\infty} x(x^2-1)e^{-x^2}\ dx = 0 \\ \|f\| &= \sqrt{\int_{-\infty}^{\infty} x^2 e^{-x^2}\ dx} = \sqrt{\frac{1}{2}\sqrt{\pi}} \end{aligned}$$

$\diamondsuit$

Laguerre inner product: Another important example is given by

$$\langle f \mid g \rangle = \int_0^\infty f(x)g(x)e^{-x}dx.$$

Here an integration by parts argument will give

$$\int_0^\infty x^n e^{-x}\ dx = n!$$

making calculations with the Laguerre inner product of polynomials not too difficult.

Example:

Let $f(x) = x$ and $g(x) = x^2 - 1$, then

$$\begin{aligned} \langle f \mid g \rangle &= \int_0^\infty x(x^2-1)e^{-x}\ dx = 3! - 1! = 5 \\ \|f\| &= \sqrt{\int_0^\infty x^2 e^{-x}\ dx} = \sqrt{2!} = \sqrt{2} \end{aligned}$$

$\diamondsuit$

Both of Laguerre and Hermite inner products occur naturally because of connections with differential equations which have been found to be important in physics.

Tchebyshev inner product: Our last example is the Tchebyshev inner product, which has important applications in numerical analysis where it is used to find the best points to use in interpolation. Here the interval used is $[-1, 1]$ and the weighting function is $w(x) = \frac{1}{\sqrt{1-x^2}}$.

Because the integration technique involved in evaluating the integrals for the Tchebyshev inner product is a bit delicate (the integrals which result are improper at both end points and require both trigonometric substitutions and even powers of $\sin(\theta)$), we include a table of values of the integral

$\int_{-1}^{1} \frac{x^n}{\sqrt{1-x^2}}\, dx$:

n	$\int \frac{x^n}{\sqrt{1-x^2}}\, dx$	$\int_{-1}^{1} \frac{x^n}{\sqrt{1-x^2}}\, dx$
0	$\arcsin(x)$	π
1	$-\sqrt{1-x^2}$	0
2	$-\frac{x\sqrt{1-x^2}}{2} + \frac{\arcsin(x)}{2}$	$\frac{\pi}{2}$
3	$\sqrt{1-x^2}\left(-\frac{2}{3} - \frac{x^2}{3}\right)$	0
4	$\sqrt{1-x^2}\left(-\frac{3x}{8} - \frac{x^4}{4}\right) + \frac{3}{8}\arcsin(x)$	$\frac{3\pi}{8}$
5	$\sqrt{1-x^2}\left(-\frac{8}{15} - \frac{4x^2}{15} - \frac{x^4}{5}\right)$	0
6	$\sqrt{1-x^2}\left(-\frac{5x}{16} - \frac{5x^3}{24} - \frac{x^5}{6}\right) + \frac{5}{16}\arcsin(x)$	$\frac{5\pi}{16}$
7	$\sqrt{1-x^2}\left(-\frac{16}{35} - \frac{8x^2}{35} - \frac{6x^4}{35} - \frac{x^6}{7}\right)$	0
8	$\sqrt{1-x^2}\left(-\frac{35x}{128} - \frac{35x^3}{192} - \frac{7x^5}{48} - \frac{x^7}{8}\right) + \frac{35}{128}\arcsin(x)$	$\frac{35\pi}{128}$

Example:

Let $f(x) = x$ and $g(x) = x^2 - 1$, then

$$\begin{aligned} \langle f \mid g \rangle &= \int_{-1}^{1} \frac{x(x^2-1)}{\sqrt{1-x^2}}\, dx = 0 \\ \|f\| &= \sqrt{\int_{-1}^{1} \frac{x^2}{\sqrt{1-x^2}}\, dx} = \sqrt{\frac{\pi}{2}} \end{aligned}$$

$\diamondsuit$

14.1.3 Cauchy-Schwarz and the Triangle Inequality

We finish this section with two important inequalities which follow from the definition of an inner product. They generalize results about dot products and distances discussed in chapter 1.

Theorem 14.1.1 (Cauchy-Schwarz Inequality) *For any inner product space*

$$\langle \vec{a} \mid \vec{b} \rangle^2 \leq \langle \vec{a} \mid \vec{a} \rangle \langle \vec{b} \mid \vec{b} \rangle.$$

PROOF:

If $\vec{b} = \vec{0}$, then the theorem is trivial, so let us consider the case where $\vec{b}$ is not $\vec{0}$, so that $\langle \vec{b} \mid \vec{b} \rangle$ is strictly positive.

The only axiom for inner products which has an inequality in it is positive definiteness, so we must find a way to use it which involves both $\vec{a}$ and $\vec{b}$. Let us apply it to a vector of the form $\vec{a} + x\vec{b}$:

$$\langle \vec{a} + x\vec{b} \mid \vec{a} + x\vec{b} \rangle \geq 0.$$

Applying the other axioms we get

$$\begin{aligned}
\langle \vec{a} + x\vec{b} \mid \vec{a} + x\vec{b} \rangle &= \langle \vec{a} + x\vec{b} \mid \vec{a} \rangle + \langle \vec{a} + x\vec{b} \mid x\vec{b} \rangle \\
&= \langle \vec{a} \mid \vec{a} \rangle + x\langle \vec{b} \mid \vec{a} \rangle + x\langle \vec{a} \mid \vec{b} \rangle + x\langle x\vec{b} \mid \vec{b} \rangle \\
&= \langle \vec{a} \mid \vec{a} \rangle + 2\langle \vec{a} \mid \vec{b} \rangle x + \langle \vec{b} \mid \vec{b} \rangle x^2.
\end{aligned}$$

If we now think of this as a quadratic expression in x and treat the inner products as coefficients, we can pull the theorem out of high school algebra. If $Ax^2 + Bx + C$ is always going to be non-negative and A is positive then there can be at most one real root. That means that the discriminant $B^2 - 4AC$, which appears under the radical in the quadratic formula, must be non-positive. Applying this in the current situation we see that

$$(2\langle \vec{a} \mid \vec{b} \rangle)^2 - 4\langle \vec{a} \mid \vec{a} \rangle \langle \vec{b} \mid \vec{b} \rangle \leq 0$$

Simple algebraic manipulation then gives the theorem. ∎

Corollary 14.1.2 (The Triangle Inequality) *In any inner product space*

$$\|\vec{a} + \vec{b}\| \leq \|\vec{a}\| + \|\vec{b}\|.$$

PROOF:

First note that we know both sides of the inequality are non-negative, so it will suffice to show that

$$\|\vec{a} + \vec{b}\|^2 \leq (\|\vec{a}\| + \|\vec{b}\|)^2.$$

This helps because

$$\|\vec{a} + \vec{b}\|^2 = \langle \vec{a} + \vec{b} \mid \vec{a} + \vec{b} \rangle,$$

which we can calculate.

$$\begin{aligned}\langle \vec{a}+\vec{b} \mid \vec{a}+\vec{b}\rangle &= \langle \vec{a} \mid \vec{a}+\vec{b}\rangle + \langle \vec{b} \mid \vec{a}+\vec{b}\rangle \\ &= \langle \vec{a} \mid \vec{a}\rangle + 2\langle \vec{a} \mid \vec{b}\rangle + \langle \vec{b} \mid \vec{b}\rangle\end{aligned}$$

Now by the Cauchy-Schwarz inequality $\langle \vec{a} \mid \vec{b}\rangle^2 \leq \|\vec{a}\|^2\|\vec{b}\|^2$, so it is certainly true that the weaker statement $\langle \vec{a} \mid \vec{b}\rangle \leq \|\vec{a}\|\|\vec{b}\|$ is true. Thus

$$\begin{aligned}\langle \vec{a}+\vec{b} \mid \vec{a}+\vec{b}\rangle &\leq \langle \vec{a} \mid \vec{a}\rangle + 2\|\vec{a}\|\|\vec{b}\| + \langle \vec{b} \mid \vec{b}\rangle \\ &\leq \|\vec{a}\|^2 + 2\|\vec{a}\|\|\vec{b}\| + \|\vec{b}\|^2 \\ &= (\|\vec{a}\| + \|\vec{b}\|)^2\end{aligned}$$

as needed. ∎

Exercises 14.1:

1. Prove in detail that the Legendre inner product is in fact an inner product on the space of continuous functions from $[-1, 1]$ to $\mathbf{R}$.

2. If $\vec{w} = [1, 2, 2, 2, 3]$ and $\vec{a} = [0, 1, 2, 3, 4]$ and $\vec{b} = [9, 3, 5, 1, -6]$ then find

   (a) $\langle \vec{a} \mid \vec{b}\rangle$ using the weighted inner product

   (b) $\|\vec{a}\|$ using the weighted inner product

   (c) $\|\vec{b} - \vec{a}\|$

3. Given that $\vec{a} = [1, 2, 4]$, $\vec{b} = [3, 7, 1]$ and $\vec{c} = [2, 0, -1]$ use dot product to find

   (a) $\langle \vec{a}|\vec{b}+\vec{c}\rangle$ and $\langle \vec{a}|\vec{b}\rangle + \langle \vec{a}|\vec{c}\rangle$

   (b) $\|\vec{a}\|\|\vec{b}\|$ and $\langle \vec{a}|\vec{b}\rangle^2$

   (c) k so that $\langle \vec{a}+k\vec{b}|\vec{c}\rangle = 0$

4. Use the fact that $\int_0^\infty x^n e^{-x}dx = n!$ to find the Laguerre inner product of $f(x) = x + 1$ and $g(x) = x^2 - 1$.

5. Using the Legendre inner product find c so that $f(x) = x + c$ and $g(x) = 1$ are orthogonal; that is, so that $\langle f \mid g \rangle = 0$.

6. Find $\|x^2\|$ using

   (a) the Legendre inner product

   (b) the Tchebyshev inner product

   (c) the Laguerre inner product

7. Find $\langle x - 2|x^2 + 1\rangle$ using

   (a) the Legendre inner product

   (b) the Tchebyshev inner product

   (c) the Laguerre inner product

8. Illustrate the triangle inequality by finding

$$\|x^2\| + \|1\| \text{ and } \|x^2 + 1\|$$

   using

   (a) the Legendre inner product

   (b) the Tchebyshev inner product

   (c) the Laguerre inner product

14.2 Orthogonal Bases

If $\mathcal{V}$ is an inner product space there is more that we can ask for in a basis. Since we have notions of orthogonality and length we can ask for the basis vectors to be pairwise orthogonal and have unit length. Part of our reason for asking for these properties is given by the next theorem.

Theorem 14.2.1 *If a set of non-zero vectors $\{\vec{a}_1, \ldots, \vec{a}_n\}$ has the property that $\langle \vec{a}_i \mid \vec{a}_j \rangle = 0$ for all $i \neq j$, then it is linearly independent.*

PROOF:

Suppose, to the contrary, that the set is dependent. Then there are scalars $k_1, \ldots k_n$ not all 0 so that

$$k_1\vec{a}_1 + \ldots + k_n\vec{a}_n = 0.$$

Suppose that k_i is non-zero. If we take an inner product on both sides with $\vec{a}_i$ we get

$$\langle k_1\vec{a}_1 \mid \vec{a}_i\rangle + \ldots + \langle k_i\vec{a}_i \mid \vec{a}_i\rangle + \ldots \langle k_n\vec{a}_n \mid \vec{a}_i\rangle = \langle \vec{0} \mid \vec{a}_i\rangle.$$

Now all of the terms $\langle k_j\vec{a}_j \mid \vec{a}_i\rangle = k_j\langle \vec{a}_j \mid \vec{a}_i\rangle = 0$ if $i \neq j$ and $\langle \vec{0} \mid \vec{v}\rangle = 0$ for any $\vec{v}$. This tells us that $\langle k_i\vec{a}_i \mid \vec{a}_i\rangle = 0$. But this tells us that $k_i\langle \vec{a}_i \mid \vec{a}_i\rangle = 0$ so either $k_i = 0$ or $\vec{a}_i = \vec{0}$, both of which we have assumed to be false. Thus the set is linearly independent. ∎

One consequence of this theorem is that when we are looking for a basis in an inner product space we can save ourselves the labor of checking for linear independence by making sure that vectors in our set are pairwise orthogonal. The next proposition tells how to do that.

Proposition 14.2.2 *Given a vector $\vec{v}$ and a vector $\vec{w}$ not parallel to $\vec{v}$,*

$$\vec{w} - \frac{\langle \vec{v} \mid \vec{w}\rangle}{\langle \vec{v} \mid \vec{v}\rangle}\vec{v} \text{ is orthogonal to } \vec{v}.$$

Proof:

Taking an inner product with $\vec{v}$ gives

$$\langle \vec{v} \mid \vec{w}\rangle - \frac{\langle \vec{v} \mid \vec{w}\rangle}{\langle \vec{v} \mid \vec{v}\rangle}\langle \vec{v} \mid \vec{v}\rangle = \langle \vec{v} \mid \vec{w}\rangle - \langle \vec{v} \mid \vec{w}\rangle = 0.$$

∎

If the vectors were parallel then

$$\vec{w} = \frac{\langle \vec{v} \mid \vec{w}\rangle}{\langle \vec{v} \mid \vec{v}\rangle}\vec{v}.$$

Now suppose that we have a set of pairwise orthogonal vectors, or an orthogonal set of vectors for short, $A = \{\vec{v}_1, \ldots \vec{v}_n\}$, and a vector $\vec{w}$ we

want to write as the sum of a vector in Span(A) and a vector orthogonal to everything in A. We know that

$$\vec{w} - \left(\frac{\langle \vec{v}_1 \mid \vec{w}\rangle}{\langle \vec{v}_1 \mid \vec{v}_1\rangle}\right)\vec{v}_1$$

is orthogonal to $\vec{v}_1$. So is $\vec{v}_2$. Thus

$$\left(\vec{w} - \left(\frac{\langle \vec{v}_1 \mid \vec{w}\rangle}{\langle \vec{v}_1 \mid \vec{v}_1\rangle}\right)\vec{v}_1\right) - \left(\left(\frac{\langle \vec{v}_2 \mid \vec{w}\rangle}{\langle \vec{v}_2 \mid \vec{v}_2\rangle}\right)\vec{v}_2\right)$$

will be orthogonal to $\vec{v}_1$ because

$$\left(\vec{w} - \left(\frac{\langle \vec{v}_1 \mid \vec{w}\rangle}{\langle \vec{v}_1 \mid \vec{v}_1\rangle}\right)\vec{v}_1\right) \text{ is orthogonal to } \vec{v}_1,$$

as is $\vec{v}_2$, and it will be orthogonal to $\vec{v}_2$ by Proposition 6.3.2. Continuing in this manner we see that

$$\vec{w} - \sum_{i=1}^{n} \frac{\langle \vec{v}_i \mid \vec{w}\rangle}{\langle \vec{v}_i \mid \vec{v}_i\rangle}\vec{v}_i$$

is orthogonal to all of the $\vec{v}_i$. This is the essence of the Gram Schmidt orthogonalization process.

Theorem 14.2.3 (Gram Schmidt Orthogonalization process) *If $(\vec{a}_1, \ldots, \vec{a}_n)$ is an ordered set of vectors and we define*

$$\begin{aligned} \vec{b}_1 &= \vec{a}_1 \\ \vec{b}_j &= \vec{a}_j - \sum_{i=1}^{j-1} \frac{\langle \vec{a}_j \mid \vec{b}_i\rangle}{\langle \vec{b}_i \mid \vec{b}_i\rangle}\vec{b}_i, \end{aligned}$$

Then the resulting set $\{\vec{b}_1, \ldots, \vec{b}_n\}$ is an orthogonal set which spans the same subspace as the set $\{\vec{a}_1, \ldots, \vec{a}_n\}$. If we omit from the set of $\vec{b}$'s those which are $\vec{0}$ we get an orthogonal basis for the subspace spanned by $(\vec{a}_1, \ldots, \vec{a}_n)$.

Proof:

We will prove this by induction on n. If $n = 1$ there is nothing to prove since any one element set is an orthogonal set. So suppose that we have the theorem for $n = k$, we need to show

that the set $\{\vec{b}_1, \ldots, \vec{b}_{k+1}\}$ is orthogonal. Since we are assuming that the set $\{\vec{b}_1, \ldots, \vec{b}_k\}$ is orthogonal, we need only show that $\vec{b}_{k+1} \in \{\vec{b}_1, \ldots, \vec{b}_k\}^\perp$. Let j be a number between 1 and k. We know that $\vec{b}_j$ is orthogonal to all the other $\vec{b}_i$ with $i < k$ so the sum

$$\sum_{\substack{i=1 \\ i \neq j}}^{k} \frac{\langle \vec{a}_{k+1} \mid \vec{b}_i \rangle}{\langle \vec{b}_i \mid \vec{b}_i \rangle} \vec{b}_i$$

is too. Thus it will suffice to prove that

$$\vec{a}_{k+1} - \frac{\langle \vec{a}_{k+1} \mid \vec{b}_j \rangle}{\langle \vec{b}_j \mid \vec{b}_j \rangle} \vec{b}_j$$

is orthogonal to $\vec{b}_j$. But this is exactly what Proposition 14.2.2 tells us.

To see that the subspace spanned by the $\vec{a}$'s is the same as the subspace spanned by the $\vec{b}$'s we need only note that each $\vec{b}_k$ is defined as a linear combination of the vectors $\vec{a}_1 \ldots \vec{a}_k$ and that it is easy to see how to write $\vec{a}_k$ as a linear combination of the vectors $\vec{b}_1$ to $\vec{b}_k$:

$$\vec{a}_k = \vec{b}_k + \sum_{i=1}^{k-1} \frac{\langle \vec{a}_j \mid \vec{b}_i \rangle}{\langle \vec{b}_i \mid \vec{b}_i \rangle} \vec{b}_i.$$

For the comment that the $\vec{b}$'s form a basis for the subspace they span if we omit zeros, recall that an orthogonal set of nonzero vectors is linearly independent. A linearly independent spanning set is a basis. ∎

The calculation carried out in the Gram Schmidt process is somewhat easier if at each point we replace $\vec{b}_i$ by the unit vector in the same direction. This process, called normalization, forces $\langle \vec{b}_i \mid \vec{b}_i \rangle$ to be 1, saving a lot of division. The result is what is called an orthonormal set of vectors and the process is called Gram Schmidt orthonormalization.

Definition 14.2.1 *A set of vectors* $\{\vec{v}_1, \ldots, \vec{v}_n\}$ *is called* **orthonormal** *if*

$$\langle \vec{v}_i \mid \vec{v}_j \rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} .$$

Example:

The standard basis $\{[1,0,0],[0,1,0],[0,0,1]\}$ is an orthonormal basis for $\mathbb{R}^3$. $\diamondsuit$

Example:

Use the Gram Schmidt process to find an orthogonal basis for $\mathbb{R}^3$ starting with the set $\{[1,1,1],[2,1,1],[2,2,1]\}$. We get

$$\begin{aligned}
\vec{b}_1 &= [1,1,1] \\
\vec{b}_2 &= [2,1,1] - \frac{[2,1,1]\cdot[1,1,1]}{[1,1,1]\cdot[1,1,1]} \\
&= [2,1,1] - \frac{4}{3}[1,1,1] \\
&= [2/3,-1/3,-1/3] \\
\vec{b}_3 &= [2,2,1] - \frac{[2,2,1]\cdot[1,1,1]}{[1,1,1]\cdot[1,1,1]}[1,1,1] \\
&\quad - \frac{[2,2,1]\cdot[2/3,-1/3,-1/3]}{(4/9+1/9+1/9)}[2/3,-1/3,-1/3] \\
&= [2,2,1-5/3][1,1,1] - 1/2[2/3,-1/3,-1/3] \\
&= [0,1/2,-1/2]
\end{aligned}$$

$\diamondsuit$

Example:

Use the Gram Schmidt process on the set $\{1,x,x^2\}$ with the inner product

$$\langle f \mid g\rangle = \int_0^1 f(x)g(x)dx.$$

This gives:

$$\begin{aligned}
\vec{b}_1 &= 1 \\
\vec{b}_2 &= x - \frac{\langle x \mid 1\rangle}{\langle 1 \mid 1\rangle}1
\end{aligned}$$

$$\begin{aligned}
&= x - \frac{\int_0^1 x dx}{\int_0^1 1 dx} \\
&= x - 1/2 \\
\vec{b}_3 &= x^2 - \frac{\langle x^2 \mid 1\rangle}{\langle 1 \mid 1\rangle}1 - \frac{\langle x^2 \mid x - 1/2\rangle}{\langle x - 1/2 \mid x - 1/2\rangle}(x - 1/2) \\
&= x^2 - \frac{\int_0^1 x^2 dx}{\int_0^1 1 dx} - \frac{\int_0^1 x^2(x - 1/2)dx}{\int_0^1 (x - 1/2)^2 dx}(x - 1/2) \\
&= x^2 - 1/3 - (x - 1/2) \\
&= x^2 - x + 1/6
\end{aligned}$$

$\diamondsuit$

This is an example of using the Gram Schmidt process to find orthogonal polynomials. There are many ways of generating certain of the families of orthogonal polynomials, some using the Gram Schmidt process, some using differential equations, some using a three term recurrence which can be derived from Gram Schmidt. Families of orthogonal polynomials have important uses in approximation theory, least squares fit of polynomials to data, and differential equations.

Exercises 14.2:

1. Find an orthonormal basis for $\mathbb{R}^2$ with dot product using the set $\{[1, 1], [0, 1]\}$ as a starting point.

2. Find an orthogonal basis for $\mathbb{R}^3$ with dot product using the set
$$\{[1, 0, -1], [0, 1, 1], [1, 1, 1]\}$$
as a starting point.

3. Find an orthonormal basis for $\mathbb{R}^2$ with weighted dot product with $\vec{w} = [1, 2]$ using the set $\{[1, 1], [0, 1]\}$ as a starting point.

