MATH 1410
ELEMENTARY LINEAR ALGEBRA

Fall 2016 Edition, University of Lethbridge

Sean Fitzpatrick

Department of Mathematics and Computer Science

University of Lethbridge

Contributing Textbooks

Precalculus

Version $\lfloor \pi \rfloor = 3$

Carl Stitz and Jeff Zeager

www.stitz-zeager.com

Fundamentals of Matrix Algebra

Third Edition, Version 3.1110

Gregory Hartman

www.vmi.edu

APEX Calculus

Version 3.0

Gregory Hartman et al

apexcalculus.com

This version of the text assembled and edited by Sean Fitzpatrick, University of Lethbridge, July, 2016.

Contents

Preface

Math 1410 preface goes here...

The book is very much a work in progress, and I will be editing it regularly. Feedback is always welcome.

Acknowledgements

First and foremost, I need to thank the authors of the textbooks that provide the source material for this text. Without their hard work, and willingness to make their books (and the source code) freely available, it would not have been possible to create an affordable textbook for this course. You can find the original textbooks at their websites:

www.stitz-zeager.com, for the *Precalculus* textbook, by Stitz and Zeager,

http://www.vmi.edu/academics/departments/applied-mathematics/affordable-textbooks-apex/, for the *Fundamentals of Matrix Algebra* textbook, by Gregory Hartman, and

apexcalculus.com, for the APEX *Calculus* textbook, by Hartman et al.

Sean Fitzpatrick
Department of Mathematics and Computer Science
University of Lethbridge
July, 2016

1: The Real Numbers

This first chapter is intended as a resource for those students needing a quick review of some of the essential high school mathematics for this course. We begin with some basic set theory terminology that may pop up from time to time, followed by a reminder on the rules for arithmetic with real numbers, and a tour of the Cartesian coordinate plane. Students who are already comfortable with these topics can feel free to jump ahead to Chapter 2.

1.1 Some Basic Set Theory Notions

Definition 1 Set

A **set** is a well-defined collection of objects which are called the 'elements' of the set. Here, 'well-defined' means that it is possible to determine if something belongs to the collection or not, without prejudice.

For example, the collection of letters that make up the word "pronghorns" is well-defined and is a set, but the collection of the worst math teachers in the world is **not** well-defined, and so is **not** a set. In general, there are three ways to describe sets. They are

One thing that student evaluations teach us is that any given Mathematics instructor can be simultaneously the best and worst teacher ever, depending on who is completing the evaluation.

Key Idea 1 Ways to Describe Sets

1. **The Verbal Method:** Use a sentence to define a set.
2. **The Roster Method:** Begin with a left brace '$\{$', list each element of the set *only once* and then end with a right brace '$\}$'.
3. **The Set-Builder Method:** A combination of the verbal and roster methods using a "dummy variable" such as x.

For example, let S be the set described *verbally* as the set of letters that make up the word "pronghorns". A **roster** description of S would be $\{p, r, o, n, g, h, s\}$. Note that we listed 'r', 'o', and 'n' only once, even though they appear twice in "pronghorns." Also, the *order* of the elements doesn't matter, so $\{o, n, p, r, g, s, h\}$ is also a roster description of S. A **set-builder** description of S is:

$$\{x \,|\, x \text{ is a letter in the word "pronghorns".}\}$$

The way to read this is: 'The set of elements x <u>such that</u> x is a letter in the word "pronghorns."' We define to sets to be equal if they have exactly the same elements, and denote this using the familiar equals sign '$=$'. Thus, we may write $S = \{p, r, o, n, g, h, s\}$ or $S = \{x \,|\, x \text{ is a letter in the word "pronghorns".}\}$. Clearly r is an element of S and q is not an element S. We express these sentiments mathematically by writing $r \in S$ and $q \notin S$.

More precisely, we have the following.

Definition 2 Notation for set inclusion

Let A be a set.

- If x is an element of A then we write $x \in A$ which is read 'x is in A'.
- If x is *not* an element of A then we write $x \notin A$ which is read 'x is not in A'.

The notation $x \in A$ can be read as "x is in A", or "x is an element of A", or "x belongs to A", with similar readings for $x \notin A$.

Now let's consider the set

$$C = \{x \mid x \text{ is a consonant in the word "pronghorns"}\}.$$

A roster description of C is $C = \{p, r, n, g, h, s\}$. Note that by construction, every element of C is also in S. We express this relationship by stating that the set C is a **subset** of the set S, which is written in symbols as $C \subseteq S$. The more formal definition is given below.

Definition 3 Subset

Given sets A and B, we say that the set A is a **subset** of the set B and write '$A \subseteq B$' if every element in A is also an element of B.