4. Find an orthogonal basis for $\mathbb{R}^3$ with dot product with $\vec{w} = [1, 3, 2]$ using the set
$$\{[1, 0, -1], [0, 1, 1], [1, 1, 1]\}$$
as a starting point.

5. Extend the set $\{[1,0,1],[2,1,-2]\}$ to an orthogonal basis for $\mathbf{R}^3$ with dot product.

6. Use the Legendre inner product

$$\langle f \mid g \rangle = \int_{-1}^{1} f(x)g(x)dx$$

to find the first three Legendre polynomials by applying the Gram Schmidt orthogonalization process to the set $\{1, x, x^2\}$.

7. Use the Laguerre inner product

$$\langle f \mid g \rangle = \int_{0}^{\infty} f(x)g(x)e^{-x}dx$$

to find the first three Laguerre polynomials by applying the Gram Schmidt process to $\{1, x, x^2\}$. You may use the fact that

$$\int_{0}^{\infty} x^n e^{-x}dx = n!.$$

8. Use the Hermite inner product

$$\langle f \mid g \rangle = \int_{-\infty}^{\infty} f(x)g(x)e^{-x^2}\ dx$$

to find the first three Hermite polynomials by applying the Gram Schmidt orthogonalization process to the set $\{1, x, x^2\}$.

9. Use the Tchebyshev inner product

$$\langle f \mid g \rangle = \int_{-1}^{1} \frac{f(x)g(x)}{\sqrt{1-x^2}}\ dx$$

to find the first three Tchebyshev polynomials by applying the Gram Schmidt orthogonalization process to the set $\{1, x, x^2\}$.

10. Show that the set $\{\sin(nx) \mid n \text{ a positive integer}\} \cup \{\cos(mx) \mid m \text{ a positive integer}\}$ is orthogonal using the inner product

$$\langle f \mid g \rangle = \int_{-\pi}^{\pi} f(x)g(x)dx.$$

This involves showing that if $n \neq m$, then

$$\langle \sin nx \mid \cos mx \rangle = 0,$$
$$\langle \sin nx \mid \sin mx \rangle = 0$$

and that f $n \neq m$, then

$$\langle \cos nx \mid \cos mx \rangle = 0.$$

The following identities may prove useful:

$$\begin{aligned} \sin nx \cos mx &= \frac{1}{2}(\sin(n+m)x + \sin(n-m)x) \\ \sin nx \sin mx &= \frac{1}{2}(\cos(n-m)x - \cos(n+m)x) \\ \cos nx \cos mx &= \frac{1}{2}(\cos(n-m)x + \cos(n+m)x) \end{aligned}$$

This orthogonal set forms the basis for Fourier series representations of periodic functions as sums of sines and cosines.

11. (Project Problem) The Gram-Schmit process can be quite cumbersome for finding orthogonal polynomials of high degree. If the inner product involved has the property that

$$\langle xf(x)|g(x)\rangle = \langle f(x)|xg(x)\rangle$$

then a recurrence with fewer terms works with less effort. What you do is let

$$\begin{aligned} u_0(x) &= 1 \\ u_1(x) &= x - \langle u_0|x\rangle u_0(x) \\ u_n(x) &= (x - \frac{\langle xu_{n-1}|u_{n-1}\rangle}{\langle u_{n-1}|u_{n-1}\rangle})u_{n-1}(x) - \frac{\langle xu_{n-1}|u_{n-2}\rangle}{\langle u_{n-2}|u_{n-2}\rangle}u_{n-2}(x) \end{aligned}$$

(a) Show that the leading coefficient of these polynomials is always 1.

(b) Show that $\langle u_n|u_{n-1}\rangle = 0$ and $\langle u_n|u_{n-2}\rangle = 0$.

(c) Explain why any polynomial $p(x)$ of degree $n-1$ or less can be written as a linear combination of the functions u_i where $i \leq n-1$.

(d) Show that for $k < n-2$ we get $\langle xu_{n-1}|u_k\rangle = 0$.

(e) Show that for $k < n-2$ we get $\langle u_n|u_k\rangle = 0$.

(f) Use this recurrence to find the Laguerre polynomials with leading coefficient 1 up to degree 5.

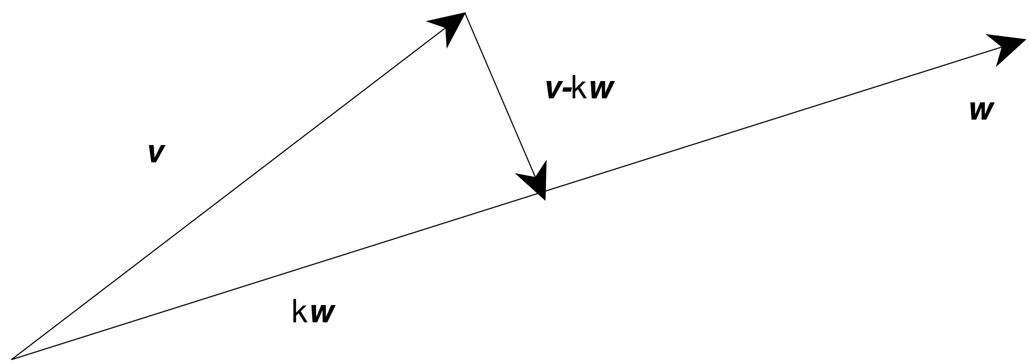


Figure 14.1: Orthogonal Projection

14.3 Projections and Best Approximations

After going to all the work that the Gram Schmidt process involves one may well ask why we care so much about orthogonal bases. For an answer let us recall one of the uses of the dot product in $\mathbb{R}^2$. Given two vectors $\vec{v}$ and $\vec{w}$ we ask for the vector in the direction of $\vec{w}$ which is closest to $\vec{v}$ in the sense that $\|\vec{v} - k\vec{w}\|$ is minimized. If we look at the situation geometrically it is clear that what we want to do is make $\vec{v} - k\vec{w}$ perpendicular to $\vec{w}$. (See Figure 14.1)

This is clear since the shortest distance from the endpoint of $\vec{v}$ to the line determined by $\vec{w}$ is along the perpendicular. Since we want $\langle \vec{w} \mid \vec{v} - k\vec{w} \rangle = 0$, we want $\langle \vec{w} \mid \vec{v} \rangle - k\langle \vec{w} \mid \vec{w} \rangle = 0$, so

$$k = \frac{\langle \vec{w} \mid \vec{v} \rangle}{\langle \vec{w} \mid \vec{w} \rangle}.$$

If we want to do a similar construction in $\mathbb{R}^3$, finding the vector in a given plane which is closest to a given vector, we drop a perpendicular from the endpoint of the vector to the plane. On the face of it, this does not appear to be as easy to do analytically as the two dimensional case was. Our approach is to find two vectors in the given plane which are themselves perpendicular and find the projection of the given vector onto each of them, then add the projections. If $\vec{w}_1$ and $\vec{w}_2$ are the perpendicular vectors in the given plane and $\vec{v}$ is the given vector then the projection is

$$\frac{\vec{v} \cdot \vec{w}_1}{\vec{w}_1 \cdot \vec{w}_1}\vec{w}_1 + \frac{\vec{v} \cdot \vec{w}_2}{\vec{w}_2 \cdot \vec{w}_2}\vec{w}_2.$$

The geometry of two space and three space leads us to suspect that if we want the vector in a subspace closest to a given vector we should take a projection along a perpendicular. In other words to find the vector in a subspace $\mathcal{W}$ closest to $\vec{v}$ we should write $\vec{v}$ as $\vec{w} + \vec{w'}$ where $\vec{w} \in \mathcal{W}$ and $\vec{w'}$ is orthogonal to everything in $\mathcal{W}$ and then use $\vec{w}$ as our best approximation. We will show that this is, in fact, the case by showing how to use an orthonormal basis to calculate distance and find projections.

Definition 14.3.1 *The projection of $\vec{v}$ onto the subspace $\mathcal{W}$ is the vector $\vec{w} \in \mathcal{W}$ such that $\vec{v} = \vec{w} + \vec{w}'$ with $\vec{w}'$ orthogonal to every vector in $\mathcal{W}$.*

Lemma 14.3.1 *If $\{\vec{b}_1, \ldots, \vec{b}_n\}$ is an orthonormal basis for $\mathcal{V}$ and $\vec{v} = \sum_{i=1}^n k_i \vec{b}_i$ then $\|\vec{v}\|^2 = \sum_{i=1}^n k_i^2$.*

PROOF:

We need only calculate $\langle \vec{v} \mid \vec{v} \rangle$. Using the properties of the inner product we find that

$$\begin{aligned} \langle \vec{v} \mid \vec{v} \rangle &= \langle \sum_{i=1}^n k_i \vec{b}_i \mid \sum_{j=1}^n k_j \vec{b}_j \rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n k_i k_j \langle \vec{b}_i \mid \vec{b}_j \rangle \\ &= \sum_{i=1}^n k_i^2 \end{aligned}$$

where the last equality follows from the fact that $\langle \vec{b}_i \mid \vec{b}_j \rangle = 0$ if $i \neq j$ and $\langle \vec{b}_i \mid \vec{b}_i \rangle = 1$. ∎

This means that if we represent our vectors in terms of an orthonormal basis, then the length looks just like the lengths we calculate using the Pythagorean theorem in Euclidean space.

Lemma 14.3.2 *If $\{\vec{b}_1, \ldots, \vec{b}_n\}$ is an orthonormal set and $\vec{v}$ is any vector in $\mathcal{V}$, then the vector*

$$\pi(\vec{v}) = \sum_{i=1}^n \langle \vec{v} \mid \vec{b}_i \rangle \vec{b}_i$$

has the property that $\vec{v} - \pi(\vec{v})$ is orthogonal to all of the $\vec{b}_i$. The vector $\pi(\vec{v})$ is the projection of $\vec{v}$ onto the subspace spanned by $\{\vec{b}_1, \ldots, \vec{b}_n\}$.

PROOF:

Again all that is involved is the calculation of a rather messy inner product:

$$
\begin{aligned}
\langle\left(\vec{v}-\sum_{i=1}^{n}\langle\vec{v}\mid\vec{b}_i\rangle\vec{b}_i\right)\mid\vec{b}_j\rangle &= \langle\vec{v}\mid\vec{b}_j\rangle-\sum_{i=1}^{n}\langle\vec{v}\mid\vec{b}_i\rangle\langle\vec{b}_i\mid\vec{b}_j\rangle \\
&= \langle\vec{v}\mid\vec{b}_j\rangle-\langle\vec{v}\mid\vec{b}_j\rangle\langle\vec{b}_j\mid\vec{b}_j\rangle \\
&\quad \text{since } \langle\vec{b}_i\mid\vec{b}_j\rangle = 0 \text{ if } i\neq j. \\
&= \langle\vec{v}\mid\vec{b}_j\rangle-\langle\vec{v}\mid\vec{b}_j\rangle \\
&= 0.
\end{aligned}
$$

∎

This lemma tells us how to find the projection of a vector onto a subspace, provided that we have an orthonormal basis for the subspace. The Gram Schmidt process tells us how to find an orthogonal basis for the subspace. Dividing each of those basis vectors by its length gives an orthonormal basis, so this information is available to us. The next theorem tells us why we want projections.

Theorem 14.3.3 *If $\mathcal{W}$ is a subspace of an inner product space $\mathcal{V}$ and $\vec{v}$ is an element of $\mathcal{V}$, then the element of $\mathcal{W}$ closest to $\vec{v}$ is the projection of $\vec{v}$ onto $\mathcal{W}$.*

PROOF:

Let $\{\vec{b}_1,\ldots,\vec{b}_{n-1}\}$ be an orthonormal basis for $\mathcal{W}$. We know that we can find one because of the Gram Schmidt process. Let $\vec{b}_n$ be the unit vector in the direction of $\vec{v}-\pi(\vec{v})$. Then $\{\vec{b}_1,\ldots,\vec{b}_n\}$ is an orthonormal set which is a basis for Span$(\mathcal{W}\cup\{\vec{v}\})$. If we write $\vec{v}$ as $\sum_{i=1}^{n}k_i\vec{b}_i$, then

$$\|\vec{v}-\pi(\vec{v})\| = \|k_n\vec{b}_n\| = \mid k_n|.$$

Any time we can write $\vec{v}$ as $\vec{w}+\vec{w}'$ with $\vec{w}$ in $\mathcal{W}$, we know that $\vec{w}'$ is in Span$(\{\vec{b}_1,\ldots,\vec{b}_n\})$. Since $\vec{w}\in\mathcal{W}$ we know that its $\vec{b}_n$ component must be 0. Thus $\vec{w}'$ must be of the form

$$\sum_{i=1}^{n-1}h_i\vec{b}_i+k_n\vec{b}_n.$$

Hence its length, calculated using Lemma 14.3.1 is

$$\sqrt{\left(\sum_{i=1}^{n-1} h_i^2\right) + k_n^2},$$

and thus is at least as big as $|\, k_n|$. This shows that the projection is the closest element of $\mathcal{W}$ to $\vec{v}$. ∎

Example:

Find the vector in the space spanned by the orthonormal set

$$\{[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}], [\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}]\}$$

closest to $[0, 1, 0]$. We want the projection of $[0, 1, 0]$ onto the subspace spanned by the given orthonormal set. It is

$$\begin{aligned}
[0,1,0] \quad - \quad & [0,1,0] \cdot [\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}][\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}] \\
- \quad & [0,1,0] \cdot [\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}][\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}] \\
= \quad & [0,1,0] - [1/3, 1/3, 1/3] \\
= \quad & [-1/3, 2/3, -1/3].
\end{aligned}$$

♢

Example:

Find the polynomial of degree 2 or less closest to x^3 using the inner product

$$\langle f \mid g \rangle = \int_0^1 f(x)g(x)dx.$$

We have the orthogonal polynomials $\{1, x - \frac{1}{2}, x^2 - x + 1/6\}$ as an example in the last section. To make this an orthonormal set we need to divide each by its length. $\langle 1 \mid 1 \rangle$ is 1, so no change is needed on the first.

$$\langle x - 1/2 \mid x - 1/2 \rangle = 1/12$$

so for our second basis vector we use $(\sqrt{12})(x-1/2)$. Continuing with Gram Schmidt

$$\begin{aligned}\langle (x^2-x+1/6) \mid (x^2-x+1/6)\rangle &= \int_0^1 (\frac{1}{36}-\frac{x}{3}+\frac{4x^2}{3}-2x^3+x^4)dx \\ &= \frac{1}{180},\end{aligned}$$

so our third basis vector is $\sqrt{180}(x^2-x+1/6)$. The projection is then

$$\begin{aligned}\int_0^1 x^3dx &+ \left(\int_0^1 x^3\sqrt{12}(x-1/2)dx\right)\sqrt{12}(x-1/2) \\ &+ \left(\int_0^1 x^3\sqrt{180}(x^2-x+1/6)dx\right)\sqrt{180}(x^2-x+1/6) \\ &= 1/4+12(\frac{1}{5}-\frac{1}{8})(x-1/2)+180(\frac{1}{6}-\frac{1}{5}+\frac{1}{24})(x^2-x+1/6) \\ &= \frac{1}{4}+\frac{9}{10}(x-\frac{1}{2})+\frac{3}{2}(x^2-x+\frac{1}{6}) \\ &= \frac{1}{20}-\frac{3x}{5}+\frac{3x^2}{2}\end{aligned}$$

The graphs are given in Figure 14.3 with the approximation graphed in black and x^3 graphed in gray.

$\diamondsuit$

Exercises 14.3:

1. Find the vector in the space spanned by $\{[1/\sqrt{2},0,-1/\sqrt{2}],[1/\sqrt{3},1/\sqrt{3},1/\sqrt{3}]\}$ closest to $[1,0,0]$.

2. Find the vector in the space spanned by $\{[1/\sqrt{5},0,0,2/\sqrt{5}],[1/2,1/2,1/2,-1/2]\}$ closest to $[1,1,1,1]$.

3. Find the vector in the space spanned by $\{[1,0,1],[2,2,1]\}$ closest to $[1,1,1]$.

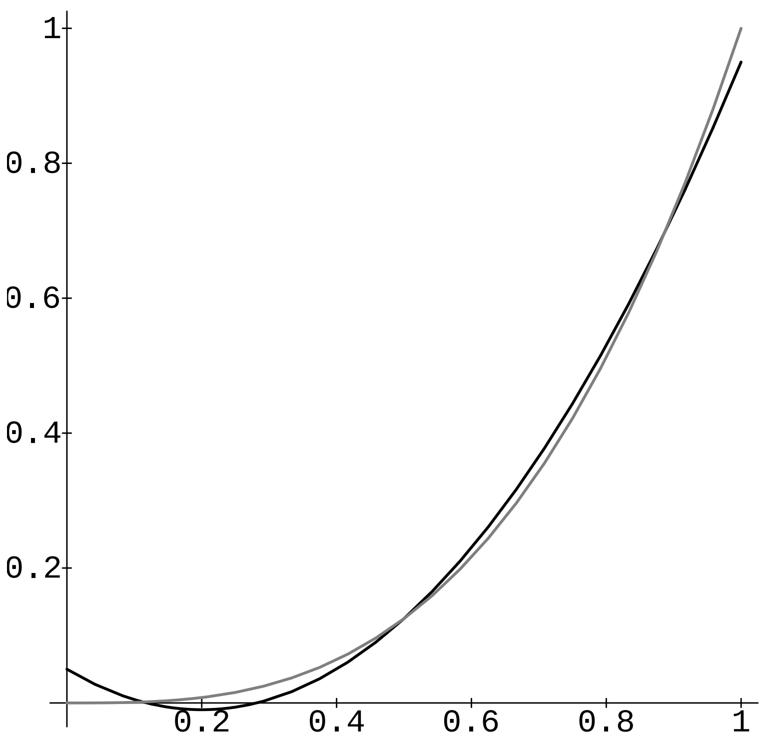


Figure 14.2: Approximating x^3

4. Find the polynomial of degree 2 closest to $\ln(x+1)$ using the inner product
$$\langle f|g\rangle = \int_0^1 f(x)g(x)dx.$$
Hint: Use integration by parts.

5. Find the polynomial of degree 2 closest to $\sin(x)$ using the Legendre inner product.

6. Find the polynomial of degree 2 closest to $\sin(x)$ using the Lageurre inner product. (This will involve some nasty integration by parts.)

7. (Project Problem) The problem of finding the *best* curve of a particular type for a set of data points occurs often in science. For instance suppose that we want to find the polynomial of degree 2 which comes closest to fitting the points (1,0), (2,1), (3,5), (4,0), (5,1) in the sense that
$$\sum_{i=1}^{5}(p(i)-y_i)^2$$
is minimized. Compare the amount of work involved in the following two approaches.

   (a) Since a polynomial of degree 4 is completely determined by its value at 5 points, the rule
   $$\langle p|q\rangle = \sum_{i=1}^{5} p(i)q(i)$$
   defines an inner product on the vector space $\mathbb{R}[x]_4$. Use the Gram Schmidt orthonormalization procedure to find an orthonormal basis for $\mathbb{R}[x]_2 < \mathbb{R}[x]_4$. Then find the projection of the polynomial of degree 4 passing through the given points onto $\mathbb{R}[x]_2$.

   (b) The least squares approximation can also be found by applying multivariable calculus. We want
   $$E(a,b,c) = \sum_{i=1}^{5}(a\ i^2 + b\ i + c - y_i)^2$$

to be minimized. This will happen when the partial derivatives with respect to a, b, and c are all zero. This gives the system of so called normal equations:

$$\begin{aligned} a\sum_{i=1}^{5} i^4 + b\sum_{i=1}^{5} i^3 + c\sum_{i=1}^{5} i^2 &= \sum_{i=1}^{5} i^2 y_i \\ a\sum_{i=1}^{5} i^3 + b\sum_{i=1}^{5} i^2 + c\sum_{i=1}^{5} i &= \sum_{i=1}^{5} i y_i \\ a\sum_{i=1}^{5} i^2 + b\sum_{i=1}^{5} i + c5 &= \sum_{i=1}^{5} y_i. \end{aligned}$$

This system can then be solved to find the coefficients of the approximating polynomial. Try it and see that you get the same result.

(c) Which of the methods in a and b would you prefer to use if you had 100 sets of data of the form $(1, y_1), (2, y_2), \ldots (5, y_5)$. Which would let you fit a polynomial of degree 3 without throwing out all of your calculations?

14.4 Approximation in Function Spaces

14.4.1 Fourier series approximations

We noted in the exercises that the set of functions

$$\{1, \sin(x), \sin(2x), \ldots, \sin(nx), \cos(x), \cos(2x), \ldots, \cos(nx)\}$$

is orthogonal with respect to the inner product

$$\langle f|g\rangle = \int_{-\pi}^{\pi} f(x)g(x)dx.$$

The norm which results from this inner product gives the root mean square measure of the size of periodic functions. This family of functions is often used to give approximations to periodic functions in terms of a fundamental frequency and its harmonics. This approximation of periodic functions by

trigonometric polynomials is the first step in analysis of signals using Fourier series.

If we limit our attention to even functions (those for which $f(-x) = f(x)$) we can use just the cosine terms, since all the integrals with $\sin(nx)$ will give 0. The n^{th} Fourier coefficient will then be

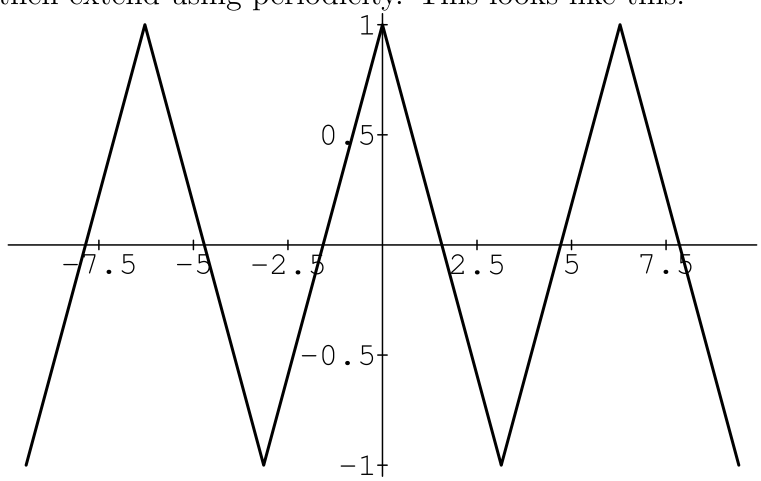
$$\frac{\langle f|\cos(nx)\rangle}{\langle\cos(nx)|\cos(nx)\rangle} = \int_{-\pi}^{\pi} f(x)\cos(nx)dx/\pi.$$

Example: Sawtooth wave

Suppose we start by looking at a sawtooth wave:

$$w(x) = \begin{cases} 1 - \frac{2}{\pi}x & \text{if } 0 \leq x \leq \pi \\ 1 + \frac{2}{\pi}x & \text{if } -\pi \leq x < 0 \end{cases}$$

and then extend using periodicity. This looks like this:



A sawtooth wave

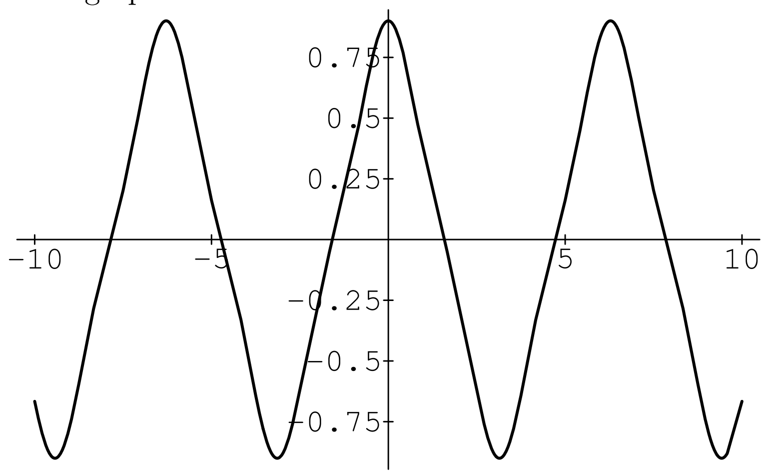
The Fourier coefficient is

$$\frac{\int_{-\pi}^{\pi} w(x)\cos(nx)dx}{\pi} = \frac{8}{n^2\pi^2} \text{ for } n \geq 1.$$

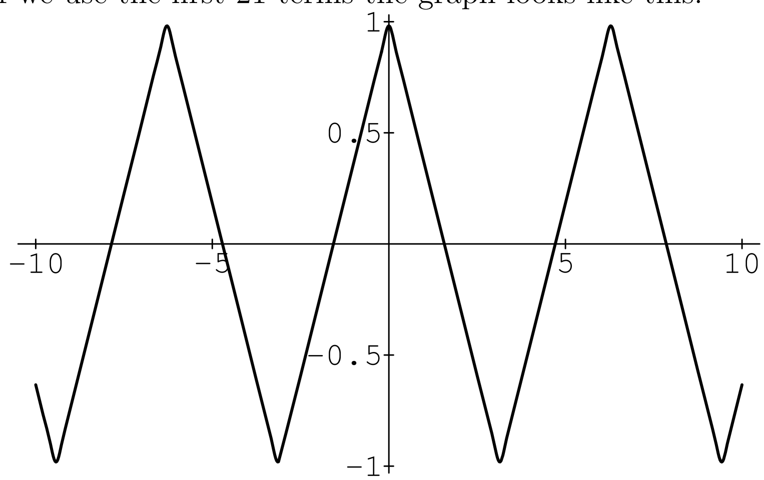
This makes it fairly easy to approximate $w(x)$ using trigonometric polynomials. If we use the first three terms, we get

$$w(x) \approx \frac{4\cos(x)}{\pi} - \frac{4\cos(3x)}{3\pi},$$

which has graph:



If we use the first 21 terms the graph looks like this:



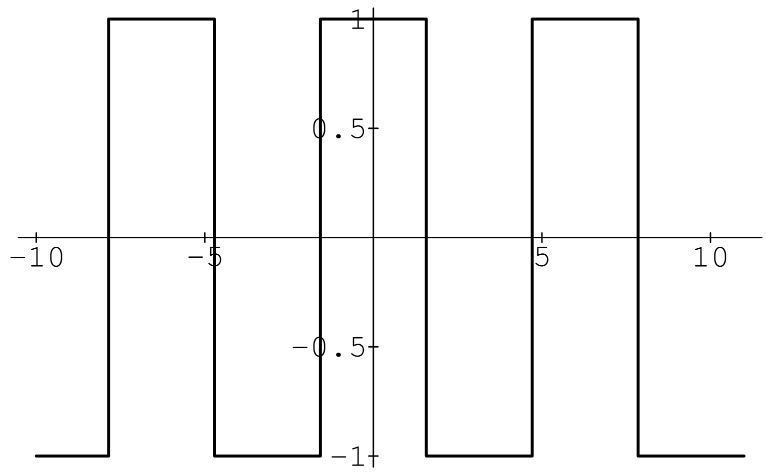
which gives a very good approximation! $\diamondsuit$

Example: Square wave

If we look at a square wave we can see one of the things which can go wrong with Fourier approximations. The square wave is periodic with period 2π and has its value on $[-\pi, \pi]$ given by

$$s(x) = \begin{cases} 1 & \text{if } -\pi/2 \leq x \leq \pi/2 \\ -1 & \text{otherwise} \end{cases}$$

This has graph (the vertical lines are artifacts of the program used to produce the graph)



A square wave.

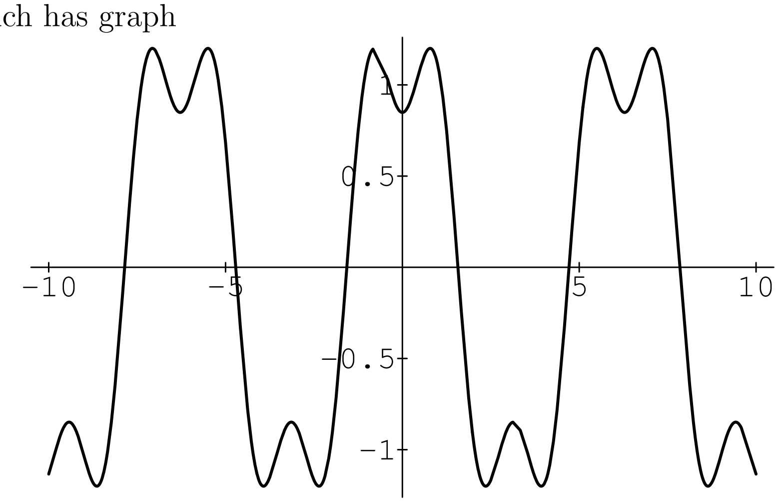
The Fourier coefficient is

$$\frac{\int_{-\pi}^{\pi} s(x)\cos(nx)dx}{\pi} = \begin{cases} 0 & \text{if } n \text{ is even} \\ \frac{4}{n\pi} & \text{if } n = 4k+1 \\ -\frac{4}{n\pi} & \text{if } n = 4k-1 \end{cases}$$

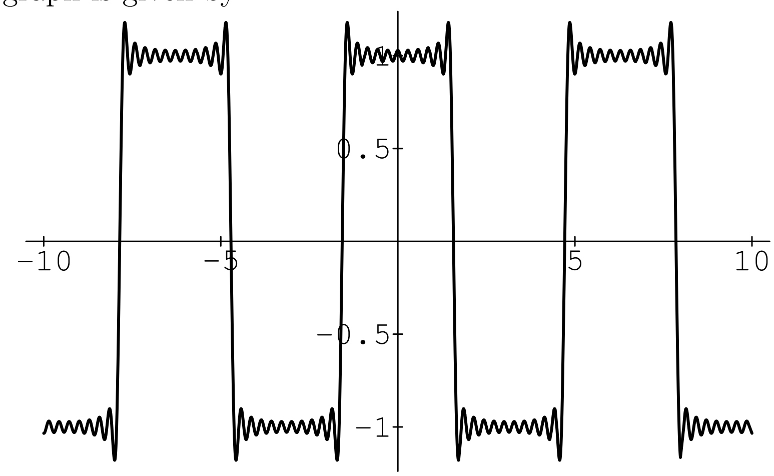
Again this makes it fairly easy to approximate $s(x)$ using trigonometric polynomials. If we use the terms up to $n = 3$ we get

$$s(x) \approx \frac{4\cos(x)}{\pi} - \frac{4\cos(3x)}{3\pi},$$

which has graph



This isn't as good an approximation as we got for the sawtooth function, so lets look at the approximation up to $n = 21$. Here the graph is given by



The "ears" at each of the jump discontinuities will not disappear if we take larger n. They illustrate the phenomenon of Gibbs overshoot; it is typical of Fourier approximations to functions with jump discontinuities. ♢

14.4.2 Wavelet approximations

Fourier series give good approximations to continuous periodic functions. They are less successful with discontinuous functions and functions with bounded support. We turn our attention next to wavelet functions, which approximate functions with bounded support much more satisfactorily.

For the purposes of this section we will use the inner product

$$\langle f|g\rangle = \int_{-\infty}^{+\infty} f(x)g(x)\ dx$$

on the vector space of square integrable functions. We will start with two simple functions and use them to generate a large orthogonal set. Let

$$\phi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Then we define translations and dilations of these functions by

$$\phi_n^m(x) = \sqrt{2^n}\phi(2^n x + m) \text{ and } \psi_n^m(x) = \sqrt{2^n}\psi(2^n x + m).$$

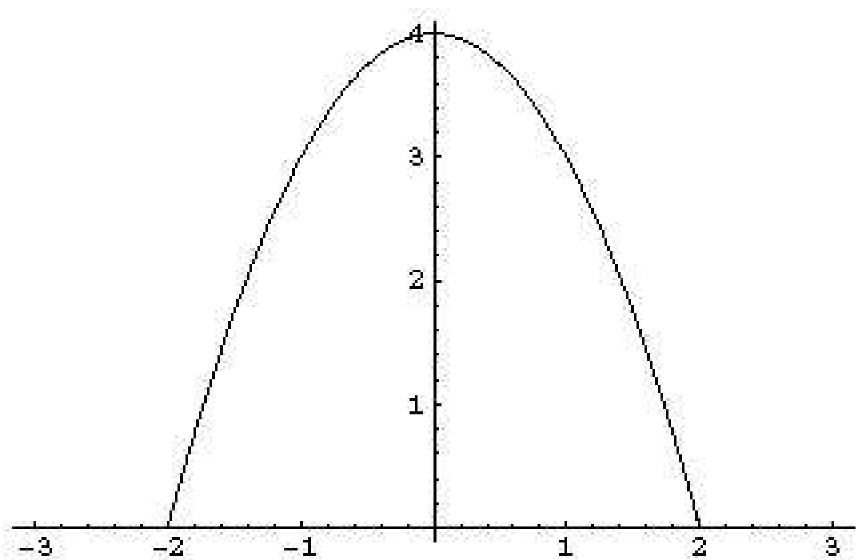
The functions ϕ_n^m are called **Haar scaling functions** and the ψ_n^m are called **Haar wavelets** . For each n the family of all ϕ_n^m and all ψ_n^m is orthonormal. Each of the ψ_{n-1}^m can be obtained as linear combinations of the ϕ_n^m. The index n gives a measure of the grain size of the approximation.

Example:

To illustrate how wavelets can be used to approximate a function with a bounded support, let us consider approximations to

$$g(x) = \begin{cases} 4 - x^2 & \text{if } -2 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

which has graph

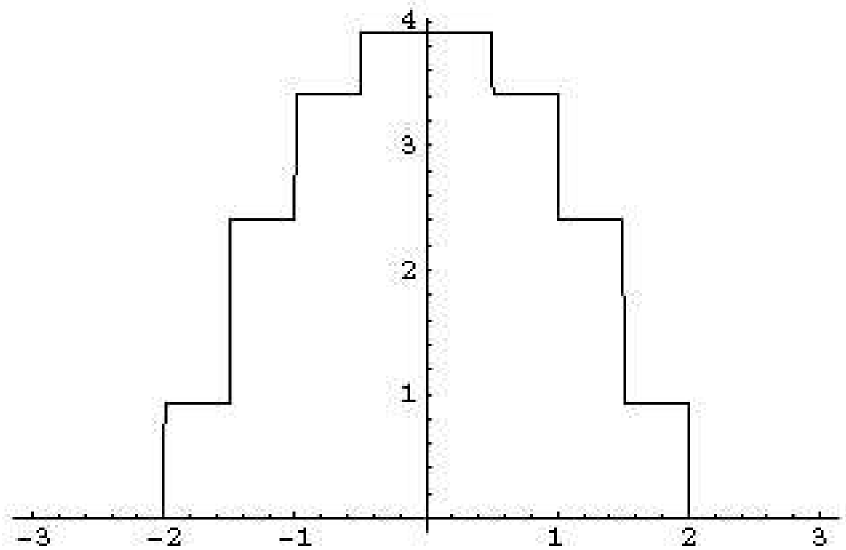


Since this function is 0 outside of [-2,2], it will not be necessary to consider wavelets which have support outside that interval. Thus the projection of g onto the subspace generated by the wavelets of grain size n will only involve a finite number of terms. In particular,

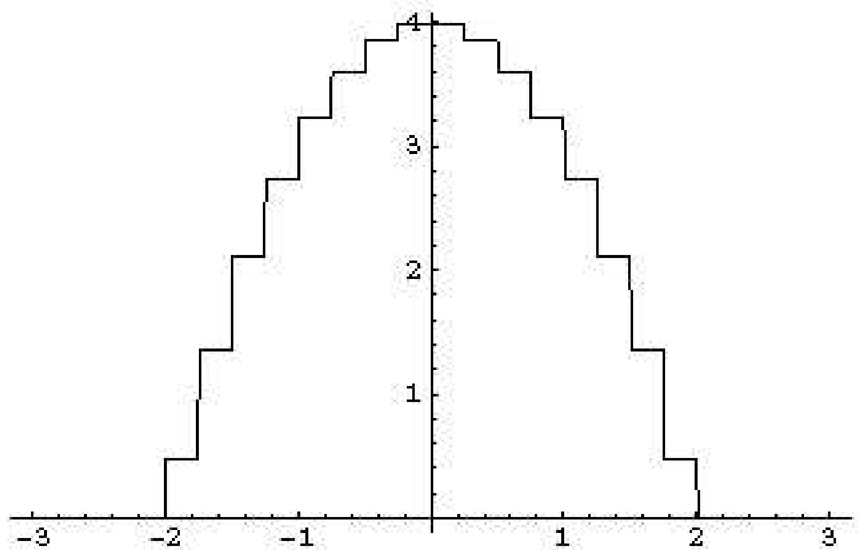
$$f(x) \approx \sum_{m=-b2^n+1}^{b2^n} \left(\langle f|\psi_n^m\rangle \psi_n^m(x) + \langle f|\phi_n^m\rangle \phi_n^m(x)\right)$$

if the support of f is contained in $[-b, b]$. This is $2^{n+1}b$ terms, so we see that one of the problems with Haar wavelet approximations is that a large number of terms are needed for a reasonable approximation. Some of the work can be eliminated by observing that the contribution of the $\phi_n^m(x)$ is precisely the whole approximation at the $n-1$ scale, so that to improve the approximation one need only add the contribution of the $\psi_n^m(x)$ terms.

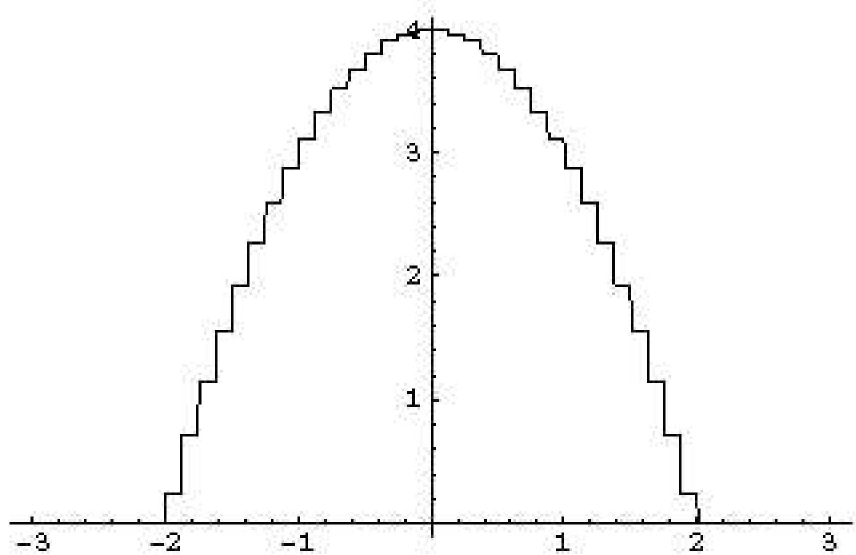
We can see how the granularity affects the quality of the approximation by looking at the approximations for several different values of n.



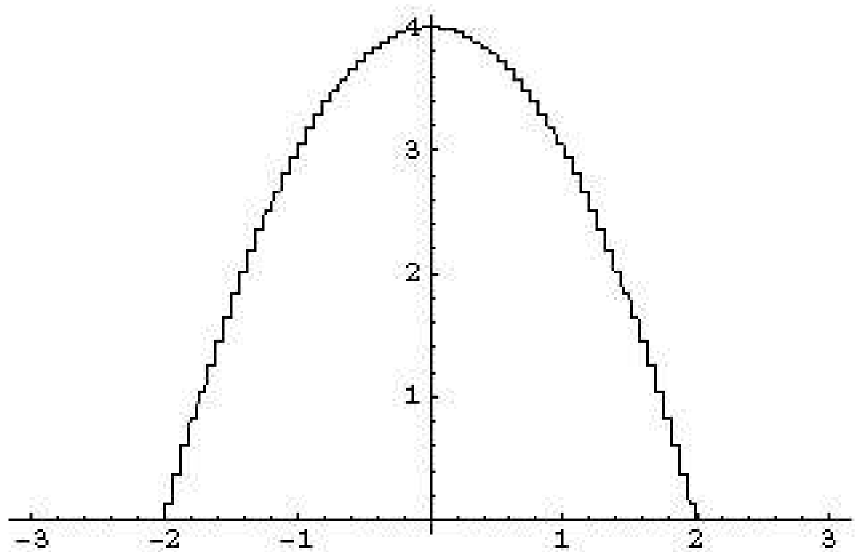
Wavelet approximation to g using $n = 0$



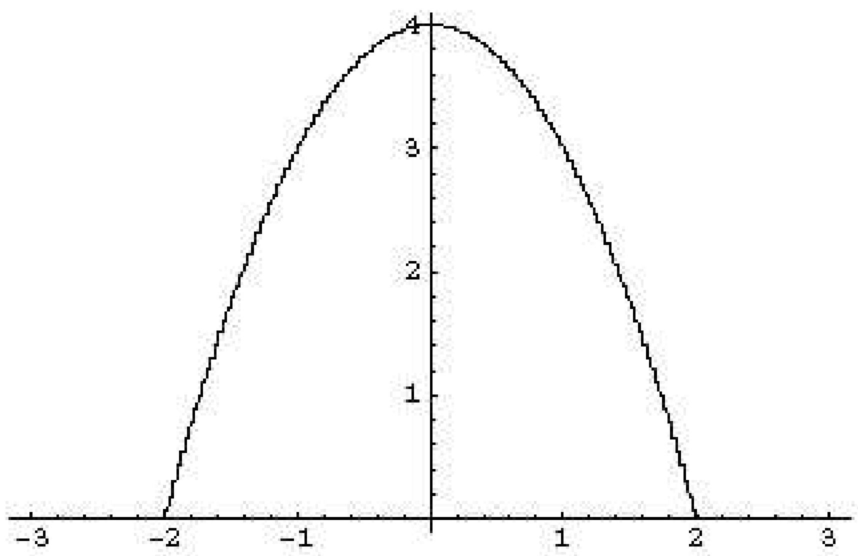
Wavelet approximation to g using $n = 1$



Wavelet approximation to g using $n = 2$



Wavelet approximation to g using $n = 3$



Wavelet approximation to g using $n = 4$

This last approximation uses 64 terms. $\diamondsuit$

Exercises 14.4:
Working with Fourier series and wavelet approximations really only makes sense if you have some computer software to do the calculations for you. If you do, try the following:

1. See what the wavelet approximation does to a function which looks like the square wave, but is 0 outside the interval from -5 to 5.

2. See how the Fourier series works on the example given for wavelets. Note that some restrictions will need to be made, since the Fourier series always give periodic functions.

3. Define the function $h(x)$ on the interval $[-\pi, \pi)$ using

$$h(x) = \begin{cases} x + \pi & \text{if } x \leq 0 \\ x - \pi & \text{otherwise} \end{cases}$$

   Use a wavelet approximation to see how well Haar wavelets handle jump discontinuities.

4. Extend the function h in the previous exercise to make it periodic with period 2π. Explore how well the Fourier series approximates this function. Note that h is not even, so the sin terms must be used as well as the cos terms.