Note that in our example above $C \subseteq S$, but not vice-versa, since $o \in S$ but $o \notin C$. Additionally, the set of vowels $V = \{a, e, i, o, u\}$, while it does have an element in common with S, is not a subset of S. (As an added note, S is not a subset of V, either.) We could, however, *build* a set which contains both S and V as subsets by gathering all of the elements in both S and V together into a single set, say $U = \{p, r, o, n, g, h, s, a, e, i, u\}$. Then $S \subseteq U$ and $V \subseteq U$. The set U we have built is called the **union** of the sets S and V and is denoted $S \cup V$. Furthermore, S and V aren't completely *different* sets since they both contain the letter 'o.' (Since the word 'different' could be ambiguous, mathematicians use the word *disjoint* to refer to two sets that have no elements in common.) The **intersection** of two sets is the set of elements (if any) the two sets have in common. In this case, the intersection of S and V is $\{o\}$, written $S \cap V = \{o\}$. We formalize these ideas below.

Definition 4 Intersection and Union

Suppose A and B are sets.

- The **intersection** of A and B is $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$
- The **union** of A and B is $A \cup B = \{x \mid x \in A \text{ or } x \in B \text{ (or both)}\}$

The key words in Definition 4 to focus on are the conjunctions: 'intersection' corresponds to 'and' meaning the elements have to be in *both* sets to be in the intersection, whereas 'union' corresponds to 'or' meaning the elements have to be in one set, or the other set (or both). In other words, to belong to the union of two sets an element must belong to *at least one* of them.

Returning to the sets C and V above, $C \cup V = \{p, r, n, g, h, s, a, e, i, o, u\}$. When it comes to their intersection, however, we run into a bit of notational awkwardness since C and V have no elements in common. While we could write $C \cap V = \{\}$, this sort of thing happens often enough that we give the set with no elements a name.

The full extent of the empty set's role will not be explored in this text, but it is of fundamental importance in Set Theory. In fact, the empty set can be used to generate numbers - mathematicians can create something from nothing! If you're interested, read about the von Neumann construction of the natural numbers or consider signing up for Math 2000.

Definition 5 **Empty set**

The **Empty Set** $\emptyset$ is the set which contains no elements. That is,

$$\emptyset = \{\} = \{x \,|\, x \neq x\}.$$

As promised, the empty set is the set containing no elements since no matter what 'x' is, '$x = x$.' Like the number '0,' the empty set plays a vital role in mathematics. We introduce it here more as a symbol of convenience as opposed to a contrivance. Using this new bit of notation, we have for the sets C and V above that $C \cap V = \emptyset$. A nice way to visualize relationships between sets and set operations is to draw a **Venn Diagram**. A Venn Diagram for the sets S, C and V is drawn in Figure 1.1.

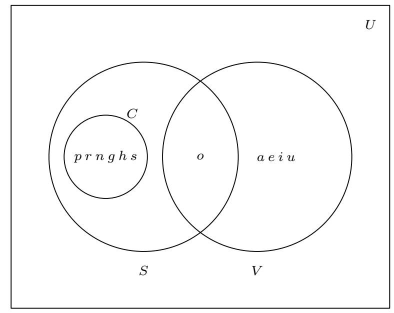
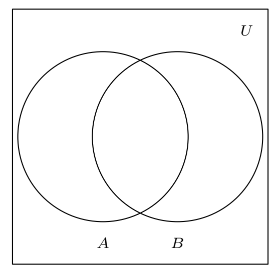


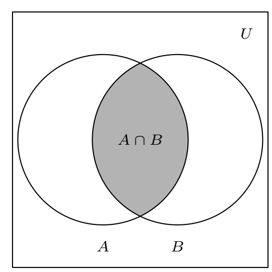
Figure 1.1: A Venn diagram for C, S, and V

In Figure 1.1 we have three circles - one for each of the sets C, S and V. We visualize the area enclosed by each of these circles as the elements of each set. Here, we've spelled out the elements for definitiveness. Notice that the circle representing the set C is completely inside the circle representing S. This is a geometric way of showing that $C \subseteq S$. Also, notice that the circles representing S and V overlap on the letter 'o'. This common region is how we visualize $S \cap V$. Notice that since $C \cap V = \emptyset$, the circles which represent C and V have no overlap whatsoever.

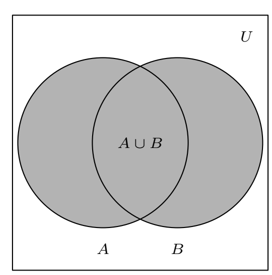
All of these circles lie in a rectangle labelled U (for 'universal' set). A universal set contains all of the elements under discussion, so it could always be taken as the union of all of the sets in question, or an even larger set. In this case, we could take $U = S \cup V$ or U as the set of letters in the entire alphabet. The usual triptych of Venn Diagrams indicating generic sets A and B along with $A \cap B$ and $A \cup B$ is given in Figure 1.2.



Sets A and B.



$A \cap B$ is shaded.



$A \cup B$ is shaded.