Chapter 15

Orthogonal Subspaces and Transformations

15.1 Orthogonal Subspaces

The existence of an inner product on a vector space opens up the possibility of many new questions. In this section we will use inner products to produce a new construction of subspaces. The construction is based on the observation that inner product with a given vector gives a linear transformation:

Proposition 15.1.1 *If $\mathcal{V}$ is an inner product space and $\vec{v}$ is any element of $\mathcal{V}$ then $\langle \ \mid \vec{v}\rangle : \mathcal{V} \to \mathbb{R}$ is a linear transformation.*

PROOF:

> First observe that $\langle \ \mid v\rangle$ is a function from $\mathcal{V}$ to $\mathbb{R}$since it takes a vector $\vec{w}$ to the real number $\langle \vec{w} \mid \vec{v}\rangle$. Linearity follows quickly from the properties of inner products:
>
> $$\begin{aligned} \langle \vec{w} + \vec{u} \mid \vec{v}\rangle &= \langle \vec{w} \mid \vec{v}\rangle + \langle \vec{u} \mid \vec{v}\rangle \\ \langle k\vec{w} \mid \vec{v}\rangle &= k\langle \vec{w} \mid \vec{v}\rangle \end{aligned}$$
>
> by the linearity axiom in our definition of an inner product. ∎

One of the ways that we can use this proposition is by describing the kernel of the linear transformation $\langle \ \mid v\rangle$. Recall that the kernel of a linear transformation is a subspace and that it is the set of all vectors sent to $\vec{0}$ by

the linear transformation. In this case we get the set of all vectors in $\mathcal{V}$ which are orthogonal to $\vec{v}$. This construction can be generalized to the subspace of vectors orthogonal to a whole set of vectors:

Definition 15.1.1 *If S is a non-empty set of vectors in an inner product space $\mathcal{V}$, then the orthogonal complement $S^\perp$ is the set of all vectors orthogonal to all of the members of S.*

Proposition 15.1.2 *For any $S \subseteq \mathcal{V}$, $S^\perp$ is a subspace of $\mathcal{V}$.*

PROOF:

$S^\perp$ is the intersection of the subspaces $Ker(\langle - \mid \vec{s}\rangle)$ where $\vec{s}$ is in S. Since $S^\perp$ is the intersection of subspaces, it is itself a subspace. ∎

Example: Finding $A^\perp$

Let $\mathcal{V} = \mathbb{R}^3$, $A = \{a_1 = [1,-1,0], a_2 = [0,2,3]\}$. If $\vec{b} = [b_1, b_2, b_3] \in A^\perp$, then we must have

$$\langle \vec{b} \mid \vec{a}_1\rangle = 0 \text{ and } \langle \vec{b} \mid \vec{a}_2\rangle = 0,$$

since any vector in $A^\perp$ must be orthogonal to every vector in Span(A). This tells us that

$$\begin{aligned} b_1 - b_2 &= 0 \\ 2b_2 + 3b_3 &= 0. \end{aligned}$$

The solution of this system is easily seen to be

$$\begin{aligned} b_1 &= b_2 \\ b_2 &= b_2 \\ b_3 &= -\frac{2}{3}b_2, \end{aligned}$$

so

$$A^\perp = \{b_2[1,1,-\frac{2}{3}]\} = \text{Span}([1,1,-\frac{2}{3}]).$$

◇

Example: Finding $A^\perp$

Let $\mathcal{V} = \mathbb{R}^3$. $A = \{\vec{a}\}$ where $\vec{a} = [1, 2, 3]$. If $\vec{b} = [b_1, b_2, b_3] \in A^\perp$, then $\langle \vec{b} \mid \vec{a} \rangle = 0$, or

$$b_1 + 2b_2 + 3b_3 = 0.$$

One way of writing the solution to this system is

$$\begin{aligned} b_1 &= -2b_2 - 3b_3 \\ b_2 &= b_2 \\ b_3 &= b_3. \end{aligned}$$

Thus,

$$\begin{aligned} A^\perp &= \{[-2b_2 - 3b_3, b_2, b_3] \mid b_2, b_3 \in \mathbb{R}\} \\ &= \{b_2[-2, 1, 0] + b_3[-3, 0, 1] \mid b_2, b_3 \in \mathbb{R}\} \\ &= S(\{[-2, 1, 0], [-3, 0, 1]\}). \end{aligned}$$

$\diamondsuit$

It will often be the case that we are interested in the orthogonal complement of $\mathcal{U}$ when $\mathcal{U}$ is a subspace of $\mathcal{V}$. When $\mathcal{V}$ is finite dimensional, it is also true that

$$\mathcal{V} = \mathcal{U} + \mathcal{U}^\perp,$$

and

$$\mathcal{U} \cap \mathcal{U}^\perp = \{\vec{0}\}.$$

Note that it is easy to show that $A \cap A^\perp = \{\vec{0}\}$ for any set A since the only vector which is orthogonal to itself is the zero vector, but the proof of the other half must await careful consideration of bases. Observe, though, that the above assertion is illustrated by the example $\mathcal{V} = \mathbb{R}^3, \mathcal{U} = xy$-plane, $\mathcal{U} =$ z-axis:

$$\begin{aligned} \mathbb{R}^3 &= (xy\text{-plane }) + (z\text{-axis}) \\ (xy\text{-plane} \cap (z\text{-axis}) &= \{\vec{0}\}. \end{aligned}$$

Example:

For another example let us consider $\mathcal{V} = \mathbb{R}^4$, let

$$\mathcal{U} = \mathrm{Span}(\{\vec{e}_1, \vec{e}_3\})$$

where $\vec{e}_1 = [1,0,0,0]$ and $\vec{e}_3 = [0,0,1,0]$. If $\vec{b} = [b_1, b_2, b_3, b_4] \in \mathcal{U}$, then $\langle \vec{b} \mid \vec{e}_1 \rangle = \langle \vec{b} \mid \vec{e}_3 \rangle = 0$, or $b_1 = b_3 = 0$. Thus

$$\begin{aligned} \mathcal{U}^{\perp} &= \{[0, b_2, 0, b_4] \mid b_2, b_4 \in \mathbb{R}\} \\ &= \{b_2[0,1,0,0] + b_4[0,0,0,1]\} \\ &= \{b_2\vec{e}_2 + b_4\vec{e}_4\} \\ &= S(\{\vec{e}_2, \vec{e}_4\}). \end{aligned}$$

Note that

$$\mathcal{U} + \mathcal{U}^{\perp} = \mathbb{R}^4$$

and

$$\mathcal{U} \cap \mathcal{U}^{\perp} = \{\vec{0}\}.$$

$\diamondsuit$

With the Gram Schmidt process we can prove that $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$ by exhibiting a basis for $\mathcal{V}$ which is the disjoint union of a basis for $\mathcal{U}$ and a basis for $\mathcal{U}^{\perp}$.

Theorem 15.1.3 *If $\mathcal{V}$ is a finite dimensional inner product space and $\mathcal{U}$ is any subspace, then $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}^{\perp}$.*

PROOF:

Choose an ordered basis for $\mathcal{U}$, say $(\vec{a}_1, \ldots, \vec{a}_m)$, and extend to a basis $(\vec{a}_1, \ldots, \vec{a}_n)$ for all of $\mathcal{V}$. Apply the Gram Schmidt process to obtain an orthogonal basis for $\mathcal{V}$, $(\vec{b}_1, \ldots, \vec{b}_n)$.

In this process the subspace spanned by the first m elements of the original basis is the same as the subspace spanned by the first m elements of the orthogonal basis, so $(\vec{b}_1, \ldots, \vec{b}_m)$ is a basis for $\mathcal{U}$. We claim that $(\vec{b}_{m+1}, \ldots, \vec{b}_n)$ is a basis for $\mathcal{U}^{\perp}$. Clearly a vector orthogonal to all of the vectors in $\mathcal{U}$ must have the coefficients of $\vec{b}_1$ to $\vec{b}_m$ all equal to 0, so it must lie in the subspace of $\mathcal{V}$ spanned by $\{\vec{b}_{m+1}, \ldots, \vec{b}_n\}$. Similarly anything in that subspace will be orthogonal to everything in $\mathcal{U}$. The set $\{\vec{b}_{m+1}, \ldots, \vec{b}_n\}$ is

linearly independent because it is a subset of a basis for $\mathcal{V}$, namely $(\vec{b}_1, \ldots, \vec{b}_n)$. It spans $\mathcal{U}^\perp$. Thus it is a basis for $\mathcal{U}^\perp$.

It is an easy exercise to show that partitioning a basis gives a direct sum decomposition. ∎

Exercises 15.1:

1. Let $\mathcal{V} = \mathbb{R}^3$. Find $A^\perp$ using ordianry dot product if
   (a) $A = \{[1, 0, 0]\}$
   (b) $A = \{[1, 2, 0]\}$
   (c) $A = \{[1, 1, 0], [0, 1, 1]\}$
   (d) $A = \{[0, 1, 1], [1, 0, 1]\}$
   (e) $A = \text{Span}(\{[1, 1, 0], [0, 1, 1]\})$
2. Let $\mathcal{V} = \mathbb{R}^4$. Find $A^\perp$ using dot product if
   (a) $A = \{[1, 1, 0, 0], [0, 0, 1, 1]\}$
   (b) $A = \{[1, 1, 0, 0], [1, 0, 1, 0]\}$
   (c) $A = \{[1, 1, 0, 0], [1, 0, 1, 0], [0, 0, 1, 1]\}$
   (d) $A = \{[1, 2, 3, 4], [0, -1, 1, 0]\}$
3. Prove that if A is not empty and $\mathcal{V}$ is an inner product space then $A \subseteq (A^\perp)^\perp$.
4. Give an example which shows that it need not be true that $A = (A^\perp)^\perp$. When does equality hold?.
5. Prove that if $\mathcal{V}$ is an inner product space and A is a subset of B then $B^\perp$ is a subset of $A^\perp$.
6. Prove that if $\mathcal{V}$ is an inner product space and A is a subset of $\mathcal{V}$ then $A^\perp = (S(A))^\perp$.

15.2 Orthogonal Transformations

If $\mathcal{V}$ and $\mathcal{W}$ are inner product spaces it is natural to ask that a map $L : \mathcal{V} \to \mathcal{W}$ preserve both the vector space structure and the inner product. In this section we will explore the properties of the matrix associated with such a map. We start with a formal definition.

Definition 15.2.1 *If $\mathcal{V}$ and $\mathcal{W}$ are inner product spaces and $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation then L is an orthogonal transformation if and only if*

$$\langle L\vec{a} \mid L\vec{b}\rangle = \langle a \mid b\rangle$$

for all $\vec{a}$, $\vec{b}$ in $\mathcal{V}$.

Example: Inclusion of a subspace

The map $L : \mathbb{R}^2 \to \mathbb{R}^3$ with $L([x, y]) = [x, y, 0]$ preserves the usual inner product (dot product) and is linear, hence it is an orthogonal transformation. ♢

Example: Rotation

The map

$$p : \mathbb{R}^2 \to \mathbb{R}^2$$

taking

$$[x, y] \text{ to } [(x + y)/\sqrt{2}, (x - y)/\sqrt{2}]$$

is orthogonal. To see this we must calculate $\langle p[a, b] \mid p[c, d]\rangle$.

$$\begin{aligned}
\langle p[a, b] \mid p[c, d]\rangle &= \langle[(a + b)/\sqrt{2}, (a - b)/\sqrt{2}] \mid [(c + d)/\sqrt{2}, (c - d)/\sqrt{2}]\rangle \\
&= \frac{(a + b)(c + d)}{2} + \frac{(a - b)(c - d)}{2} \\
&= \frac{ac + ad + bc + bd}{2} + \frac{ac - ad - bc + bd}{2} \\
&= ac + bd \\
&= \langle[a, b] \mid [c, d]\rangle.
\end{aligned}$$

♢

Example: A non-orthogonal transformation

The map $s : \mathbb{R}^2 \to \mathbb{R}^2$ which takes $[x, y]$ to $[x, 2y]$ is not orthogonal. To see this note that

$$\langle [1,1] \mid [1,1] \rangle = 2$$

but

$$\langle s([1,1]) \mid s([1,1]) \rangle = \langle [1,2] \mid [1,2] \rangle = 5.$$

$\diamondsuit$

This last example illustrates the following proposition:

Proposition 15.2.1 *Orthogonal transformations preserve length.*

PROOF:

If $\vec{a}$ is a vector in an inner product space $\mathcal{V}$ then

$$\|\vec{a}\,\| = \sqrt{\langle \vec{a} \mid \vec{a} \rangle}.$$

We want to show $\|L\vec{a}\| = \|\vec{a}\|$ when L is orthogonal. Now

$$\begin{aligned} \|L\vec{a}\| &= \sqrt{\langle L\vec{a} | L\vec{a} \rangle} \\ &= \sqrt{\langle \vec{a} | \vec{a} \rangle} \\ &= \|a\| \end{aligned}$$

using the definition of length for the outside equalities and the definition of orthogonal transformations for the inside equality. ∎

Proposition 15.2.2 *Orthogonal transformations take orthonormal sets of vectors to orthonormal sets of vectors.*

PROOF:

We need to show that if $\langle \vec{a} \mid \vec{b} \rangle = 0$ then $\langle L\vec{a} \mid L\vec{b} \rangle = 0$ and that if $\langle \vec{a} \mid \vec{a} \rangle = 1$ then $\langle L\vec{a} \mid L\vec{a} \rangle = 1$. Both are direct consequences of the definition of orthogonal transformation. ∎

The most natural kind of basis for us to use in exploring the properties of the matrix associated with an orthogonal transformation is an orthonormal basis $\{\vec{b}_i\}_{i=1,\ldots,n}$. We know from the previous proposition that the vectors $L(\vec{b}_1),\ldots,L(\vec{b}_n)$ will be orthonormal. The next lemma will help us identify orthogonal transformations by looking at the columns in the corresponding matrix.

Lemma 15.2.3 *If $\{\vec{d}_i\}_{i=1,n}$ is an orthonormal basis for the inner product space $\mathcal{W}$ then the inner product*

$$\langle\sum_{i=1}^{n} a_i\vec{d}_i \mid \sum_{j=1}^{n} b_j\vec{d}j\rangle = [a_1,\ldots,a_n]\cdot[b_1,\ldots,b_n].$$

Proof:

This is a straightforward, if tedious, calculation:

$$\langle\sum_{i=1}^{n} a_i\vec{d}_i \mid \sum_{j=1}^{n} b_j\vec{d}_j\rangle = \sum_{i=1}^{n} a_i\langle\vec{d}_i \mid \sum_{j=1}^{n} b_j\vec{d}_j\rangle$$

by linearity of the inner product in the first variable. This is turn equals

$$\sum_{i=1}^{n} a_i \sum_{j=1}^{n} b_j\langle\vec{d}_i \mid \vec{d}_j\rangle$$

by linearity in the second variable. Combining the double sum we get

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_ib_j\langle\vec{d}_i \mid \vec{d}_j\rangle.$$

Now $\langle\vec{d}_i \mid \vec{d}_j\rangle = 0$ if $i \neq j$ and $\langle\vec{d}_i \mid \vec{d}_i\rangle = 1$ so this double sum collapses to a single sum

$$\sum_{i=1}^{n} a_ib_i.$$

This is just $[a_1,\ldots,a_n]\cdot[b_1,\ldots,b_n]$. ∎

Theorem 15.2.4 *If $L : \mathcal{V} \to \mathcal{W}$ is an orthogonal transformation and $(\vec{b}_i)_{i=1,m}$ and $(\vec{d}_j)_{j=1,n}$ are ordered orthonormal bases for $\mathcal{V}$ and $\mathcal{W}$ then the matrix for L with respect to $(\vec{b}_i)$ and $(\vec{d}_j)$ has orthonormal columns.*

PROOF:

> Since L preserves the inner product it takes the basis to an orthonormal set. The column vectors of the associated matrix are obtained by representing the $L\vec{b}_i$ in terms of the basis for $\mathcal{W}$. Then Lemma 15.2.3 tells us how to calculate $\langle L\vec{b}_i \mid L\vec{b}_j\rangle$ using these coefficients. The result is that the column vectors must be orthonormal. ∎

The converse of this theorem is also true.

Theorem 15.2.5 *If $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation with associated matrix $\mathbf{L}$ using the ordered orthonormal bases $(\vec{b}_i)$ and $(\vec{d}_j)$, then if $\mathbf{L}$ has orthonormal columns then L is an orthogonal transformation.*

PROOF:

> Since the columns of $\mathbf{L}$ are the images of basis vectors, the fact that $\mathbf{L}$ has orthonormal columns tells us that the the basis $(\vec{b}_i)$ is taken to an orthonormal set by L. To show that $\langle L\vec{x} \mid L\vec{y}\rangle = \langle \vec{x} \mid \vec{y}\rangle$ we first write $\vec{x}$ and $\vec{y}$ as linear combinations of basis elements:
>
> $$\vec{x} = \sum_{i=1}^{m} x_i\vec{b}_i \text{ and } \vec{y} = \sum_{j=1}^{m} y_j\vec{b}_j.$$
>
> Then
>
> $$L\vec{x} = \sum_{i=1}^{m} x_iL\vec{b}_j \text{ and } L\vec{y} = \sum_{j=1}^{m} y_jL\vec{b}_j$$
>
> using linearity of L. Then we calculate
>
> $$\begin{aligned}\langle L\vec{x} \mid L\vec{y}\rangle &= \langle\sum_{i=1}^{m} x_iL\vec{b}_i \mid \sum_{j=1}^{m} y_jL\vec{b}_j\rangle \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m} x_iy_j\langle L\vec{b}_i \mid L\vec{b}_j\rangle\end{aligned}$$

using linearity of the inner product. Since the $L\vec{b_i}$ form an orthonormal set we replace this with the single sum

$$\sum_{i=1}^{m} x_i y_i.$$

Our earlier Lemma tells us that this is $\langle \vec{x} \mid \vec{y} \rangle$ as needed. ∎

This theorem gives us a useful tool for identifying orthogonal transformations. To see whether a given linear transformation is orthogonal one need only represent it with respect to orthonormal bases and see if the resulting matrix has orthonormal columns.

This leads to the following definition.

Definition 15.2.2 *A square matrix* $\mathbf{A}$ *is called orthogonal if its column vectors form an orthonormal set.*

By reason of tradition we restrict the term orthogonal matrix to square matrices. For such matrices it can also be shown that the rows form an orthonormal set. The transpose of a matrix $\mathbf{M} = [[m_{ij}]]$ is the matrix $\mathbf{M}^t = [[m_{ji}]]$ obtained by reflecting through the main diagonal. The fact that an orthogonal matrix has orthonormal columns then tells us that if $\mathbf{M}$ is orthogonal $\mathbf{M}^t\mathbf{M}$ is an identity matrix. For square matrices this will tell us that $\mathbf{M}^{-1} = \mathbf{M}^t$.

Example: A rotation

The matrix for the rotation by 45 degrees with respect to the standard basis $([1,0],[0,1])$ is

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix},$$

which is an orthogonal matrix. ♢

Matrices with orthonormal columns also figure in the QR factorization given by the Gram-Schmit orthonormalization process:

Theorem 15.2.6 (QR Factorization) *If* $\mathbf{M}$ *is a matrix with linearly indpendent columns, then there are matrices* $\mathbf{Q}$*, which has orthonormal columns, and* $\mathbf{R}$*, which is upper triangular, with* $\mathbf{M} = \mathbf{QR}$

PROOF:

The columns of $\mathbf{M}$ form a linearly independent set, so the Gram-Schmidt orthonormalization process can be used to find column vectors $\vec{q}_k$ such that

1. Each $\vec{q}_k$ is a linear combination of the first k columns of $\mathbf{M}$
2. $\text{Span}\{\vec{q}_i | i = 1 \dots k\} = \text{Span}\{\text{the first k columns of } \mathbf{M}\}$
3. The vectors $\vec{q}_k$ form an orthomormal set.
4. The matrix $\mathbf{Q}$ has the vectors $\vec{q}_k$ as its column vectors.

We can write the k^{th} column of $\mathbf{M}$ as a linear combination of $\vec{q}_1 \dots \vec{q}_k$, so there are numbers $r_{i,j}$ with

$$m_{i,k} = \sum_{j=1}^{k} q_{i,j} r_{j,k}.$$

This gives us an upper triangular matrix $\mathbf{R} = [[r_{i,j}]]$ with $\mathbf{M} = \mathbf{QR}$.

The easiest way to find $\mathbf{R}$ once you have $\mathbf{Q}$ is by noting that $\mathbf{R} = \mathbf{IR} = \mathbf{Q}^t\mathbf{QR} = \mathbf{Q}^t\mathbf{M}$. ∎

Exercises 15.2:

1. Are the following orthogonal matrices?

   (a) $\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$

   (b) $\begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & -1 \\ -1 & 2 & 0 \end{bmatrix}$

   (c) $\begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{-12}{13} & \frac{5}{13} \end{bmatrix}$

2. Prove that the composition of two orthogonal transformations is orthogonal.

3. In each of the following the inner product is given by dot product. Which are orthogonal transformations?

   (a) $f : \mathbb{R}^3 \to \mathbb{R}^3$ with $f[x, y, z] = [-x, -y, z]$
   (b) $g : \mathbb{R}^3 \to \mathbb{R}^3$ with $g[x, y, z] = [\frac{x+y+z}{\sqrt{3}}, \frac{x+z}{\sqrt{2}}, \frac{x+y}{\sqrt{2}}]$
   (c) $h : \mathbb{R}^2 \to \mathbb{R}^2$ with $h[x, y] = [x \cos(\Theta) + y \sin(\Theta), x \sin(\Theta) - y \cos(\Theta)]$
   (d) $i : \mathbb{R}^2 \to \mathbb{R}^3$ with $i[x, y] = [x/\sqrt{2}, x/\sqrt{2}, y]$

4. Give four examples of orthogonal matrices.

15.3 Rational Canonical Forms for Orthogonal Matrices

Throughout this section we assume that $\mathcal{V}$ is a finite dimensional vector space equipped with an inner product. Our object is to show that orthogonal linear transformations from $\mathcal{V}$ to itself have a rational canonical form. That is, buy proper choice of basis, the matrix will be a block diangonal matrix with each block a companion matrix.

Theorem 15.3.1 *If $L : \mathcal{V} \to \mathcal{V}$ preserves inner products, then if $\mathcal{U}$ is L-invariant, then so is $\mathcal{U}^\perp$.*

PROOF:

First note that since L preserves inner products, it must be one-to-one. Thus since $\mathcal{U} < \mathcal{V}$, which is finite dimensional, L must map $\mathcal{U}$ onto $\mathcal{U}$.

Now suppose $\vec{v} \in \mathcal{U}^\perp$ and $\vec{u} \in \mathcal{U}$. Then $\vec{u} = L(\vec{u}_1)$ for some $\vec{u}_1 \in \mathcal{U}$ from the fact that L is onto. But $\langle \vec{v} | \vec{u}_1 \rangle = 0$, so we would then get

$$0 = \langle L(\vec{v}) | L(\vec{u}_1) \rangle = \langle L(\vec{v}) | \vec{u} \rangle$$

Thus $L(\vec{v}) \in \mathcal{U}^\perp$, since $\vec{u}$ was an arbitrarily chosen member of $\mathcal{U}$.

∎

Theorem 15.3.2 *If $L : \mathcal{V} \to \mathcal{V}$ preserves inner products, then $\mathcal{V}$ can be written as the direct sum of L-cyclic subspaces.*

PROOF:

> We prove this by strong induction on the dimension of $\mathcal{V}$. If $\vec{V}$ is of dimension 1 it is L cyclic, so the theorem is true.
>
> Now suppose that $\dim(\mathcal{V}) > 1$. First pick any vector $\vec{v} \in \mathcal{V}$ and let $\mathcal{U}_1$ be the L-cyclic subspace it generates. The dimension of $\mathcal{U}_1$ will be at least one, so the dimension of $\mathcal{U}_1^\perp$ will be less than the dimension of $\mathcal{V}$. Since L preserves the inner product, its restriction to $\mathcal{U}_1^\perp$ is an inner product preserving map onto $\mathcal{U}_1^\perp$. That means, by the induction hypothesis, that $\mathcal{U}_1^\perp$ can be written as the direct sum of L-cyclic subspaces. Since $\mathcal{V} = \mathcal{U} \oplus \mathcal{U}_1^\perp$, this will give us $\mathcal{V}$ as the direct sum of L-cyclic subspaces as needed.
>
> ∎

Recall (from Chapter 11 section 2) that a basis can be chosen so that the matrix for a linear transformation L on an L-cyclic subspace is a companion matrix. This gives the following corollary:

Corollary 15.3.3 *Any orthogonal matrix is similar to a block diagonal matrix in which each block is a companion matrix.*

Now one of the nice properties of companion matrices is that their characteristic equations can be obtained by inspection: if the companion matrix $\mathbf{M}$ is built using a basis $(\vec{v_0}, L(\vec{v_0}), L^2(\vec{v_0}), \ldots, L^{n-1}(\vec{v_0}))$ where $L^n(\vec{v_0}) = a_0\vec{v_0} + a_1L(\vec{v_0}) + \ldots + a_{n-1}L^{n-1}(\vec{v_0})$ then the characteristic polynomial of $\mathbf{M}$ is $p(x) = (-1)^{n+1}(a_0 + a_1x + \ldots + a_{n-1}x^{n-1} - x^n)$. Now suppose we define

$$p(L) = (-1)^{n+1}(a_0\mathrm{Id} + a_1L + a_2L^2 + \ldots a_(n-1)L^{n-1} - L^n$$

as a linear transformation. We know by definition that $p(L)(\vec{v_0}) = \vec{0}$. Furthermore, linearity tells us that $p(L)(L(\vec{v})) = L(p(L)(\vec{v}))$, so we will get $p(L)(L^k(\vec{v_0})) = \vec{0}$ for each k. Since the companion matrix is found using a basis of vectors of the form $L^k(\vec{v_0})$, this tells us that $p(L)$ takes each basis vector to $\vec{0}$ and thus must be the zero transformation. This tells us that any companion matrix satisfies its characteristic polynomial. Since we have shown that any orthogonal matrix can be written as a direct sum of companion matrices we get the same result for orthogonal matrices.

Indeed, rational canonical forms can be found for all matrices, not just orthogonal ones, though the decomposition into direct summands is a bit more delicate. This form makes the following theorem evident:

Theorem 15.3.4 (Cayley-Hamilton) *Every linear transformation satisfies its characteristic equation.*

Exercises 15.3:
For each of the following matrices find the rational canonical form:

1. $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

2. $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{sqrt2} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{sqrt2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$

3. $\begin{bmatrix} \frac{1}{2} & 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{12}} \\ -\frac{1}{2} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{12}} \\ -\frac{1}{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{12}} \\ \frac{1}{2} & 0 & 0 & \frac{3}{\sqrt{12}} \end{bmatrix}$

Appendix A

Definitions of basic concepts

B-coordinate representation of a vector [1.2.2,8.2.2]:

If $B = (\vec{b}_1, \vec{b}_n)$ is an ordered basis for $\mathcal{V}$, then the B-coordinate representation of $\vec{a} = k_1\vec{b}_1 + \ldots + k_n\vec{b}_n$ is the column vector $\begin{bmatrix} k_1 \\ \vdots \\ k_n \end{bmatrix}$.

basis for $\mathbb{R}^2$ [1.1]:

A basis for $\mathbb{R}^2$ is a pair of vectors $\vec{b}_1$ and $\vec{b}_2$ such that any vector in $\mathbb{R}^2$ can be written in exactly one way as $k_1\vec{b}_1 + k_2\vec{b}_2$.

basis [8.2.1]:

A basis for a vector space $\mathcal{V}$ is a set B of vectors which is linearly independent and which spans $\mathcal{V}$.

characteristic polynomial of a matrix [13.1.3]:

The characteristic polynomial of the matrix $\mathbf{M}$ is

$$p_{\mathbf{M}}(\lambda) = \det(\mathbf{M} - \lambda\mathbf{I})$$

codomain [5.2.1]:

The codomain of a function $L : \mathcal{V} \to \mathcal{W}$ is $\mathcal{W}$.

column rank of a matrix [10.2.4]:

The column rank of a matrix is the dimension of its column space.

column space of a matrix [2.3.1,6.2.4,10.2.3]:

If $\mathbf{M}$ is a $m \times n$ matrix then the column space of $\mathbf{M}$ is the subspace of $\mathbb{R}^m$ spanned by the column vectors of $\mathbf{M}$.

companion matrix [11.2]:

The companion matrix for the polynomial $p(x) = x^n - (a_0 + a_1 x + \ldots + a_{n-1} x^{n-1}$ is the matrix

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & a_0 \\ 1 & 0 & \cdots & 0 & a_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_{n-1} \end{bmatrix}$$

dependent set [8.1.1]:

A set A of vectors in $\mathcal{V}$ is linearly dependent if and only if $\vec{0}$ can be written as a linear combination of vectors in A in which there are nonzero coefficients.

determinant of a 2×2 matrix [2.3.3,Exercises 2.1#6]:

The determinant of a 2×2 matrix is given by the formula

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

determinant of a matrix [Exercises 2.3#4,12.1.1]:

A determinant function

$$\det : n \times n\text{-matrices} \to \mathbb{R}$$

is a function with the following properties:

1. det is multiplicative: $\det(\mathbf{AB}) = \det \mathbf{A} \det \mathbf{B}$

2. $\det(\mathbf{A}^t) = \det(\mathbf{A})$
3. det is multilinear in rows
4. det is not identiacally 0

dimension [8.2.3]:

The dimension of a finite dimensional vector space is the number of elements in a basis for that space.

direct sum of subspaces [6.3.2]:

Let $\mathcal{V}$ be a vector space and let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of V which satisfy

$$\mathcal{V} = \mathcal{U} + \mathcal{W} \text{ and } \mathcal{U} \cap \mathcal{W} = \{\vec{0}\},$$

then $\mathcal{V}$ is said to be the **direct sum** of $\mathcal{U}$ and $\mathcal{W}$; this relation is indicated by writing

$$\mathcal{V} = \mathcal{U} \oplus \mathcal{W}.$$

domain [5.2.1]:

The domain of a function $L : \mathcal{V} \to \mathcal{W}$ is $\mathcal{V}$.

dot product [T]:

he dot product of vectors $\vec{a} = [a_1, \ldots, a_n]$ and $\vec{b} = [b_1, \ldots, b_n]$ is defined

1. geometrically as $\vec{a} \cdot \vec{b} = \|\vec{a}\| \; \|\vec{b}\| \cos(\theta)$ where θ is the angle between the vectors
2. algebraically as $\vec{a} \cdot \vec{b} = \sum_{i=1}^{n} a_i b_i$

dual space $\mathcal{V}^*$ [Example 6.1.6]:

$\mathcal{V}^* = \mathrm{Hom}(\mathcal{V}, \mathbb{R})$

eigenvalue [13.1.1]:

A number λ such that there is a non-zero vector $\vec{v}$ with

$$L(\vec{v}) = \lambda \vec{v}$$

eigenvector [13.1.1]:

A non-zero vector $\vec{v}$ such that $L(\vec{v}) = \lambda\vec{v}$.

elementary row operations [2.4,7.1]:

The elementary row operations are:

1. Swap rows i and j : $R_i \leftrightarrow R_j$
2. Multiply row i by r (where $r \neq 0$): rR_i
3. Add r times row i to row j: $R_j + rR_i$

equal algebraic vectors [1.1.2]:

Two vectors $\vec{a} = [a_1, a_2]$ and $\vec{b} = [b_1, b_2]$ are equal if $a_1 = b_1$ and $a_2 = b_2$.

equal geometric vectors [1.1.1]:

Two vectors will be considered equal if they have the same direction and magnitude or if both are the zero vector.

field [5.1.1]:

A field is a set F equipped with two binary operations $+ : F \times F \to F$ and $\times : F \times F \to F$ satisfying the following axioms for all a, b, and $c \in F$:

Closure:	$a + b \in F$	$a \times b \in F$
Associativity:	$(a + b) + c = a + (b + c)$	$(a \times b) \times c = a \times (b \times c)$
Commutativity:	$a + b = b + a$	$a \times b = b \times a$
Identity:	$\exists_{0 \in F} \forall_a (a + 0 = a)$	$\exists_{1 \in F} \forall_a (a \times 1 = a)$
Inverses:	$\forall_{a \in F} \exists_{-a} (-a + a = 0)$	$\forall_{a \neq 0} \exists_{\frac{1}{a} \in F} (a \times \frac{1}{a} = 1)$
Distributive:	$a \times (b + c) = (a \times b) + (a \times c)$	

finite dimensional [8.2.2]:

A vector space is finite dimensional if it has a finite basis.

formal power series [5.1]:

A formal power series is an expression of the form $\sum_{i=0}^{\infty} a_1 x^i$. There is no requirement of convergence.

Fourier series [14.4]:

The Fourier series up to n^{th} harmonics for a periodic function is its projection onto the subspace spanned by the functions $\sin(mx)$ and $\cos(mx)$ for $m \leq n$.

Hom$(\mathcal{V}, \mathcal{W})$ [Example 6.1.6]:

$\mathrm{Hom}(\mathcal{V}, \mathcal{W})$ is the vector space of linear transformations from $\mathcal{V}$ to $\mathcal{W}$.

homogeneous [Example 6.1.8, Example 6.1.9]:

A homogeneous equation is one in which the constant term is 0.

identity linear transformation [2.2]:

The identity linear transformation $Id : \mathcal{V} \to \mathcal{V}$ has value $Id(\vec{v}) = \vec{v}$.

identity matrix [2.2]:

The identity matrix $\mathbf{I}$ has 1's on the main diagonal and 0's elsewhere. It has the property that $\mathbf{IM} = \mathbf{M}$ and $\mathbf{MI} = \mathbf{M}$ for all $\mathbf{M}$ for which the products are defined.

image of a linear transformation [2.3.1,6.1]:

The image of a linear transformation $L : \mathcal{V} \to \mathcal{W}$ is the subspace of $\mathcal{W}$ consisting of all vectors of the form $L(\vec{v})$.

inconsistent system [Example 7.1.2]:

A system of equations with no solutions is called inconsistent.

independent [8.1.1]:

A set A of vectors in $\mathcal{V}$ is linearly independent if whenever a linear combination of elements of A has

$$c_1\vec{a}_1 + \ldots + c_n\vec{a}_n = \vec{0}$$

then all of the $c_i = 0$.

inner product [14.1.1]:

An inner product on a vector space $\mathcal{V}$ over the reals is a function from $\mathcal{V} \times \mathcal{V}$ to $\mathbb{R}$ whose value at $(\vec{a}, \vec{b})$ is denoted $\langle \vec{a} \mid \vec{b} \rangle$, satisfying the following axioms:

Symmetry:	$\langle \vec{a} \mid \vec{b} \rangle = \langle \vec{b} \mid \vec{a} \rangle$
Positive definiteness:	$\langle \vec{a} \mid \vec{a} \rangle \geq 0$ with equality if and only if $\vec{a} = \vec{0}$
Linearity:	$\langle \vec{a} \mid k\vec{b} \rangle = k\langle \vec{a} \mid \vec{b} \rangle$
	$\langle \vec{a} \mid \vec{b} + \vec{c} \rangle = \langle \vec{a} \mid \vec{b} \rangle + \langle \vec{a} \mid \vec{c} \rangle$

inverse of a linear transformation [10.1.1]:

The inverse of a linear transformation $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation $L^{-1} : \mathcal{W} \to \mathcal{V}$ such that $L \circ L^{-1} = id_{\mathcal{W}}$ and $L^{-1} \circ L = id_{\mathcal{V}}$.

inverse of a matrix [10.1.2]:

The inverse of a square matrix $\mathbf{M}$ isd a matrix $\mathbf{M}^{-1}$ such that $\mathbf{M}\mathbf{M}^{-1} = \mathbf{M}^{-1}\mathbf{M} = \mathbf{I}$.

invertible [10.1.1]:

A matrix or linear transformation is said to be invertible if and only if it has an inverse.

kernel of a linear transformation [2.3.2,6.1]:

The set of vectors $\vec{v}$ such that $L(\vec{v}) = \vec{0}$ is the kernel of L. It is a subspace of the domain of L.

L-cyclic subspace [11.2.1]:

The L-cyclic subspace of $\mathcal{V}$ generated by $\vec{v}$ is

$$C^L_{\mathcal{V}} = \mathrm{Span}(\{\vec{v}, L(\vec{v}), \ldots, L^{n-1}(\vec{v})\})$$

length of a vector in $\mathbb{R}^2$ [1.1.2]:

The length of the vector $\vec{a} = [a_1, a_2]$ is $\|\vec{a}\| = \sqrt{a_1^2 + a_2^2}$.

linear combination [2.3,6.2.1]:

Let A be a nonempty subset of vectors in $\mathcal{V}$. A linear combination of vectors in A is a vector $\vec{b}$ of the form $\vec{b} = c_1\vec{a}_1 + \ldots + c_m\vec{a}_m$, where $c_1, \ldots, c_m$ are scalars and $\vec{a}_1, \ldots, \vec{a}_m \in A$.

linear transformation [2.1.1,5.2.1]:

A function $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation if $L(\vec{v}_1 + \vec{v}_2) = L(\vec{v}_1) + L(\vec{v}_2)$ and $L(k\vec{v}) = kL(\vec{v})$.

L-invariant subspace [11.2.2]:

A subspace $\mathcal{U} < \mathcal{V}$ is called L-invariant it whenever $\vec{u} \in \mathcal{U}$ then $L(\vec{u}) \in \mathcal{U}$.

lower triangular matrix [10.3.1]:

A matrix with $a_{ij} = 0$ whenever $i < j$ is called lower triangular.

matrix for a linear transformation with respect to ordered bases on domain and codomain [2.2,9.1]:

The matrix for the linear transformation L with respect to the ordered basis $(\vec{d}_1, \ldots, \vec{d}_n)$ for the domain and $(\vec{c}_1, \ldots, \vec{c}_m)$ for the codomain has l_{ij} given by the coefficient of $\vec{c}_i$ in $L(\vec{b}_j)$.

matrix [2.2]:

A matrix is a rectangular array of numbers.

minor [12.2]:

The ij-minor of a matrix $\mathbf{M}$ is the matrix $\mathbf{M_{ij}}$ obtained by omitting the i^{th} row and the j^{th} column of $\mathbf{M}$.

nilpotent [11.2.3]:

A linear transformation L is nilpotent with index of nilpotence n if L^n is identically $\vec{0}$ but L^{n-1} is not.

norm [14.1.2]:

If $\langle \ \mid \ \rangle$ is an inner product on $\mathcal{V}$ and $\vec{a}$ is a vector in $\mathcal{V}$, then the norm of $\vec{a}$, denoted $\|\vec{a}\|$, is $\sqrt{\langle \vec{a} \mid \vec{a} \rangle}$.

nullity of a linear transformation [10.2.2]:

The nullity of a linear transformation $L : \mathcal{V} \to \mathcal{W}$ is the dimension of $\mathrm{Ker}(L)$.

orthogonal basis [14.2]:

An orthogonal basis is a basis whose elements are mutually orthogonal.

orthogonal complement $S^\perp$ [15.1.1]:

If S is a non-empty set of vectors in an inner product space $\mathcal{V}$, then the orthogonal complement $S^\perp$ is the set of all vectors orthogonal to all of the members of S.

orthogonal matrix [15.2.2]:

A square matrix $\mathbf{A}$ is called orthogonal if its column vectors form an orthonormal set.

orthogonal transformation [15.2.1]:

If $\mathcal{V}$ and $\mathcal{W}$ are inner product spaces and $L : \mathcal{V} \to \mathcal{W}$ is a linear transformation then L is an orthogonal transformation if and only if

$$\langle L\vec{a} \mid L\vec{b}\rangle = \langle a \mid b\rangle$$

for all $\vec{a}$, $\vec{b}$ in $\mathcal{V}$.

orthogonal [14.1.3]:

Two vectors $\vec{v}$ and $\vec{w}$ are said to be orthogonal if $\langle \vec{v}|\vec{w}\rangle = 0$.

orthonormal set [14.2.1]:

A set of vectors $\{\vec{v}_1, \dots, \vec{v}_n\}$ is called orthonormal if

$$\langle \vec{v}_i \mid \vec{v}_j\rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} .$$

pivot [7.1.2]:

If $\mathbf{M}$ is a matrix with $m_{ij} \neq 0$ then the pivot on the ij position of $\mathbf{M}$ is the sequence of row operations $\dfrac{1}{a_{ij}}R_i$ then $R_k - m_{kj}R_i$ for all $k \neq i$. It results in a new matrix with a 1 in the ij position and 0 in the rest of the jth column.

product of a matrix and a column vector [2.2,5.2]:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} ak_1 + bk_2 \\ ck_1 + dk_2 \end{bmatrix}$$

product of two matrices [2.2,9.3.1]:

The product $\mathbf{GF}$ of an $m \times n$ matrix $\mathbf{F}$ and a $p \times m$ matrix $\mathbf{G}$ is the $p \times n$ matrix with ij entry given by

$$\sum_{k=1}^{m} g_{ik} f_{kj}.$$

projection of a vector onto subspace [14.3.1]:

The projection of $\vec{v}$ onto the subspace $\mathcal{W}$ is the vector $\vec{w} \in \mathcal{W}$ such that $\vec{v} = \vec{w} + \vec{w}'$ with $\vec{w}'$ orthogonal to every vector in $\mathcal{W}$.

rank of a linear transformation [10.2.1]:

The rank of a linear transformation $L : \mathcal{V} \to \mathcal{W}$ is the dimension of ImL.

rank of a matrix [Theorem 10.2.5]:

The row rank and the column rank of a matrix are equal, hence we call either number the rank of $\mathbf{M}$.

redundant system [Example 7.1.3]:

A system of equations in which one or more of the equations provides no new information is called redundant.

row rank of a matrix [10.2.4]:

The row rank of a matrix $\mathbf{M}$ is the dimension of the row space of $\mathbf{M}$.

row space of a matrix [6.2.3,10.2.3]:

If $\mathbf{M}$ is a $m \times n$ matrix then the row space of $\mathbf{M}$ is the subspace of $\mathbb{R}^n$ spanned by the row vectors of $\mathbf{M}$.

row-reduced echelon form [7.1.1]:

A matrix is in row-reduced echelon form if and only if

1. The first nonzero entry in each row is a 1
2. The first nonzero entry in a row appears to the right of the first nonzero entry in the row above it
3. All other entries in the column of that first nonzero entry in the row are 0.
4. All rows with only 0 entries are at the bottom.

scalar multiple of a matrix [2.2,9.2]:

If A is an $m \times n$ matrix then kA has ij-entry ka_{ij}.

similar matrices [2.2.1,11.1.1,11.2]:

If two matrices represent the same linear transformation with respect to different choices of basis we say that they are similar matrices. This is usually restricted to square matrices using the same basis for domain and codomain; in this case $\mathbf{A} \sim \mathbf{B}$ if and only if there is an invertible matrix $\mathbf{P}$ with $\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P}$.