Figure 1.2: Venn diagrams for intersection and union

(The reader may well wonder if there is an ultimate universal set which contains *everything*. The short answer is 'no'. Our definition of a set turns out to be overly simplistic, but correcting this takes us well beyond the confines of this course. If you want the longer answer, you can begin by reading about Russell's Paradox on Wikipedia.)

In the next section, we will review the algebraic properties of the real number system. Other properties of the real numbers (such as the order property that allows us to picture the set of real numbers as a "number line") are essential to Calculus, but not that important for Linear Algebra, so we will leave it to other courses (such as Math 1010 or Math 1560) to handle the discussion of these topics.

1.2 Real Number Arithmetic

In this section we list the properties of real number arithmetic. We will focus on those aspects that are most needed for Math 1410 (things like the algebraic axioms, and working with fractions), and gloss over those that are more calculus related (exponents, roots, etc.). Students wanting more detail than what is provided here might want to consult the resources available on the *Math Basics* page on Moodle. In particular, since this is an *algebra* textbook, we will assume that the reader has encountered the real number system before, and omit a definition, focusing instead on the algebraic properties of real numbers. We begin with the axioms for addition of real numbers.

The set of real numbers is denoted usually denoted $\mathbb{R}$. A careful definition of $\mathbb{R}$ is actually quite complicated, and even most calculus classes choose not to include it. If you want to *really* understand what the real numbers are, you'll need to take a course in Real Analysis, such as Math 3500.

Definition 6 Properties of Real Number Addition

- **Closure:** For all real numbers a and b, $a+b$ is also a real number.
- **Commutativity:** For all real numbers a and b, $a+b=b+a$.
- **Associativity:** For all real numbers a, b and c, $a+(b+c)=(a+b)+c$.
- **Identity:** There is a real number '0' so that for all real numbers a, $a+0=a$.
- **Inverse:** For all real numbers a, there is a real number $-a$ such that $a+(-a)=0$.
- **Definition of Subtraction:** For all real numbers a and b, $a-b=a+(-b)$.

Next, we give real number multiplication a similar treatment. Recall that we may denote the product of two real numbers a and b a variety of ways: ab, $a\cdot b$, $a(b)$, $(a)(b)$ and so on. We'll refrain from using $a\times b$ for real number multiplication in this text.

While it is true that $ab=ba$ for any pair of real numbers a and b, there are plenty of algebraic systems where this property does not hold. In particular, in Chapter 3 we'll learn how to multiply *matrices* A and B, and see that in most cases, the matrix product AB is **not** equal to the product BA. (We'll also see that such products are not even guaranteed to be *defined*!) Students are so used to the commutative property for multiplication of real numbers that it can be difficult to make the adjustment to the non-commutative nature matrix multiplication, and this is the source of many common errors in a course like Math 1410.

Definition 7 Properties of Real Number Multiplication

- **Closure:** For all real numbers a and b, ab is also a real number.
- **Commutativity:** For all real numbers a and b, $ab=ba$.
- **Associativity:** For all real numbers a, b and c, $a(bc)=(ab)c$.
- **Identity:** There is a real number '1' so that for all real numbers a, $a\cdot 1=a$.
- **Inverse:** For all real numbers $a\neq 0$, there is a real number $\frac{1}{a}$ such that $a\left(\frac{1}{a}\right)=1$.
- **Definition of Division:** For all real numbers a and $b\neq 0$, $a\div b=\frac{a}{b}=a\left(\frac{1}{b}\right)$.

While most students (and some faculty) tend to skip over these properties or give them a cursory glance at best, it is important to realize that the properties stated above are what drive the symbolic manipulation for all of Algebra. When listing a tally of more than two numbers, $1 + 2 + 3$ for example, we don't need to specify the order in which those numbers are added. Notice though, try as we might, we can add only two numbers at a time and it is the associative property of addition which assures us that we could organize this sum as $(1+2)+3$ or $1+(2+3)$. This brings up a note about 'grouping symbols'. Recall that parentheses and brackets are used in order to specify which operations are to be performed first. In the absence of such grouping symbols, multiplication (and hence division) is given priority over addition (and hence subtraction). For example, $1 + 2 \cdot 3 = 1 + 6 = 7$, but $(1 + 2) \cdot 3 = 3 \cdot 3 = 9$. As you may recall, we can 'distribute' the 3 across the addition if we really wanted to do the multiplication first: $(1 + 2) \cdot 3 = 1 \cdot 3 + 2 \cdot 3 = 3 + 6 = 9$. More generally, we have the following.

Definition 8 The Distributive Property and Factoring

For all real numbers a, b and c:

- **Distributive Property:** $a(b+c) = ab+ac$ and $(a+b)c = ac+bc$.
- **Factoring:** $ab + ac = a(b + c)$ and $ac + bc = (a + b)c$.