Span(A) [6.2.2]:

If A is not empty then the span of A, written Span(A), is the set of all linear combinations of vectors in A. If A is empty Span(A) is the vector space $\{\vec{0}\}$.

spanning set [6.2.2]:

If Span$(A) = \mathcal{V}$ then A is a spanning set for $\mathcal{V}$.

stochastic matrix [3.1 Exercise 22]:

A matrix with all entries non-negative in which the columns add up to 1.

subspace spanned by a set [2.3,6.2.2]:

Let A be a subset of $\mathcal{V}$. If A is not empty then the span of A, written Span(A), is the set of all linear combinations of vectors in A. If A is empty Span(A) is the vector space $\{\vec{0}\}$.

subspace [2.3,6.1.1]:

A subspace $\mathcal{W}$ of a vector space V is a subset which is a vector space using the same addition and scalar multiplication as in $\mathcal{V}$.

sum of subspaces [6.3.1]:

$$\mathcal{W}_1 + \mathcal{W}_2 = \{\vec{a} \in \mathcal{V} \mid \vec{a} = \vec{b}_1 + \vec{b}_2, \vec{b}_1 \in \mathcal{W}_1, \vec{b}_2 \in \mathcal{W}_2\}.$$

sum of two matrices [2.2,9.2]:

If A and B are $m \times n$ matrices then $A+B$ has ij-entry $a_{ij}+b_{ij}$.

symmetric matrix [Exercises 9.3#3]:

A symmetric matrix is a matrix with $m_{ij} = m_{ji}$ for all i, j.

transpose of a matrix [2.3 Exercise 21b]:

The transpose switches rows and columns:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^t = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

upper triangular matrix [10.3.1]:

A matrix with $a_{ij} = 0$ whenever $i > j$ is called upper triangular.

vector in the plane [1.1.2]:

A vector $\vec{a}$ in the plane is an ordered pair of real numbers $\vec{a} = [a_1, a_2]$.

vector space [5.1.2]:

A vector space over a field F (whose elements are called scalars) is a set $\mathcal{V}$ (whose elements are called vectors) which has two operations: $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and scalar multiplication $F \times \mathcal{V} \to \mathcal{V}$ (usually indicated by juxtaposition) which are required to satisfy the following axioms for all vectors $\vec{a}, \vec{b}, \vec{c}$, and scalars h and k:

Closure:	both $\vec{a} + \vec{b}$ and $k\vec{a}$ are vectors
Commutativity of $+$:	$\vec{a} + \vec{b} = \vec{b} + \vec{a}$
Associativity of $+$:	$(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$
Identity for $+$:	There is a unique vector $\vec{0}$ with $\vec{0} + \vec{a} = \vec{a}$ for all $\vec{a}$.
Inverses for $+$:	For each $\vec{a}$ there is a unique $-\vec{a}$ so that $\vec{a} + -\vec{a} = \vec{0}$
Absorption:	$h(k\vec{a}) = (hk)\vec{a}$
Distributivity:	$(h + k)\vec{a} = (h\vec{a}) + (k\vec{a})$
	$h(\vec{a} + \vec{b}) = (h\vec{a}) + (h\vec{b})$
Identity for scalars:	$1\vec{a} = \vec{a}$

zero vector [1.1.3]:

The zero vector $\vec{0}$ is the vector such that $\vec{v} + \vec{0} = \vec{v}$ for all $\vec{v}$.

$\mathbb{Z}_2$ [5.3]:

The field of integers using arithmetic modulo 2.

Appendix B

Answers to selected exercises

Chapter 1
Section 1.1

1. (a) $\vec{a} + 5\vec{c} = [28, -55]$

   (c) $\pi\vec{a} + 3(\vec{b} + \vec{d}) = [3\pi - 3, 5\pi + 12]$

3. Look at $-_1\vec{v} + \vec{v} + -_2\vec{v}$. The associative law tells us that

$$(-_1\vec{v} + \vec{v}) + -_2\vec{v} = -_1\vec{v} + (\vec{v} + -_2\vec{v}).$$

The facts that $-_1\vec{v}$ and $-_2\vec{v}$ are inverses for $\vec{v}$ tell us that

$$(-_1\vec{v} + \vec{v}) = \vec{0} \text{ and } (\vec{v} + -_2\vec{v}) = \vec{0}$$

so

$$-_2\vec{v} = \vec{0} + -_2\vec{v} = -_1\vec{v} + \vec{0} = -_1\vec{v}.$$

Section 1.2

1. (a) $[3, 0] = 2[1, 1] - 1[-1, 2]$

   (c) $[2, 5] = 3[1, 1] + 1[-1, 2]$

   (e) $[a, b] = x[1, 1] + y[-1, 2]$ has the unique solution

$$\begin{aligned} x &= \frac{2a + b}{3} \\ y &= \frac{b - a}{3} \end{aligned}$$

   Thus $\{[1, 1], [-1, 2]\}$ is a basis.

3. $c = 2$ and $d = 1$. We can also let $c = 0$ and $d = 0$.

5. $\vec{v} = \frac{-b}{a}\vec{w}$

7. Certainly if we let $x = 0$ and $y = 0$ we get a solution. Since exercise 5 proved that if the vectors are nonzero and nonparallel, then there is at most one solution, this is sufficient.

Chapter 2

Section 2.1

1. Not linear: $L([0,0]) = [3,-2]$ instead of $[0,0]$.

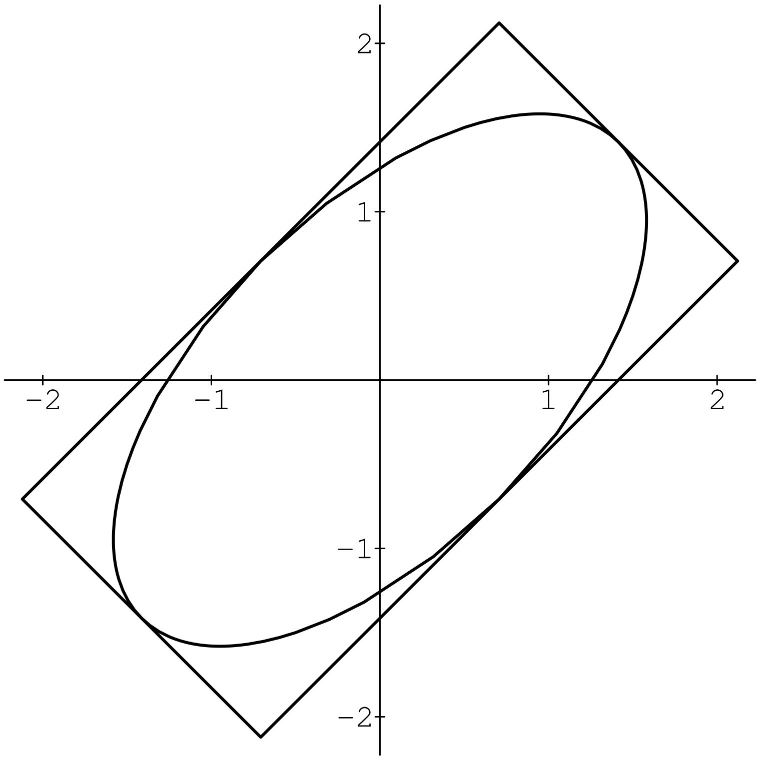
3. Linear

5. Not linear: $2L([1,1]) = [0,4]$ but $L([2,2]) = [0,8]$.

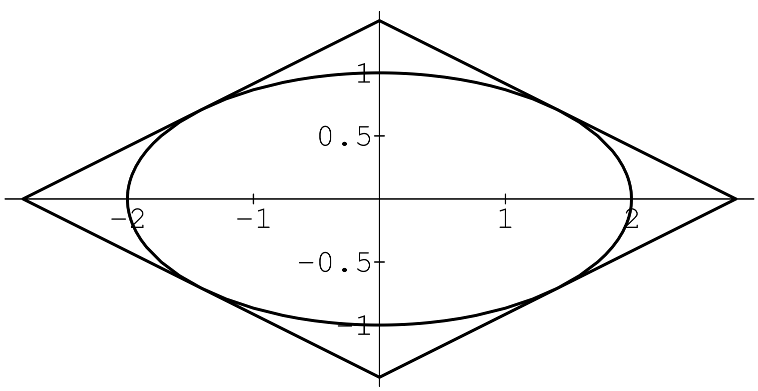
7. Not linear: $L([1,2]+[2,1]) = L([3,3]) = [0,0]$ but $L([1,2])+L([2,1]) = [1,1]+[1,1] = [2,2]$.

9. Linear

11. Let us look at what happens to a unit square:



Stretch first then rotate



Rotate first then stretch

12. If $L([a_1, b_1]) = L([a_2, b_2])$ then $[a_1, 2b_1] = [a_2, 2b_2]$, from which we conclude that $a_1 = a_2$ and $b_1 = b_2$, so L is one to one. To show it is onto we need to see that any vector $[a, b]$ is in the image of L: now $[a, b] = L([a, \frac{b}{2}])$, so L is onto. The inverse is the map taking $[a, b]$ to $[a, \frac{b}{2}]$.

Section 2.2

1. $\mathbf{AC} = \begin{bmatrix} 0 & 1 \\ -5 & -3 \end{bmatrix}$

3. $\mathbf{BC} = \begin{bmatrix} -1 & -2 \\ -7 & 1 \end{bmatrix}$

5. $\mathbf{B}(\mathbf{CA}) = \begin{bmatrix} 5 & -10 \\ -10 & 5 \end{bmatrix}$

7. $(\mathbf{A}+\mathbf{B})\mathbf{C} = \begin{bmatrix} -1 & -1 \\ -12 & -2 \end{bmatrix}$

9. (a) $\begin{bmatrix} 3 & 1 \\ -1 & 1 \end{bmatrix}$

   (b) $\begin{bmatrix} 2 & 2 \\ 0 & 2 \end{bmatrix}$

   (c) $\begin{bmatrix} 4 & 1 \\ -4 & 0 \end{bmatrix}$

   (d) $\begin{bmatrix} \frac{23}{19} & \frac{9}{19} \\ \frac{-25}{19} & \frac{53}{19} \end{bmatrix}$

11. (a) $\begin{bmatrix} 2 & 3 \\ -5 & -8 \end{bmatrix}$

   (b) $\begin{bmatrix} -2 & 9 \\ 1 & -4 \end{bmatrix}$

   (c) $\begin{bmatrix} 5 & 3 \\ -18 & -11 \end{bmatrix}$

   (d) $\begin{bmatrix} -\frac{144}{19} & \frac{37}{19} \\ -\frac{107}{19} & \frac{30}{19} \end{bmatrix}$

13. If we think of the matrix as representing a linear transformation L with respect to the standard basis, then the columns tell us where the standard basis vectors go. Since any vector in $\mathbb{R}^2$ can be written as $x\vec{c}_1 + y\vec{c}_2$ where the $\vec{c}_i$ are the column vectors, any vector in $\mathbb{R}^2$ can be written as $L([x, y])$. Thus L is onto.

17. This proof is just like the proof that $\mathbb{R}^2$ is a vector space, but it uses four components instead of two, since there are four positions in a 2×2 matrix.

Section 2.3

1. $\text{Span}(\{[0,0]\}) = \{[0,0]\}$

3. $\text{Span}(\{[1,2]\}) = \{[x,2x] | x \in \mathbb{R}\}$

5. $\text{Span}(\{[1,3],[1,2]\}) = \mathbb{R}^2$

7. $\text{Span}(\{[1,2],[-2,4],[3,-6]\}) = \mathbb{R}^2$

9. For these problems many answers are possible, including
   (a) $L([x,y]) = [x+y, x-y]$
   (b) $L([x,y]) = [0,0]$
   (c) $L([x,y]) = [x - \frac{1}{3}y, 0]$

11. Determinant = 6 so matrix has an inverse

13. Determinant = 0 so matrix does not have an inverse

15. Determinant = .04 so matrix has an inverse

17. Determinant = 5 so matrix has an inverse

19. Determinant = 14 so matrix has an inverse

21. If $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ then we can conclude that

$$\frac{bc - ad}{b} x = 0 \text{ and } y = -\frac{a}{b} x$$

Since $ad - bc = 0$ this puts no restriction on x.

22. (a) $\text{Ker}(L) = \{[0,0]\}$ and $\text{Im}(L) = \mathbb{R}^2$ There is an inverse $L^{-1}[x,y] = \frac{1}{9}[x+2y, 4x-y]$
    (c) $\text{Ker}(Z) = \mathbb{R}^2$ and $\text{Im}(Z) = \{[0,0]\}$ No inverse

Section 2.4

1. $\begin{bmatrix} 2 & -\frac{3}{4} \\ -1 & \frac{1}{2} \end{bmatrix}$

3. $\begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ -\frac{1}{2} & \frac{1}{4} \end{bmatrix}$

5. $\begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix}^{-1}$ does not exist

7. L^{-1} does not exist

9. L^{-1} does not exist

14. The algorithm using row operations takes 6 multiplications, 2 reciprocals, and 3 additions. The method using the formula takes 6 multiplications and one addition, so **for the 2 by 2 case** it is less work.

Chapter 3
Section 3.1

1. The eigenvalues are the roots of $(2-\lambda)(3-\lambda) - 1 \cdot 0 = 0$, so we get $\lambda = 2, 3$. An eigenvector for 2 is $[1, 0]$; for 3 we get $[1, 1]$. The system will expand fairly rapidly in the direction of both eigenvectors.

3. The eigevalues are the roots of $(2-\lambda)(1-\lambda) - 12 = 0$, giving $\lambda = 5$ or $\lambda = -2$. Corresponding eigenvectors are $[1, 1]$ and $[1, -\frac{4}{3}]$. We get expansion in the direction of the eigenvector for 5 and diverging oscillation in the direction of the eigenvector for -2.

5. Here the eigenvalue is .2 obtained twice. Eigenvectors all have the form $[x, 0]$. Iterations get closer to the origin.

7. The characteristic equation is $\lambda^2 - 3\lambda + 4 = 0$ which has complex roots $\lambda = \frac{3 \pm \sqrt{7}i}{2}$. These have modulus bigger than 1, so the system spirals outward.

9–16 In all cases the value of the characteristic polynomial applied to the original matrix is the zero matrix.

17. The eigenvalues are complex with modulus less than 1. The real part is negative.

19. The eigenvalues are complex with modulus greater than 1.

21. Both eigenvalues are real, positive, and less than 1.

Section 3.2

1. $\begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ using the basis $\{[1,0],[1,-\frac{1}{3}]\}$

3. $\begin{bmatrix} -2 & 3 \\ 1 & 1 \end{bmatrix} \sim \begin{bmatrix} \frac{-1-\sqrt{21}}{2} & 0 \\ 0 & \frac{-1+\sqrt{21}}{2} \end{bmatrix}$ using the basis $\{[1,\frac{3-\sqrt{21}}{6}],[1,\frac{3+\sqrt{21}}{6}]\}$

5. $\begin{bmatrix} 2 & -3 \\ 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ using the basis $\{[1,0],[1,\frac{1}{3}]\}$

7. $\begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \sim \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ using the basis $\{[1,1],[1,2]\}$

9. $\begin{bmatrix} 1 & -1 \\ 1 & 3 \end{bmatrix} \sim \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ using the basis $\{[-1,1],[2,-1]\}$

11. $\begin{bmatrix} 1 & 4 \\ -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$ using the basis $\{[0,2],[4,0]\}$

13. Since
$$k\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} k\lambda_1 & 0 \\ 0 & k\lambda_2 \end{bmatrix}$$
giving eigenvalues $k\lambda_1$ and $k\lambda_2$ and
$$k\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} k\lambda & k \\ 0 & k\lambda \end{bmatrix}$$
giving eigenvalues $k\lambda$ twice, and since
$$k\begin{bmatrix} a & b \\ -b & a \end{bmatrix} = \begin{bmatrix} ka & kb \\ -kb & ka2 \end{bmatrix}$$
giving eigenvalues $ka \pm kb\, i$, multiplying by k multiplies the eigenvalues by k.

15. (a) Suppose $\mathbf{M}$ is the identity matrix, which has 1 as an eigenvalue, and that

$$\mathbf{N} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

with eigenvalues 2 and 3. Then $\mathbf{MN} = \mathbf{N}$ which does not have 1 as an eigenvalue.

(b) The eigenvalues for

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$$

are 1 and 2. The eigenvalues for

$$\mathbf{N} = \begin{bmatrix} 3 & 1 \\ 0 & 4 \end{bmatrix}$$

are 3 and 4. The eigenvalues for

$$\mathbf{MN} = \begin{bmatrix} 3 & 1 \\ 3 & 9 \end{bmatrix}$$

are $6 \pm 2\sqrt{3}$, not obtained as a product of the eigenvalues of $\mathbf{M}$ and $\mathbf{N}$.

Chapter 4
Section 4.1

1. $\cos(\theta) = \frac{3}{\sqrt{10}}$

3. $\cos(\theta) = \frac{3}{\sqrt{14}}$

5. $\cos(\theta) = 1$

7. $\cos(\theta_x) = \frac{3}{5}$ and $\cos(\theta_y) = \frac{4}{5}$

9. $\cos(\theta_x) = \frac{1}{\sqrt{2}}$ and $\cos(\theta_y) = \frac{1}{\sqrt{2}}$

11. $\cos(A) = \frac{58}{\sqrt{73}\sqrt{61}}, \cos(B) = \frac{-33}{\sqrt{73}\sqrt{18}}, \cos(C) = \frac{3}{\sqrt{61}\sqrt{18}}$

13. $c = \frac{3}{4}$

15. $k = \frac{14}{20}$

17. Recall that $\|[a_1, a_2]\| = \sqrt{a_1^2 + a_2^2}$ so $\|[a_1, a_2]\|^2 = a_1^2 + a_2^2 = [a_1, a_2] \cdot [a_1, a_2]$.

19. We calculate

$$\begin{aligned} \|\vec{a} + \vec{b}\|^2 &= (\vec{a} + \vec{b}) \cdot (\vec{a} + \vec{b}) \\ &= (\vec{a} + \vec{b}) \cdot \vec{a} + (\vec{a} + \vec{b}) \cdot \vec{b} \\ &= \vec{a} \cdot \vec{a} + 2\vec{a} \cdot \vec{b} + \vec{b} \cdot \vec{b} \\ &= \|\vec{a}\|^2 + 2\vec{a} \cdot \vec{b} + \|\vec{b}\|^2 \\ &\leq \|\vec{a}\|^2 + 2\|va\|\|\vec{b}\| + \|\vec{b}\|^2 \\ &= (\|\vec{a}\| + \|\vec{b}\|)^2 \end{aligned}$$

Since both $\|\vec{a}\| + \|\vec{b}\|$ and $\|\vec{a} + \vec{b}\|$ are non-negative, we can conclude that $\|\vec{a} + \vec{b}\| \leq \|\vec{a}\| + \|\vec{b}\|$.

Section 4.2

1. $\overrightarrow{\text{proj}}_{\vec{u}}(\vec{v}) = \frac{11}{5}[1, 2] = [\frac{11}{5}, \frac{22}{5}]$

3. $\overrightarrow{\text{proj}}_{\vec{u}}(\vec{v}) = \frac{3}{29}[5, -2] = [\frac{15}{29}, -\frac{6}{29}]$

5. $\overrightarrow{\text{proj}}_{\vec{u}}(\vec{v}) = \frac{-7}{5}[-2, -1] = [\frac{14}{5}, \frac{7}{5}]$

7. $[3, 4] = [\frac{11}{5}, \frac{22}{5}] + [\frac{4}{5}, -\frac{2}{5}]$

9. $[1, 1] = [\frac{15}{29}, -\frac{6}{29}] + [\frac{14}{29}, \frac{35}{29}]$

11. $[2, 3] = [\frac{14}{5}, \frac{7}{5}] + [-\frac{4}{5}, \frac{8}{5}]$

13. Let $\vec{a} = [a_1, a_2], \vec{v} = [v_1, v_2]$, and $\vec{w} = [w_1, w_2]$ then

    (a) In general
    $$\vec{a} = \frac{v_2 a_1 - v_1 a_2}{w_1 v_2 - w_2 v_1}\vec{v} + \frac{w_2 a_1 - w_1 a_2}{v_1 w_2 - v_2 w_1}\vec{w}$$

    (b) For orthogonal vectors
    $$\vec{a} = \frac{\vec{v} \cdot \vec{a}}{\vec{v} \cdot \vec{v}}\ \vec{v} + \frac{\vec{w} \cdot \vec{a}}{\vec{w} \cdot \vec{w}}\ \vec{w}.$$

Chapter 5
Section 5.1

1. This is essentially $\mathbb{R}^{\mathbb{N}}$ which is a vector space as in Theorem 5.1.5.

3. Since the sum of two polynomials which have only even powers also has only even powers, we get closure. The zero polynomial has only even powers of x, so identity is OK. All of the other properties follow as for polynomials without the restriction.

5. This is $\mathbb{R}^3$ written with a different encoding.

7. We define + by $\vec{0}+\vec{0}=\vec{0}$ and scalar multiplication by $k\vec{0}=\vec{0}$.

Commutativity	$\vec{0}+\vec{0}=\vec{0}+\vec{0}$
Associativity	$(\vec{0}+\vec{0})+\vec{0}=\vec{0}+(\vec{0}+\vec{0})=\vec{0}$
Identity	$\vec{0}+\vec{0}=\vec{0}$
Inverses	$\vec{0}+\vec{0}=\vec{0}$
Absorption	$k(h\vec{0})=(kh)\vec{0}=\vec{0}$
Distributive	$(k+h)\vec{0}=k\vec{0}+h\vec{0}=\vec{0}$
	$k(\vec{0}+\vec{0})=k\vec{0}+k\vec{0}=\vec{0}+\vec{0}=\vec{0}$
Identity for multiplication	$1\vec{0}=\vec{0}$

We have checked, for each axiom, the only cases which occur. Thus $\{\vec{0}\}$ is a vector space.

9. We define **R**, the vector space, by using the + from the field as + for vectors and x from the field gives scalar multiplication. Commutativity for + in vectors is precisely commutativity for + in the field. Associativity for + in vectors is precisely associativity for + in the field. The associative law for multiplication in the field yields the absorption law for multiplication by a scalar. The distributive law $k(a+b)=ka+kb$ is the distributive law for the field. The distributive law $(k+h)a=ka+ha$ follows from it by commutativity of multiplication. Identity for multiplication in the field gives identity for scalar multiplication in the vector space.