Warning: A common source of errors for beginning students is the misuse (that is, lack of use) of parentheses. When in doubt, more is better than less: redundant parentheses add clutter, but do not change meaning, whereas writing $2x + 1$ when you meant to write $2(x + 1)$ is almost guaranteed to cause you to make a mistake. (Even if you're able to proceed correctly in spite of your lack of proper notation, this is the sort of thing that will get you on your grader's bad side, so it's probably best to avoid the problem in the first place.)

It is worth pointing out that we didn't really need to list the Distributive Property both for $a(b+c)$ (distributing from the left) and $(a+b)c$ (distributing from the right), since the commutative property of multiplication gives us one from the other. Also, 'factoring' really is the same equation as the distributive property, just read from right to left. These are the first of many redundancies in this section, and they exist in this review section for one reason only - in our experience, many students *see* these things differently so we will list them as such.

It is hard to overstate the importance of the Distributive Property. For example, in the expression $5(2 + x)$, without knowing the value of x, we cannot perform the addition inside the parentheses first; we must rely on the distributive property here to get $5(2 + x) = 5 \cdot 2 + 5 \cdot x = 10 + 5x$. The Distributive Property is also responsible for combining 'like terms'. Why is $3x + 2x = 5x$? Because $3x + 2x = (3 + 2)x = 5x$.

We continue our review with summaries of other properties of arithmetic, each of which can be derived from the properties listed above. First up are properties of the additive identity 0.

Theorem 1 Properties of Zero

Suppose a and b are real numbers.

- **Zero Product Property:** $ab = 0$ if and only if $a = 0$ or $b = 0$ (or both)

  Note: This not only says that $0 \cdot a = 0$ for any real number a, it also says that the *only* way to get an answer of '0' when multiplying two real numbers is to have one (or both) of the numbers be '0' in the first place.

- **Zeros in Fractions:** If $a \neq 0$, $\dfrac{0}{a} = 0 \cdot \left(\dfrac{1}{a}\right) = 0$.

  Note: The quantity $\dfrac{a}{0}$ is undefined.

The Zero Product Property drives most of the equation solving algorithms in Algebra because it allows us to take complicated equations and reduce them to simpler ones. For example, you may recall that one way to solve $x^2 + x - 6 = 0$ is by factoring the left hand side of this equation to get $(x-2)(x+3) = 0$. From here, we apply the Zero Product Property and set each factor equal to zero. This yields $x - 2 = 0$ or $x + 3 = 0$ so $x = 2$ or $x = -3$. This type of calculation is key to finding the eigenvalues of a matrix, as we'll see in Section 8.1.

We now continue with a review of arithmetic with fractions.

Key Idea 2 Properties of Fractions

Suppose a, b, c and d are real numbers. Assume them to be nonzero whenever necessary; for example, when they appear in a denominator.

- **Identity Properties:** $a = \dfrac{a}{1}$ and $\dfrac{a}{a} = 1$.
- **Fraction Equality:** $\dfrac{a}{b} = \dfrac{c}{d}$ if and only if $ad = bc$.
- **Multiplication of Fractions:** $\dfrac{a}{b} \cdot \dfrac{c}{d} = \dfrac{ac}{bd}$. In particular: $\dfrac{a}{b} \cdot c = \dfrac{a}{b} \cdot \dfrac{c}{1} = \dfrac{ac}{b}$
- **Division of Fractions:** $\dfrac{a}{b} \Big/ \dfrac{c}{d} = \dfrac{a}{b} \cdot \dfrac{d}{c} = \dfrac{ad}{bc}$.

  In particular: $1 \Big/ \dfrac{a}{b} = \dfrac{b}{a}$ and $\dfrac{a}{b} \Big/ c = \dfrac{a}{b} \Big/ \dfrac{c}{1} = \dfrac{a}{b} \cdot \dfrac{1}{c} = \dfrac{a}{bc}$
- **Addition and Subtraction of Fractions:** $\dfrac{a}{b} \pm \dfrac{c}{b} = \dfrac{a \pm c}{b}$.
- **Equivalent Fractions:** $\dfrac{a}{b} = \dfrac{ad}{bd}$, since $\dfrac{a}{b} = \dfrac{a}{b} \cdot 1 = \dfrac{a}{b} \cdot \dfrac{d}{d} = \dfrac{ad}{bd}$
- **'Reducing' Fractions:** $\dfrac{a\cancel{d}}{b\cancel{d}} = \dfrac{a}{b}$, since $\dfrac{ad}{bd} = \dfrac{a}{b} \cdot \dfrac{d}{d} = \dfrac{a}{b} \cdot 1 = \dfrac{a}{b}$.

  In particular, $\dfrac{ab}{b} = a$ since $\dfrac{ab}{b} = \dfrac{ab}{1 \cdot b} = \dfrac{a\cancel{b}}{1 \cdot \cancel{b}} = \dfrac{a}{1} = a$ and $\dfrac{b-a}{a-b} = \dfrac{(-1)\cancel{(a-b)}}{\cancel{(a-b)}} = -1$.