11. We can think of $\mathbb{C}$ as numbers of the form $a+b\,i$. This is just another way to write $\mathbb{R}^2$.

13. This fails to satisfy the identity axiom for scalar multiplication. All of the other axioms are satisfied.

15. This fails to satisfy the identity axiom for scalar multiplication.

17. We consider the axioms in order:

Closure under $\oplus$	holds	
Closure under scalar multipication	holds	
Associativity of $\oplus$	holds	
Commutativity of $\oplus$	holds	
Absorption	fails	$.3 \cdot (30 \cdot .8) = .3 \neq .24 = .9 \cdot .8$
Identity under $\oplus$	holds	
Identity for scalar multiplication	holds	
Distributivity	holds	

Section 5.2

1. We show that both addition and scalar multiplication are preserved. Let $\vec{v} = [x, y]$ and $\vec{u} = [s, t]$ then

$$\begin{aligned} L(\vec{u} + \vec{v}) &= L([s + x, t + y]) \\ &= 3(s + x) - 4(t + y) \\ &= 3s + 3x - 4t - 4y \\ &= (3s - 4t) + (3x - 4y) \\ &= L(\vec{u}) + L(\vec{v}) \\ L(k\vec{u}) &= L([ks, kt]) \\ &= 3(ks) - 4(kt) \\ &= k(3s - 4t) \\ &= kL(\vec{u}) \end{aligned}$$

3. We show that both addition and scalar multiplication are preserved.

$$\begin{aligned} J(x_1 + x_2) &= [x_1 + x_2, 0] \\ &= [x_1, 0] + [x_2, 0] \\ &= J(x_1) + J(x_2) \\ J(kx) &= [kx, 0] \\ &= k[x, 0] \\ &= kJ(x) \end{aligned}$$

The function taking x t $[x, 1]$ does not preserve addition, so it is not linear. The function taking x to $[x, 2x]$ is linear.

5. We show that both addition and scalar multiplication are preserved. Let $\vec{p} = \begin{bmatrix} ab & \\ c & d \end{bmatrix}$ and $\vec{q} = \begin{bmatrix} s & t \\ u & v \end{bmatrix}$ then

$$
\begin{aligned}
M(\vec{p}+\vec{q}) &= M(\begin{bmatrix} a+sb+t & \\ c+u & d+v \end{bmatrix}) \\
&= \begin{bmatrix} a+s+2(b+t) \\ c+u+2(d+v) \end{bmatrix} \\
&= \begin{bmatrix} a+2b+s+2t) \\ c+2d+u+2v) \end{bmatrix} \\
&= \begin{bmatrix} a+2b \\ c+2d) \end{bmatrix} + \begin{bmatrix} s+2t \\ u+2v \end{bmatrix} \\
&= M(\vec{p}) + L(\vec{q}) \\
M(k\vec{p}) &= M(\begin{bmatrix} ka & 2kb \\ kc & kd \end{bmatrix}) \\
&= \begin{bmatrix} ka+2kb) \\ kc+2kd \end{bmatrix} \\
&= kM(\vec{u})
\end{aligned}
$$

7.

$$
\begin{aligned}
g([x,y,z]+[x',y',z']) &= g[x+x', y+y', z+z'] \\
&= x+x'+y+y'-2z-2z') \\
&= g[x,y,z]+g[x',y',z'] \\
g(k[x,y,z]) &= g[kx, ky, kz] \\
&= kx+ky-2kz \\
&= k(x+y-2z) \\
&= kg[x,y,z]
\end{aligned}
$$

9. $k(p+q) = p(x+1)+q(x+1) = (p+q)(x+1)$ and $k(rp) = (rp)(x+1) = r(p(x+1)) = rk(p)$ so k is linear.

11. $m[0,0] = [1,0,0] \neq \vec{0}$ so this is not linear

13. If $L(\vec{a})$ and $L(\vec{b})$ both equal $\vec{0}$ then $L(\vec{a}+\vec{b}) = L(\vec{a}) + L(\vec{b}) = \vec{0}+\vec{0} = \vec{0}$ and $L(k\vec{a}) = kL(\vec{a}) = k\vec{0} = \vec{0}$.

15. If L is 1-1 then $L(\vec{a}) = L(\vec{b})$ implies $\vec{a} = \vec{b}$ so if $L(\vec{a}) = \vec{0}$ then $\vec{a} = \vec{0}$ since we know $L(\vec{0}) = \vec{0}$. Now suppose $L(\vec{a}) = \vec{0} \Rightarrow \vec{a} = \vec{0}$. Then if $L(\vec{b}) = L(\vec{b}')$ we know $L(\vec{b}-\vec{b}') = \vec{0}$ so $\vec{b}-\vec{b}' = \vec{0}$ and $\vec{b} = \vec{b}'$.

Section 5.3

1.

$$\begin{aligned}
((a+bi)(c+di))(e+fi) &= ((ac-bd)+(ad+bc)i)(e+fi) \\
&= ((ace-bde-adf-bcf)+(ade+bce+acf-bdf)i \\
(a+bi)((c+di)(e+fi)) &= (a+bi)((ce-df)+(ed+fc)i) \\
&= (ace-adf-bed-bfc)+(aed+afi+bce-bdf)i \\
&= (ace-bde-adf-bcf)+(ade+bce+acf-bdf)i
\end{aligned}$$

so associativity holds.

3.

$$\begin{aligned}
((a_1+b_1\ i)+(a_2+b_2\ i))^* &= ((a_1+a_2)+(b_1+b_2)\ i)^* \\
&= ((a_1+a_2)-((b_1+b_2)\ i) \\
&= (a_1-b_1\ i)+(a_2-b_2\ i) \\
&= (a_1+b_1\ i)^*+(a_2+b_2\ i)^* \\
((a_1+b_2\ i)\times(a_2+b_2\ i))^* &= ((a_1a_2-b_1b_2)+(a_1b_2+a_2b_1)\ i)^* \\
&= ((a_1a_2-b_1b_2)-(a_1b_2+a_2b_1)\ i) \\
&= ((a_1a_2-(-b_1)(-b_2))+(a_1(-b_2)+a_2(-b_1))\ i) \\
&= (a_1-b_1\ i)\times(a_2-b_2\ i) \\
&= (a_1+b_1)^*\times(a_2+b_2\ i)^*
\end{aligned}$$

5. Since $\mathbb{R}$ is a subfield of $\mathbb{C}$, scalar mulitplication properties for real scalars follow from those for complex scalars.

7.

a	b	c	ab + ac	a(b+c)
1	1	1	0	0
1	1	0	1	1
1	0	1	1	1
1	0	0	0	0
0	1	1	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	0	0

Chapter 6
Section 6.1

1. (a) subspace
   (c) subspace
   (e) subspace
   (g) subspace

3. a) and b) are subspaces

5. Since the meaning of the operations is the same in all three sets all we need to do to show that $\mathcal{U} < \mathcal{W}$ is to show that it is a subset of $\mathcal{W}$.

7. functions with integral 0 (there isn't much more you can say about them)

9. This is the kernel of a linear transformation.

11. Any scalar multiple of a periodic function of period p is also periodic with period p. The sum of two functions which are periodic with period p also is periodic with period p. Thus the set of functions of period p is a subspace. Notice that we could not say the same for functions of prime period p or just periodic functions without specifying the period.

Section 6.2

1. (a) $Span(\{[1,2,3],[1,2,4]\} = \{[h, 2h, k] | h, k \in \mathbb{R}\}$
   (c) $\{[h-k, h, 2h+3k] | h, k \in \mathbb{R}\}$ note that [1,2,7] is of this form.

3. It is clear that $Span(A_1) \subseteq Span(A)$ since every linear combination of elements of A_1 *is* a linear combination of elements of A. Now suppose we have an element of $Span(A)$, say

$$\begin{aligned}\sum_{i=1}^{m} k_i\vec{a}_i &= \sum_{i=1}^{m-1} k_i\vec{a}_i + k_m(\sum_{i=1}^{m-1} c_i\vec{a}_i \\ &= \sum_{i=1}^{m-1}(k_i + k_m c_i)\vec{a}_i\end{aligned}$$

This shows that $Span(A) \subseteq Span(A_1)$ finishing the proof.

5. $\text{Span}(S) = \mathbb{R}^{\mathbb{N}}$

7. Eventually constant sequences form a subspace, so $\text{Span}(S) = S$.

9. A counterexample: in $\mathbb{R}^2$ let $S = \{[1,0]\}$ and $T = \{[0,1]\}$. Then $\text{Span}(S \cup T) = \mathbb{R}^2$ but $\text{Span}(S) \cup \text{Span}(T)$ is just the axes.

11. Suppose $\vec{v} = \sum_{i=1}^{n} a_i\vec{s}_i$ and $\vec{v} = \sum_{k=1}^{m} b_k\vec{t}_k$ are two different ways to get $\vec{v}$ as a linear combination of elements of S. Then the zero linear combination (all elements of S have coefficient 0 and the non-trivial linear combination $\sum_{i=1}^{n} a_i\vec{s}_i - \sum_{k=1}^{m} b_k\vec{t}_k$ both give $\vec{0}$.

Section 6.3

1. Since the difference of codewords is a codeword, it suffices to show that the non-zero codewords all have at least 3 bits non-zero. The 16 codewords are [0,0,0,0,0,0,0], [1,1,0,1,0,0,1], [0,1,0,1,0,1,0], [1,0,0,0,0,1,1], [1,0,0,1,1,0,0], [0,1,0,0,1,0,1], [1,1,0,0,1,1,0], [0,0,0,1,1,1,1], [1,1,1,0,0,0,0], [0,0,1,1,0,0,1], [1,0,1,1,0,1,0], [0,1,1,0,0,1,1], [0,1,1,1,1,0,0], [1,0,1,0,1,0,1], [0,0,1,0,1,1,0], and [1,1,1,1,1,1,1] all of which have at least three bits 1 except for the zero word.

2. (c) Correcting errors gives the string 0000000 1110000 1110000 1110000 which decodes as 0000 1000 1000 1000 .

3. Codewords for the Hamming(8,4) code are the solutions to the system of equations.

$$x_1 + x_2 + x_5 = 0$$

$$\begin{aligned} x_3 + x_4 + x_6 &= 0 \\ x_1 + x_3 + x_7 &= 0 \\ x_2 + x_4 + x_8 &= 0. \end{aligned}$$

Adding two such solutions gives a solution, $[0,0,0,0,0,0,0]$ is a solution so scalar multiplication stays in solutions. All other properties follow from those for $\mathbf{Z}_2^8$. No two bits occur together in all equations in which they occur at all. Thus the pattern of equations which fail determines which bit is wrong. For instance, suppose we get 10111101, then

$$\begin{aligned} x_1 + x_2 + x_5 &= 0 \\ x_3 + x_4 + x_6 &= 1 \\ x_1 + x_3 + x_7 &= 0 \\ x_2 + x_4 + x_8 &= 0. \end{aligned}$$

Thus the bit in error must be one which occurs only in the second equation. Hence it is bit 6 and the message was 1011.

Section 6.4

1. $\mathcal{U} + \mathcal{W} = \{[a_1, a_2, 0] | a_1, a_2 \in \mathbb{R}\}$ and $\mathcal{U} \cap \mathcal{W} = \{[0,0]\}$.

3. $\mathcal{U} + \mathcal{W} = \mathbb{R}^3$ and $\mathcal{U} \cap \mathcal{W} = \{[0,t,0] | t \in \mathbb{R}\}$.

5. $\mathcal{U} + \mathcal{W} = \mathbb{R}^3$ and $\mathcal{U} \cap \mathcal{W} = \{[2t, t, -3t] | t \in \mathbb{R}\}$.

7. $\mathcal{U} + \mathcal{W} = \mathbb{R}[x]$ and $\mathcal{U} \cap \mathcal{W} = \{\text{constant polynomials}\}$

9. $\mathcal{U} + \mathcal{W} = \mathbb{R}^{\mathbb{N}}$ and $\mathcal{U} \cap \mathcal{W} = \{\text{the constant sequence } 0\}$

11. It is true that $\mathcal{U} \cup \mathcal{W}$ is a subspace if $\mathcal{U} < \mathcal{W}$ or $\mathcal{W} < \mathcal{U}$

Chapter 7
Section 7.1

1. $x_1 = 1$ $x_2 = 2$ $x_3 = 0$

3. $x_1 = 0$ $x_2 = 1$ and $x_3 = 1$.

5.

$$\begin{array}{lcl} x_1 & = & 2 - x_4 \\ x_2 & = & 4 \\ x_3 & = & -3 - x_4 \\ x_4 & & \text{arbitrary} \end{array}$$

7. $x_1 = 2$, $x_2 = 5$, $x_3 = -3$

9. $x_1 = -1$, $x_2 = 10$, $x_3 = 2$

11. $x_1 + 2x_3 = 3$, $x_2 + x_3 = 4$, x_3 arbitrary

13. no solutions

15. no solutions

17. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

19. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Section 7.2

1. $[1, 1, 3]x + [-2, 1, -3]y = [1, 2, 3]$ has no solutions, so no.

3. $[6, 5, 6] = 2[1, 2, 4] - 3[0, 1, 2] + 4[1, 1, 1]$, so yes.

5. no

7. $[0, 1, 0, 1] = [1, 1, 0, 0] - [1, 0, 1, 0] + [0, 0, 1, 1]$. Yes.

9. $x^4 + x^2 + 1 = (x^4 + x^3 + 1) - (x^3 + 1) + (x^2 + x + 1) - x$

11. If the polynomial is $A + Bx + Cx^2 + Dx^3$ then four distinct roots x_1, x_2, x_3, x_4 give four equations in the four unknowns A, B, C, D:

$$\begin{aligned} A + Bx_1 + Cx_1^2 + Dx_1^3 &= 0 \\ A + Bx_2 + Cx_2^2 + Dx_2^3 &= 0 \\ A + Bx_3 + Cx_3^2 + Dx_3^3 &= 0 \\ A + Bx_4 + Cx_4^2 + Dx_4^3 &= 0 \end{aligned}$$

Row reduction gives:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 \\ 1 & x_2 & x_2^2 & x_2^3 & 0 \\ 1 & x_3 & x_3^2 & x_3^3 & 0 \\ 1 & x_4 & x_4^2 & x_4^3 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 \\ 0 & 1 & x_2 + x_1 & x_2^2 + x_2x_1 + x_1^2 & 0 \\ 0 & 0 & 1 & x_3 + x_2 + x_1 & 0 \\ 0 & 0 & 0 & x_4 - x_3 & 0 \end{bmatrix}$$

Thus the only solution is the trivial one.

Section 7.3

1. If the polynomial is $A + Bx + Cx^2 + Dx^3$ then four distinct roots x_1, x_2, x_3, x_4 give four equations in the four unknowns A, B, C, D:

$$\begin{aligned} A + Bx_1 + Cx_1^2 + Dx_1^3 &= 0 \\ A + Bx_2 + Cx_2^2 + Dx_2^3 &= 0 \\ A + Bx_3 + Cx_3^2 + Dx_3^3 &= 0 \\ A + Bx_4 + Cx_4^2 + Dx_4^3 &= 0 \end{aligned}$$

Row reduction gives:

$$\begin{aligned} \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 \\ 1 & x_2 & x_2^2 & x_2^3 & 0 \\ 1 & x_3 & x_3^2 & x_3^3 & 0 \\ 1 & x_4 & x_4^2 & x_4^3 & 0 \end{bmatrix} &\rightsquigarrow \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 \\ 0 & x_2 - x_1 & x_2^2 - x_1^2 & x_2^3 - x_1^3 & 0 \\ 0 & x_3 - x_1 & x_3^2 - x_1^2 & x_3^3 - x_1^3 & 0 \\ 0 & x_4 - x_1 & x_4^2 - x_1^2 & x_4^3 - x_1^3 & 0 \end{bmatrix} \\ &\rightsquigarrow \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 \\ 0 & 1 & x_2 + x_1 & x_2^2 + x_2x_1 + x_1^2 & 0 \\ 0 & 1 & x_3 + x_1 & x_3^2 + x_3x_1 + x_1^2 & 0 \\ 0 & 1 & x_4 + x_1 & x_4^2 + x_4x_1 + x_1^2 & 0 \end{bmatrix} \end{aligned}$$

$$\rightsquigarrow \quad \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 \\ 0 & 1 & x_2+x_1 & x_2^2+x_2x_1+x_1^2 & 0 \\ 0 & 0 & x_3-x_2 & x_3^2-x_2^2+(x_3-x_2)x_1 & 0 \\ 0 & 0 & x_4-x_2 & x_4^2-x_2^2+(x_4-x_2)x_1 & 0 \end{bmatrix}$$

$$\rightsquigarrow \quad \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 \\ 0 & 1 & x_2+x_1 & x_2^2+x_2x_1+x_1^2 & 0 \\ 0 & 0 & 1 & x_3+x_2+x_1 & 0 \\ 0 & 0 & 1 & x_4+x_2+x_1 & 0 \end{bmatrix}$$

$$\rightsquigarrow \quad \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 \\ 0 & 1 & x_2+x_1 & x_2^2+x_2x_1+x_1^2 & 0 \\ 0 & 0 & 1 & x_3+x_2+x_1 & 0 \\ 0 & 0 & 0 & x_4-x_3 & 0 \end{bmatrix}$$

Thus the only solution is the trivial one.

3. This involves solving the system of equations

$$\begin{aligned} a_1+a_2+a_3+a_4 &= 3h \\ ha_2+2ha_3+3ha_4 &= \frac{9h^2}{2} \\ h^2a_2+4h^2a_3+9h^2a_4 &= 9h^3 \\ h^3a_2+8h^3a_3+27h^3a_4 &= \frac{81h^4}{4} \\ h^4a_2+16h^4a_3+81h^4a_4+24c &= \frac{243h^5}{5} \end{aligned}$$

which has solution $a_1 = \frac{3h}{8}, a_2 = \frac{9h}{8}, a_3 = \frac{9h}{8}, a_4 = \frac{3h}{8}, c = -\frac{3h^5}{80}$.

5. General form is

$$Ax^3+Bx^2+Cx+D = y$$

For $(0,0)$ to be on the curve says

$$A0+B0+C0+D = 0.$$

For $(1,0))$ to be on the curve says

$$A+B+C+D = 0.$$

For $(2,4)$ to be on the curve says

$$8A+4B+2C+D = 4.$$

For $(3, 0)$ to be on the curve says

$$27A + 9B + 3C + D = 0.$$

Solving this system of equations gives $A = -2$, $B = 8$, $C = -6$ and $D = 0$. The cubic is y = -2x3 + 8x2 - 6x + 0

7.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{-1}{4} & 1 & \frac{-3}{4} & 0 & 0 & 0 \\ 0 & \frac{-1}{4} & 1 & \frac{-3}{4} & 0 & 0 \\ 0 & 0 & \frac{-1}{4} & 1 & \frac{-3}{4} & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \frac{27}{40} \\ 0 & 0 & 1 & 0 & 0 & \frac{9}{10} \\ 0 & 0 & 0 & 1 & 0 & \frac{39}{40} \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

9. $\dfrac{2}{x-1} - \dfrac{1}{(x-1)^2} + \dfrac{3}{x^2+1} + \dfrac{x-2}{(x^2+1)^2}$

11. $h_m = \frac{8}{13}$

Chapter 8
Section 8.1

1. Independent

3. Dependent

5. Independent

7. Dependent

9. Independent

11. Independent

13. Independent

15. (a) The linear combination of elements of A which demonstrates its dependence is also a non-trivial linear combination of vectors in B giving $\vec{0}$.

    (b) This is logically equivalent to a) using a contrapositive argument.

17. Suppose that $c_1\vec{a}_1 + \ldots + c_n\vec{a}_n = 0$ with the $\vec{a}_i$ in $S_1 \cup S_2$ and the $c_i \neq 0$. Then since both S_1 and S_2 are independent, there must be $\vec{a}_i$ from both sets. If we assume that $\vec{a}_1 \ldots \vec{a}_m$ are from S_1 and $\vec{a}_{m+1} \ldots \vec{a}_n$ are from S_2, then

$$\vec{v} = \sum_{i=1}^{m} c_i\vec{a}_i = \sum_{j=m+1}^{n} -c_j\vec{a}_j$$

is in $\mathrm{Span}(S_1) \cap \mathrm{Span}(S_2)$ and is thus $\vec{0}$. Independence of S_1 tells us that all of the $c_i = 0$ for $i = 1 \ldots m$ and independence of S_2 tells us that $c_j = 0$ for $j = m+1 \ldots n$. Thus we have a contradiction, so no nontrivial linear combination can give $\vec{0}$.

To see that it is not enough to ask for $S_1 \cap S_2 = \emptyset$ let $\mathcal{V} = \mathbb{R}^2$, $S_1 = \{[1,0],[0,1]\}$ and $S_2 = \{[1,1],[1,-1]\}$. Both S_1 and S_2 are independent and $S_1 \cap S_2 = \emptyset$, but $S_1 \cup S_2 = \{[1,0],[0,1],[1,1],[1,-1]\}$ is not independent.

19. For ease in numbering, let us assume that the set of columns involved is all of the $m \times n$ matrix $\mathbf{A}$(otherwise we will just need to avoid talking about certain indices). If the columns are dependent then there will be constants c_i, not all 0, such that

$$\sum_{i=1}^{n} c_i\mathbf{A}_{.i} = \vec{0}$$

which is the same as asking that for each $j = 1, \ldots, m$

$$\sum_{i=1}^{n} c_i a_{ji} = 0.$$

This is a homogeneous system of equations. Elementary row operations do not change the solutions to the system, so they cannot change independence of the column vectors.

21. Each non-zero row of a row reduced echelon form matrix has a leading 1 which is the only non-zero entry in its column, hence there is no way to get that entry as a linear combination of the corresponding entries of the other rows.

Section 8.2

1. Yes

3. No, not independent

5. Yes

7. No, doesn't span

9. Yes

11. If $\vec{v}_1, \ldots, \vec{v}_n$ is a basis for $\mathcal{V}$ over $\mathbb{C}$, then $\vec{v}_1, \ldots, \vec{v}_n, i\vec{v}_1, \ldots, i\vec{v}_n$ is a basis for $\mathcal{V}$ over $\mathbb{R}$.