It's always worth remembering that **division is the same as multiplication by the reciprocal**. You'd be surprised how often this comes in handy.

Note: A common denominator is **not** required to **multiply** or **divide** fractions!

Note: A common denominator **is** required to **add or subtract** fractions!

Note: The *only* way to change the denominator is to multiply both it and the numerator by the same nonzero value because we are, in essence, multiplying the fraction by 1.

We reduce fractions by 'cancelling' common factors - this is really just reading the previous property 'from right to left'.
Caution: We may only cancel common **factors** from both numerator and denominator: we can cancel the twos in $\dfrac{3(2)}{5(2)}$, but **not** in $\dfrac{3+2}{5+2}$.

Next up is a review of the arithmetic of 'negatives'. On page 4 we first introduced the dash which we all recognize as the 'negative' symbol in terms of the additive inverse. For example, the number -3 (read 'negative 3') is defined so that $3 + (-3) = 0$. We then defined subtraction using the concept of the additive inverse again so that, for example, $5 - 3 = 5 + (-3)$.

It might be junior high (elementary?) school material, but arithmetic with fractions is one of the most common sources of errors among university students. If you're not comfortable working with fractions, we strongly recommend seeing your instructor (or a tutor) to go over this material until you're completely confident that you understand it. Experience (and even formal educational studies) suggest that your success handling fractions corresponds pretty well with your overall success in passing your Mathematics courses.

Key Idea 3 Properties of Negatives

Given real numbers a and b we have the following.

- **Additive Inverse Properties:** $-a = (-1)a$ and $-(-a) = a$
- **Products of Negatives:** $(-a)(-b) = ab$.
- **Negatives and Products:** $-ab = -(ab) = (-a)b = a(-b)$.
- **Negatives and Fractions:** If b is nonzero, $-\dfrac{a}{b} = \dfrac{-a}{b} = \dfrac{a}{-b}$ and $\dfrac{-a}{-b} = \dfrac{a}{b}$.
- **'Distributing' Negatives:** $-(a + b) = -a - b$ and $-(a - b) = -a + b = b - a$.
- **'Factoring' Negatives:** $-a - b = -(a + b)$ and $b - a = -(a - b)$.

In this text we do not distinguish typographically between the dashes in the expressions '$5 - 3$' and '-3' even though they are mathematically quite different. In the expression '$5 - 3$,' the dash is a *binary* operation (that is, an operation requiring *two* numbers) whereas in '-3', the dash is a *unary* operation (that is, an operation requiring only one number). You might ask, 'Who cares?' Your calculator does - that's who! In the text we can write $-3 - 3 = -6$ but that will not work in your calculator. Instead you'd need to type $^{-}3 - 3$ to get -6 where the first dash comes from the '$+/-$' key.

An important point here is that when we 'distribute' negatives, we do so across addition or subtraction only. This is because we are really distributing a factor of -1 across each of these terms: $-(a+b) = (-1)(a+b) = (-1)(a) + (-1)(b) = (-a) + (-b) = -a - b$. Negatives do not 'distribute' across multiplication: $-(2\cdot 3) \neq (-2)\cdot(-3)$. Instead, $-(2\cdot 3) = (-2)\cdot(3) = (2)\cdot(-3) = -6$. The same sort of thing goes for fractions: $-\frac{3}{5}$ can be written as $\frac{-3}{5}$ or $\frac{3}{-5}$, but not $\frac{-3}{-5}$. It's about time we did a few examples to see how these properties work in practice.

Example 1 Arithmetic with fractions

Perform the indicated operations and simplify. By 'simplify' here, we mean to have the final answer written in the form $\frac{a}{b}$ where a and b are integers which have no common factors. Said another way, we want $\frac{a}{b}$ in 'lowest terms'.

1. $\dfrac{1}{4} + \dfrac{6}{7}$
2. $\dfrac{5}{12} - \left(\dfrac{47}{30} - \dfrac{7}{3}\right)$
3. $\dfrac{\dfrac{12}{5} - \dfrac{7}{24}}{1 + \left(\dfrac{12}{5}\right)\left(\dfrac{7}{24}\right)}$
4. $\dfrac{(2(2)+1)(-3-(-3)) - 5(4-7)}{4 - 2(3)}$
5. $\left(\dfrac{3}{5}\right)\left(\dfrac{5}{13}\right) - \left(\dfrac{4}{5}\right)\left(-\dfrac{12}{13}\right)$

Solution

1. It may seem silly to start with an example this basic but experience has taught us not to take much for granted. We start by finding the lowest common denominator and then we rewrite the fractions using that new denominator. Since 4 and 7 are **relatively prime**, meaning they have no

factors in common, the lowest common denominator is $4 \cdot 7 = 28$.

$$\begin{aligned}
\frac{1}{4} + \frac{6}{7} &= \frac{1}{4} \cdot \frac{7}{7} + \frac{6}{7} \cdot \frac{4}{4} && \text{Equivalent Fractions} \\
&= \frac{7}{28} + \frac{24}{28} && \text{Multiplication of Fractions} \\
&= \frac{31}{28} && \text{Addition of Fractions}
\end{aligned}$$

The result is in lowest terms because 31 and 28 are relatively prime so we're done.

2. We could begin with the subtraction in parentheses, namely $\frac{47}{30} - \frac{7}{3}$, and then subtract that result from $\frac{5}{12}$. It's easier, however, to first distribute the negative across the quantity in parentheses and then use the Associative Property to perform all of the addition and subtraction in one step. The lowest common denominator for all three fractions is 60.

We could have used $12 \cdot 30 \cdot 3 = 1080$ as our common denominator but then the numerators would become unnecessarily large. It's best to use the *lowest* common denominator.

$$\begin{aligned}
\frac{5}{12} - \left(\frac{47}{30} - \frac{7}{3}\right) &= \frac{5}{12} - \frac{47}{30} + \frac{7}{3} && \text{Distribute the Negative} \\
&= \frac{5}{12} \cdot \frac{5}{5} - \frac{47}{30} \cdot \frac{2}{2} + \frac{7}{3} \cdot \frac{20}{20} && \text{Equivalent Fractions} \\
&= \frac{25}{60} - \frac{94}{60} + \frac{140}{60} && \text{Multiplication of Fractions} \\
&= \frac{71}{60} && \text{Addition and Subtraction of Fractions}
\end{aligned}$$

The numerator and denominator are relatively prime so the fraction is in lowest terms and we have our final answer.

3. What we are asked to simplify in this problem is known as a 'complex' or 'compound' fraction. Simply put, we have fractions within a fraction. The longest division line (also called a 'vinculum') performs the same sort of grouping function as parentheses:

$$\frac{\dfrac{12}{5} - \dfrac{7}{24}}{1 + \left(\dfrac{12}{5}\right)\left(\dfrac{7}{24}\right)} = \frac{\left(\dfrac{12}{5} - \dfrac{7}{24}\right)}{\left(1 + \left(\dfrac{12}{5}\right)\left(\dfrac{7}{24}\right)\right)}.$$

The first step to simplifying a compound fraction like this one is to see if you can simplify the little fractions inside it. There are two ways to proceed. One is to simplify the numerator and denominator separately, and then use the fact that division is the same thing as multiplication by the reciprocal, as follows:

$$\frac{\left(\frac{12}{5}-\frac{7}{24}\right)}{\left(1+\left(\frac{12}{5}\right)\left(\frac{7}{24}\right)\right)} = \frac{\left(\frac{12}{5}\cdot\frac{24}{24}-\frac{7}{24}\cdot\frac{5}{5}\right)}{\left(1\cdot\frac{120}{120}+\left(\frac{12}{5}\right)\left(\frac{7}{24}\right)\right)} \quad \text{Equivalent Fractions}$$

$$= \frac{288/120-35/120}{120/120+84/120} \quad \text{Multiplication of fractions}$$

$$= \frac{253/120}{204/120} \quad \text{Addition and subtraction of fractions}$$

$$= \frac{253}{\cancel{120}}\cdot\frac{\cancel{120}}{204} \quad \text{Division of fractions and cancellation}$$

$$= \frac{253}{204}$$

Since $253 = 11 \cdot 23$ and $204 = 2 \cdot 2 \cdot 3 \cdot 17$ have no common factors our result is in lowest terms which means we are done.

While there is nothing wrong with the above approach, we can also use our Equivalent Fractions property to rid ourselves of the 'compound' nature of this fraction straight away. The idea is to multiply both the numerator and denominator by the lowest common denominator of each of the 'smaller' fractions - in this case, $24 \cdot 5 = 120$.

$$\frac{\left(\frac{12}{5}-\frac{7}{24}\right)}{\left(1+\left(\frac{12}{5}\right)\left(\frac{7}{24}\right)\right)} = \frac{\left(\frac{12}{5}-\frac{7}{24}\right)\cdot 120}{\left(1+\left(\frac{12}{5}\right)\left(\frac{7}{24}\right)\right)\cdot 120} \quad \text{Equivalent Fractions}$$

$$= \frac{\left(\frac{12}{5}\right)(120)-\left(\frac{7}{24}\right)(120)}{(1)(120)+\left(\frac{12}{5}\right)\left(\frac{7}{24}\right)(120)}$$

Distributive Property

$$= \frac{\frac{12\cdot 120}{5}-\frac{7\cdot 120}{24}}{120+\frac{12\cdot 7\cdot 120}{5\cdot 24}} \quad \text{Multiply fractions}$$

$$= \frac{\frac{12\cdot 24\cdot\cancel{5}}{\cancel{5}}-\frac{7\cdot 5\cdot\cancel{24}}{\cancel{24}}}{120+\frac{12\cdot 7\cdot\cancel{5}\cdot\cancel{24}}{\cancel{5}\cdot\cancel{24}}} \quad \text{Factor and cancel}$$

$$= \frac{(12\cdot 24)-(7\cdot 5)}{120+(12\cdot 7)}$$

$$= \frac{288-35}{120+84} = \frac{253}{204},$$

which is the same as we obtained above.