13. To see independence, suppose that
$$\sum_{i=0}^{n} c_i f_i = 0.$$
Then
$$\sum_{i=0}^{n} c_i f_i(k) = c_k = 0$$
for each k.

    The sequences which can be obtained as linear combinations of f_i have only a finite number of non-zero values.

15. Since any vector in $\vec{v} \in \mathcal{V}$ can be written as
$$\begin{aligned} \vec{v} &= \vec{w} + \vec{u} \\ &= \sum_{i=1}^{n} x_i \vec{b}_i + \sum_{j=1}^{m} y_j \vec{c}_j \end{aligned}$$
$\mathcal{V}$ is spanned by $\{\vec{b}_1, \ldots, \vec{b}_n, \vec{c}_1, \ldots, \vec{c}_m\}$. Since $\mathcal{W} \cap \mathcal{U} = \{\vec{0}\}$ and both $\{\vec{b}_1, \ldots, \vec{b}_n\}$ and $\{\vec{c}_1, \ldots, \vec{c}_m\}$ are independent, $\{\vec{b}_1, \ldots, \vec{b}_n, \vec{c}_1, \ldots, \vec{c}_m\}$is linearly independent. Thus it is a basis for $\mathcal{V}$.

Section 8.3

1. $\{[3, -1], [1, 0]\}$

3. $\{x - 1, 1, x^2\}$

5. $\{[1,1,-1,1],[2,-1,0,1],[1,0,0,0],[0,1,0,0]\}$

7. $\{[1,-2,3],[-4,5,6],[7,8,-9]\}$

9. $\{1, x+1, x^2+1, x^3+x^2\}$

11. Basis for $\mathcal{U} \cap \mathcal{W}$ is

$$\{[1,2,3,1,2],[1,0,1,0,1\}$$

Basis for $\mathcal{U} + \mathcal{W}$ is

$$\{[1,2,3,1,2],[1,0,1,0,1],[1,1,1,1,1],[1,2,3,4,5],[1,1,1,1,2]\}$$

13. If B is a basis and $B' = B \setminus \{\vec{b}\}$ is linearly independent then $\vec{b}$ is not in $S(B)$. But B is a spanning set, giving a contradiction.

15. Consider the xy plane and the yz plane in $\mathbf{R}^3$, dim (xy plane) = dim (yz plane) = 2 but the two planes are not the same.

17. Since the non-zero rows in a row reduced echelon form matrix are independent and span the row space of that matrix, they form a basis for the row space of the row reduced echelon form of $\mathbf{M}$. But that is the same as the row space of $\mathbf{M}$, so it gives the desired basis.

18. Row reduction gives

(a) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ so the basis is $\{[1,0,-1],[0,1,2]\}$

(c) $\begin{bmatrix} 1 & 4 & 3 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 4 & 5 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ so the basis is $\{[1,0,0],[0,1,0],[0,0,1]\}$

Chapter 9
Section 9.1

1. $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix} = \begin{bmatrix} 11 \\ -3 \end{bmatrix}$

3. $\begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -3 \\ 7 \end{bmatrix} = \begin{bmatrix} 14 \\ -3 \end{bmatrix}$
$11[1,0]+-3[0,1] = [11,-3]$, $11[1,0]+-3[0,1] = [11,-3]$, and $14[1,0]+ -3[1,1] = [11,-3]$, so these all give the same answer.

5. $\begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -3 \\ 0 & 0 & 3 & 3 \end{bmatrix}$ Now $[2,1,-3,4] = 0[1,0,0,0]+3[1,1,0,0]-1[1,2,3,0]+ 4[0,0,0,1]$ so we calculate

$$\begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -3 \\ 0 & 0 & 3 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 3 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 10 \\ -8 \\ 9 \end{bmatrix}$$

Giving $10[1,0,0] - 8[1,1,0] + 9[0,1,1] = [2,1,9]$.

7. $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 7 \\ -1 \\ 6 \\ -4 \end{bmatrix}$

9. $\begin{bmatrix} 9 & 11 & 13 \\ 14 & 14 & 18 \\ 13 & 15 & 17 \\ 6 & 6 & 8 \end{bmatrix}$ is the matrix.

11. $T[1,2,3] = 1[3,2] + 2[1,4] + 3[0,0] = [5,10]$

13. $T[1,2,3,4] = T[1,1,1,1] + T[0,1,1,1] + T[0,0,1,1] + T[0,0,0,1] = [1,2] + [2,2] + [3,5] + [0,-3] = [6,7]$

Section 9.2

1. (a) The matrix for f is $\begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 3 \end{bmatrix}$ and the one for g is $\begin{bmatrix} 0 & -2 \\ 3 & 1 \\ 6 & 0 \end{bmatrix}$

   (b) $f+g([x,y]) = [x-y, 2x+2y, 6x+3y]$ and $3g([x,y]) = [-6y, 9x+3y, 18x]$

   (c) routine

3.

$$
\begin{aligned}
\mathbf{AB} &= \begin{bmatrix} 10 & 15 \\ 3 & 18 \\ 22 & 43 \end{bmatrix} \\
\mathbf{AC} &= \begin{bmatrix} 18 & 10 \\ 23 & 13 \\ 48 & 31 \end{bmatrix} \\
\mathbf{B}+\mathbf{C} &= \begin{bmatrix} 1 & 3 \\ 1 & 1 \\ 5 & 2 \\ 5 & 7 \end{bmatrix} \\
\mathbf{A}(\mathbf{B}+\mathbf{C}) &= \begin{bmatrix} 28 & 25 \\ 26 & 31 \\ 70 & 74 \end{bmatrix} \\
\mathbf{AB}+\mathbf{AC} &= \begin{bmatrix} 28 & 25 \\ 26 & 31 \\ 70 & 74 \end{bmatrix}
\end{aligned}
$$

3. (a) $x^2 + 5xy + 4y^2$

   (c) $x^2 - xy + y^2$

   (d) $\begin{bmatrix} 2 & -2 \\ -2 & 1 \end{bmatrix}$

5. $\begin{bmatrix} 4 & 8 \\ 12 & 16 \end{bmatrix} \begin{bmatrix} 0 & 1 & 4 \\ 1 & 0-2 & \end{bmatrix}$ is defined $\begin{bmatrix} 0 & 1 & 4 \\ 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} 4 & 8 \\ 12 & 16 \end{bmatrix}$ is not.
   Even if both are defined they need not give the same answer:

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

but

$$\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}.$$

7.

$$
\begin{aligned}
f(g+h)(\vec{v}) &= f(g(\vec{v}) + h(\vec{v})) \\
&= f(g(\vec{v})) + f(h(\vec{v})) \\
&= (fg) + (fh)(\vec{v})
\end{aligned}
$$

Chapter 10
Section 10.1

1. $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$ When you multiply on the right this swaps columns 3 and 5.

3. $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ When you multiply on the right this multiplies column 4 by -2.

5. $\begin{bmatrix} \frac{1}{10} & \frac{1}{10} & \frac{7}{10} \\ -\frac{3}{20} & \frac{7}{20} & -\frac{11}{20} \\ \frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{bmatrix}$

7. $\begin{bmatrix} -\frac{3}{10} & -\frac{1}{5} & \frac{3}{5} \\ \frac{3}{10} & -\frac{4}{5} & -\frac{2}{5} \\ \frac{1}{10} & \frac{2}{5} & -\frac{1}{5} \end{bmatrix}$

9. $\begin{bmatrix} -\frac{1}{14} & \frac{1}{7} & \frac{13}{14} \\ \frac{2}{21} & \frac{1}{7} & -\frac{4}{7} \\ \frac{1}{14} & -\frac{1}{7} & \frac{1}{14} \end{bmatrix}$

11. If $L^{-1}(\vec{u}_1) = \vec{v}_1$ then $L(\vec{v}_1) = \vec{u}_1$ and if $L^{-1}(\vec{u}_2) = \vec{v}_2$ then $L(\vec{v}_2) = \vec{u}_2$. Now since L preserves sums

$$\begin{aligned} L(\vec{v}_1 + \vec{v}_2) &= L(\vec{v}_1) + L(\vec{v}_2) \\ &= \vec{u}_1 + \vec{u}_2 \\ &\text{so} \\ \vec{v}_1 + \vec{v}_2 &= L^{-1}(\vec{u}_1 + \vec{u}_1) \\ L^{-1}(\vec{u}_1) + L^{-1}(\vec{u}_2) &= L^{-1}(\vec{u}_1 + \vec{u}_1) \end{aligned}$$

13. Since L is one to one it will take a basis of $\mathcal{V}$ to a linearly independent set. Extend that set to a basis for $\mathcal{W}$. Define the left inverse M by

taking the basis vector $L(\vec{v}_i)$ to $\vec{v}_i$ and all basis vectors which are not of this form to $\vec{0}$.

Section 10.2

1. rank $= 3$

3. rank $= 3$

5. rank $= 5$

7. rank $= 3$ nullity $= 1$

9. rank $= 3$, nullity $= 0$

11. The nullity is the dimension of the kernel of L. If $r \neq 0$ then the set solutions to $rL(\vec{v}) = \vec{0}$ are exactly the same as the solutions to $L(\vec{v}) = \vec{0}$. Hence the kernels of rL and L are the same, so their dimensions must be as well.

13. If $M(\vec{v}) = \vec{0}$, then it is also true that $L \circ M(\vec{v}) = \vec{0}$ so $\text{Ker}(M) \subseteq \text{Ker}(L \circ M)$. Thus $\dim(\text{Ker}(M)) \leq \dim(\text{Ker}(L \circ M))$.

15. The following examples will do:

    (a) $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & -2 \\ 3 & -4 \end{bmatrix}$ gives $\mathbf{A} + \mathbf{B} = \begin{bmatrix} 2 & 0 \\ 6 & 0 \end{bmatrix}$ which has rank 1.

    (b) $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 0 & 2 \\ 0 & 4 \end{bmatrix}$ gives $\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

Section 10.3

1. $\begin{bmatrix} 2 & 8 \\ -4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ -4 & 19 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 2 & 4 & 12 \\ 1 & 6 & -14 \\ 3 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 4 & 0 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 6 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{bmatrix}$

Chapter 11
Section 11.1

1. $\begin{bmatrix} \frac{5}{13} & \frac{1}{13} \\ -\frac{3}{13} & \frac{2}{13} \end{bmatrix}$

3. $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \mathcal{V} = \mathbb{Z}_2^3$

5. $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -1 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$

7. $\begin{bmatrix} \frac{9}{2} & -\frac{7}{2} & \frac{9}{2} \\ -2 & 5 & -5 \\ \frac{1}{2} & -\frac{1}{2} & \frac{3}{2} \end{bmatrix}$

9. $\begin{bmatrix} 0 & 3 & 5 & 10 \\ -1 & -5 & -7 & -10 \\ 2 & 6 & 9 & 13 \\ 0 & -1 & -1 & -3 \end{bmatrix}$

11. $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 & 1 \\ 0 & 2 & -1 & 6 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 6 & 0 \\ 0 & -1 & 6 & 1 \\ 2 & 1 & -3 & 1 \end{bmatrix}$

13. $\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 5 & -6 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{5}{2} & 4 & \frac{5}{2} & \frac{7}{2} \\ \frac{13}{2} & 2 & \frac{7}{2} & \frac{9}{2} \\ \frac{17}{2} & 3 & \frac{11}{2} & \frac{11}{2} \end{bmatrix}$

15. $\begin{bmatrix} -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \\ \frac{3}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{5}{4} & \frac{1}{4} & -\frac{3}{4} \end{bmatrix} \begin{bmatrix} 1 & 3 & -2 & 1 \\ 2 & 2 & 5 & -6 \\ 0 & 2 & -1 & 6 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{bmatrix} = \begin{bmatrix} -1 & \frac{19}{4} & 1 & -\frac{23}{4} \\ 1 & \frac{31}{4} & 1 & -\frac{27}{4} \\ -2 & \frac{21}{2} & 11 & -\frac{25}{2} \\ 5 & -\frac{111}{4} & -11 & \frac{111}{4} \end{bmatrix}$

Section 11.2

1. $\begin{bmatrix} 1 & 4 \\ 3 & 6 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 3 & -6 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7 & -24 \\ 3 & -6 \end{bmatrix}$

3. $\begin{bmatrix} 3 & 12 & 12 \\ -1 & 0 & 0 \\ 1 & 6 & 9 \end{bmatrix} \sim \begin{bmatrix} 1 & 4 & 4 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ -1 & 4 & 0 \\ 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 24 & 12 \\ 0 & 6 & 3 \\ 1 & 2 & 3 \end{bmatrix}$

5. Basis $= \{[1,0,0],[1,1,1],[1,2,3]\}$ Matrix $= \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -3 \\ 0 & 1 & 3 \end{bmatrix}$

7. Basis $= \{[1,0,0],[1,1,1],[1,2,3]\}$ Matrix $= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$

9. Basis $= \{[1,0,0,0],[0,1,1,0],[0,0,0,1],[0,0,1,0]\}$ Matrix $= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

11. Basis $= \{[1,0,0,0,0],[0,0,1,0,0],[0,0,0,1,1],[0,1,0,0,0],[0,0,0,0,2]\}$

    Matrix $= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

Chapter 12
Section 12.1

1. -4

3. 24

5. 40

7. -18

9. $-a$

11. $abcd$

12. A matrix $\mathbf{A}$ is invertible if and only if its row reduced echelon form is the identity, which has determinant 1. Since the operations which reduce a matrix to row reduced echelon form do not change a determinant from 0 to non-zero, this tells us that if $\det(\mathbf{A}) \neq 0$ then $\mathbf{A}$ is invertible.

Section 12.2 1-11 as in section 12.1

Chapter 13

Section 13.1

1. (a) $\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 12 \\ 8 \end{bmatrix} = 4\begin{bmatrix} 3 \\ 2 \end{bmatrix}$

2. (a) $[1, 0, 0]$

   (c) $[-3, 1, 0]$

3. (a) $(1 - x)(5 - x)(9 - x))$

   (c) $(1 - x)((5 - x)(9 - x) - 48)$

4. (a)

Eigenvalue	eigenvector
1	[1,0,0]
5	[1,2,0]
9	[3,6,4]

   (c)

Eigenvalue	eigenvector
1	[1,0,0]
$7 - 2\sqrt{13}$	$[0, -1 - \sqrt{13}, 4]$
$7 + 2\sqrt{13}$	$[0, -1 + \sqrt{13}, 4]$

6. If λ is an eigenvalue for $\mathbf{M}$ with eigenvector $\vec{v}$ then λ^n is an eigenvalue for $\mathbf{M}^n$ with the same eigenvector.

8. If 0 is an eigenvalue of L then L is not invertable because the eigenvector for 0 is a nonzero member of the kernel of L, so L is not one to one. If L is not invertable then the rank nullity theorem tells us that the kernel must be nontrivial, giving an eigenvector for 0.

Section 13.2

1. $2 - i, 2 + i, 4$

3. $1, 2 + i, 2 - i, 3$

5. $3, 2 \pm \sqrt{7}$

7. -1

9. $0, 1$

11. $\begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix} \sim \begin{bmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ Using the basis $([-1,0,1],[-2,1,1],[2,3,1])$
so that $\begin{bmatrix} -2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{4}{3} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} -1 & -2 & 2 \\ 0 & 1 & 3 \\ 1 & 1 & 1 \end{bmatrix}$

13. The only eigenvalue of this matrix is 1. The vector $[1,-2,1]$ is an eigenvector. If we extend to a basis by taking $([1,-2,1],[0,1,0],[0,0,1])$ we get the matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -3 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

If we now restrict our attention to the lower right corner we get

$$\begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

This tells us that

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -3 \\ 0 & 1 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Section 13.3

1. (a)

n	Approximate λ	Approximate eigenvector
1	3	$[1, -\frac{1}{3}]$
3	-2.77778	$[-0.76, 1]$
20	-3.82843	$[-0.414212, 1]$

(c)

n	Approximate λ	Approximate eigenvector
1	5	$[0.6, 0, 1]$
3	4.3333	$[.582418, -.131868, 1]$
20	4.2924	$[.547514, -.197375, 1]$

Chapter 14
Section 14.1

1. All that remains to be proved in detail are the axioms $< f|g >=< g|f >$ and $< kf|g >= k < f|g >$ and $< f|g+h >=< f|g > + < f|h >$.

$$< f|g >= \int_{-1}^{1} f(x)g(x)dx = \int_{-1}^{1} g(x)f(x)dx$$

by commutativity in $\mathbf{R}$.

$$< kf|g >= \int_{-1}^{1} kf(x)g(x)dx = k\int_{-1}^{1} f(x)g(x)dx = k < f|g >$$

by the linearity of the integration process.

$$\begin{aligned} < f|g+h > &= \int_{-1}^{1} f(x)(g(x)+h(x))dx \\ &= \int_{-1}^{1} f(x)g(x) + f(x)h(x)dx \\ &= \int_{-1}^{1} f(x)g(x)dx + \int_{-1}^{1} f(x)h(x)dx \\ &= < f|g > + < f|h > . \end{aligned}$$

3. (a)

$$\begin{aligned} \langle\vec{a}|\vec{b}+\vec{c}\rangle &= [1,2,4]\cdot([3,7,1]+[2,0,-1]) \\ &= [1,2,4]\cdot[5,7,0] = 19 \\ \langle\vec{a}|\vec{b}\rangle + \langle\vec{a}|\vec{c}\rangle &= [1,2,4]\cdot[3,7,1]+[1,2,4]\cdot[2,0,-1] \\ &= 21-2 = 19 \end{aligned}$$

   (b) $\|\vec{a}\|\ \|\vec{b}\| = \sqrt{1239} \approx 35.19943181$ and $\langle\vec{a}|\vec{b}\rangle^2 = 21$

   (c) $k = \frac{2}{5}$

5. $< x+c|1 >= \int_{-1}^{1} x + cdx = \frac{x^2}{2} + cx \Big|_{-1}^{1} = 2c = 0$ when $c = 0$.

7. (a) $\displaystyle\int_{-1}^{1} x^3 - 2x^2 + x - 2\ dx = -4\frac{2}{3}$

   (b) $\displaystyle\int_{-1}^{1} \frac{x^3 - 2x^2 + x - 2}{\sqrt{1-x^2}}\ dx = -3\pi$

(c) 1

Section 14.2

1.

$$\begin{aligned}
\vec{b}_1 &= [1,1] \\
\vec{b}_1* &= \frac{[1,1]}{\|[1,1\|} \\
&= [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] \\
\vec{b}_2 &= [0,1] - ([0,1] \cdot [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}])[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}] \\
&= [0,1] - [\frac{1}{2}, \frac{1}{2} \\
&= [-\frac{1}{2}, \frac{1}{2}] \\
\vec{b}_2* &= [-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}].
\end{aligned}$$

3.

$$\vec{b}_1 = [1,1]/\sqrt{3} = [\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}] \tag{B.1}$$

$$\vec{b}_2 = \frac{[0,1] - \frac{\langle [1,1] | [0,1] \rangle}{3}[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}]}{\|[0,1] - \frac{\langle [1,1] | [0,1] \rangle}{3}[\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}]\|} = [-\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}] \tag{B.2}$$

5. $\{[1,0,1],[2,1,-2],[1,-4,1]\}$ will do

7.

$$\begin{aligned}
p_1 &= 1 \\
p_2 &= x - \frac{\int_0^\infty x\ e^{-x} dx}{\int_0^\infty e^{-x} dx} 1 \\
&= x - \frac{1!}{0!} 1 \\
&= x - 1
\end{aligned}$$

$$\begin{aligned} p_3 &= x^2 - \frac{\int_0^\infty x^2\ e^{-x}dx}{\int_0^\infty e^{-x}dx}1 - \frac{\int_0^\infty (x^3-x)e^{-x}dx}{\int_0^\infty (x^2-2x+1)e^{-x}dx}(x-1) \\ &= x^2 - \frac{2!}{0!}1 - \frac{3!-1!}{2!-2\ 1!+0!}(x-1) \\ &= x^2-2-5(x-1) \\ &= x^2-5x+3 \end{aligned}$$

9. $1, x-1, x^2-4x+2$

Section 14.3

1. $\frac{1}{\sqrt{2}}[1/\sqrt{2}, 0, -1/\sqrt{2}] + \frac{1}{\sqrt{3}}[1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}] = [\frac{5}{6}, \frac{1}{3}, \frac{-1}{6}]$

3. $[\frac{11}{9}, \frac{16}{9}, \frac{7}{9}]$

5. $3(\sin(1)-\cos(1))x$

Chapter 15
Section 15.1

1. (a) $\{[1,0,0]\}^\perp = \{[0,h,k] | h,k \in \mathbb{R}\}$
   (c) $\{[1,1,0],[0,1,1]\}^\perp = \{[h,-h,h] | h \in \mathbb{R}\}$
   (e) $(\text{Span}\{[1,1,0],[0,1,1]\})^\perp = \{[h,-h,h] | h \in \mathbb{R}\}$

3. By definition if $\vec{x} \in A^\perp$ and $\vec{a} \in A$ then $< \vec{a}|\vec{x} >= 0$, but $< \vec{a}|\vec{x} >=< \vec{x}|\vec{a} >$, so $\vec{a} \in (A^\perp)^\perp$. Thus $A \subseteq A^{\perp\perp}$

5. If $\mathcal{A} < \mathcal{B}$ then if $< \vec{v}|\vec{x} >= 0$ for all $\vec{x} \in \mathcal{B}$, it is certainly true that $< \vec{v}|\vec{x} >= 0$ for all $\vec{x} \in \mathcal{A}$, since $\vec{x} \in \mathcal{A} \Rightarrow \vec{x} \in \mathcal{B}$. This shows $\mathcal{B} < \mathcal{A}$.

Section 15.2

1. (a) Orthogonal
   (c) Not orthogonal

3. (a) yes
   (c) yes

Index