4. This fraction may look simpler than the one before it, but the negative signs and parentheses mean that we shouldn't get complacent. Again we note that the division line here acts as a grouping symbol. That is,

$$\frac{(2(2)+1)(-3-(-3))-5(4-7)}{4-2(3)} = \frac{((2(2)+1)(-3-(-3))-5(4-7))}{(4-2(3))}$$

This means that we should simplify the numerator and denominator first, then perform the division last. We tend to what's in parentheses first, giving multiplication priority over addition and subtraction.

$$\begin{aligned}\frac{(2(2)+1)(-3-(-3))-5(4-7)}{4-2(3)} &= \frac{(4+1)(-3+3)-5(-3)}{4-6}\\ &= \frac{(5)(0)+15}{-2} = \frac{15}{-2}\\ &= -\frac{15}{2} \quad \text{Properties of Negatives}\end{aligned}$$

Since $15 = 3 \cdot 5$ and 2 have no common factors, we are done.

5. In this problem, we have multiplication and subtraction. Multiplication takes precedence so we perform it first. Recall that to multiply fractions, we do *not* need to obtain common denominators; rather, we multiply the corresponding numerators together along with the corresponding denominators. However, when we perform the subtraction, we *do* need a common denominator, so we will resist the temptation to cancel the fives in the first term straight away.

$$\begin{aligned}\left(\frac{3}{5}\right)\left(\frac{5}{13}\right)-\left(\frac{4}{5}\right)\left(-\frac{12}{13}\right) &= \frac{3\cdot 5}{5\cdot 13}-\frac{4\cdot(-12)}{5\cdot 13} && \text{Multiply fractions}\\ &= \frac{15}{65}-\frac{-48}{65}\\ &= \frac{15}{65}+\frac{48}{65} && \text{Properties of Negatives}\\ &= \frac{15+48}{65} && \text{Add numerators}\\ &= \frac{63}{65}\end{aligned}$$

Since $64 = 3 \cdot 3 \cdot 7$ and $65 = 5 \cdot 13$ have no common factors, our answer $\frac{63}{65}$ is in lowest terms and we are done.

Of the issues discussed in the previous set of examples none causes students more trouble than simplifying compound fractions. We presented two different methods for simplifying them: one in which we simplified the overall numerator and denominator and then performed the division and one in which we removed the compound nature of the fraction at the very beginning. We encourage the reader to go back and use both methods on each of the compound fractions presented. Keep in mind that when a compound fraction is encountered in the rest of the text it will usually be simplified using only one method and we may not choose your favourite method. Feel free to use the other one in your notes.

Exercises 1.2

Problems

In Exercises 1 – 20, perform the indicated operations and simplify.

1. $5 - 2 + 3$

2. $5 - (2 + 3)$

3. $\dfrac{2}{3} - \dfrac{4}{7}$

4. $\dfrac{3}{8} + \dfrac{5}{12}$

5. $\dfrac{5 - 3}{-2 - 4}$

6. $\dfrac{2(-3)}{3 - (-3)}$

7. $\dfrac{2(3) - (4 - 1)}{2^2 + 1}$

8. $\dfrac{4 - 5.8}{2 - 2.1}$

9. $\dfrac{1 - 2(-3)}{5(-3) + 7}$

10. $\dfrac{5(3) - 7}{2(3)^2 - 3(3) - 9}$

11. $\dfrac{2((-1)^2 - 1)}{((-1)^2 + 1)^2}$

12. $\dfrac{(-2)^2 - (-2) - 6}{(-2)^2 - 4}$

13. $\dfrac{3 - \frac{4}{9}}{-2 - (-3)}$

14. $\dfrac{\frac{2}{3} - \frac{4}{5}}{4 - \frac{7}{10}}$

15. $\dfrac{2\left(\frac{4}{3}\right)}{1 - \left(\frac{4}{3}\right)^2}$

16. $\dfrac{1 - \left(\frac{5}{3}\right)\left(\frac{3}{5}\right)}{1 + \left(\frac{5}{3}\right)\left(\frac{3}{5}\right)}$

17. $\left(\dfrac{2}{3}\right)^{-5}$

18. $3^{-1} - 4^{-2}$

19. $\dfrac{1 + 2^{-3}}{3 - 4^{-1}}$

20. $\dfrac{3 \cdot 5^{100}}{12 \cdot 5^{98}}$

1.3 The Cartesian Coordinate Plane

As a warm-up for the discussions of vectors and three-dimensional geometry yet to come, we will make a quick review of the **Cartesian Coordinate Plane**. Imagine two real number lines crossing at a right angle at 0 as shown in Figure 1.3.

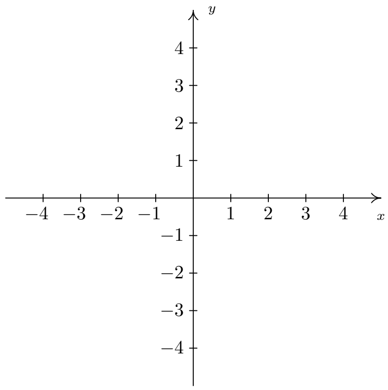


Figure 1.3: The Cartesian coordinate plane

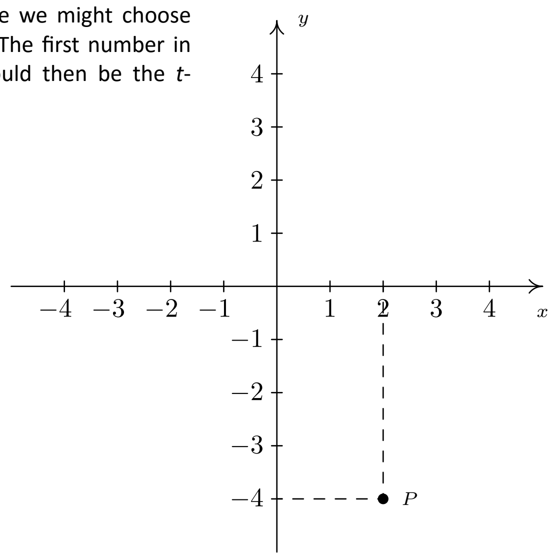
Usually extending off towards infinity is indicated by arrows, but here, the arrows are used to indicate the *direction* of increasing values of *x* and *y*.

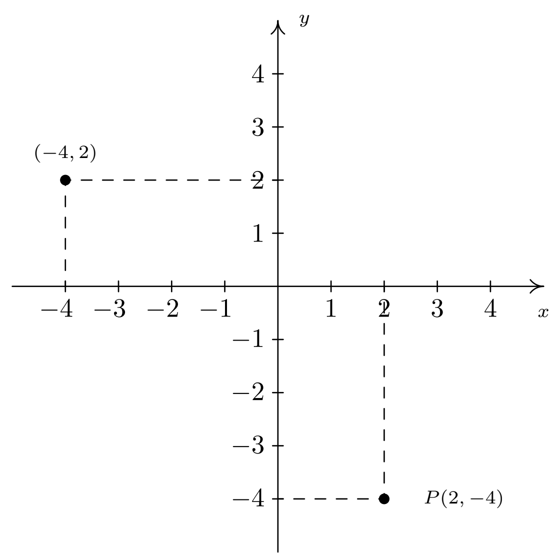
The horizontal number line is usually called the ***x*-axis** while the vertical number line is usually called the ***y*-axis**. As with the usual number line, we imagine these axes extending off indefinitely in both directions. Having two number lines allows us to locate the positions of points off of the number lines as well as points on the lines themselves.

The Cartesian Plane is named in honour of René Descartes.

For example, consider the point *P* on the next page. To use the numbers on the axes to label this point, we imagine dropping a vertical line from the *x*-axis to *P* and extending a horizontal line from the *y*-axis to *P*. This process is sometimes called 'projecting' the point *P* to the *x*- (respectively *y*-) axis. We then describe the point *P* using the **ordered pair** $(2, -4)$. The first number in the ordered pair is called the **abscissa** or ***x*-coordinate** and the second is called the **ordinate** or ***y*-coordinate**. Taken together, the ordered pair $(2, -4)$ comprise the **Cartesian coordinates** of the point *P*. In practice, the distinction between a point and its coordinates is blurred; for example, we often speak of 'the point $(2, -4)$.' We can think of $(2, -4)$ as instructions on how to reach *P* from the **origin** $(0, 0)$ by moving 2 units to the right and 4 units downwards. Notice that the order in the ordered pair is important $-$ if we wish to plot the point $(-4, 2)$, we would move to the left 4 units from the origin and then move upwards 2 units, as below on the right.

The names of the coordinates can vary depending on the context of the application. If, for example, the horizontal axis represented time we might choose to call it the *t*-axis. The first number in the ordered pair would then be the *t*-coordinate.





When we speak of the Cartesian Coordinate Plane, we mean the set of all possible ordered pairs (x, y) as *x* and *y* take values from the real numbers. Below is a summary of important facts about Cartesian coordinates.

Key Idea 4 Important Facts about the Cartesian Coordinate Plane

- (a, b) and (c, d) represent the same point in the plane if and only if $a = c$ and $b = d$.
- (x, y) lies on the x-axis if and only if $y = 0$.
- (x, y) lies on the y-axis if and only if $x = 0$.
- The origin is the point $(0, 0)$. It is the only point common to both axes.

The letter O is almost always reserved for the origin.

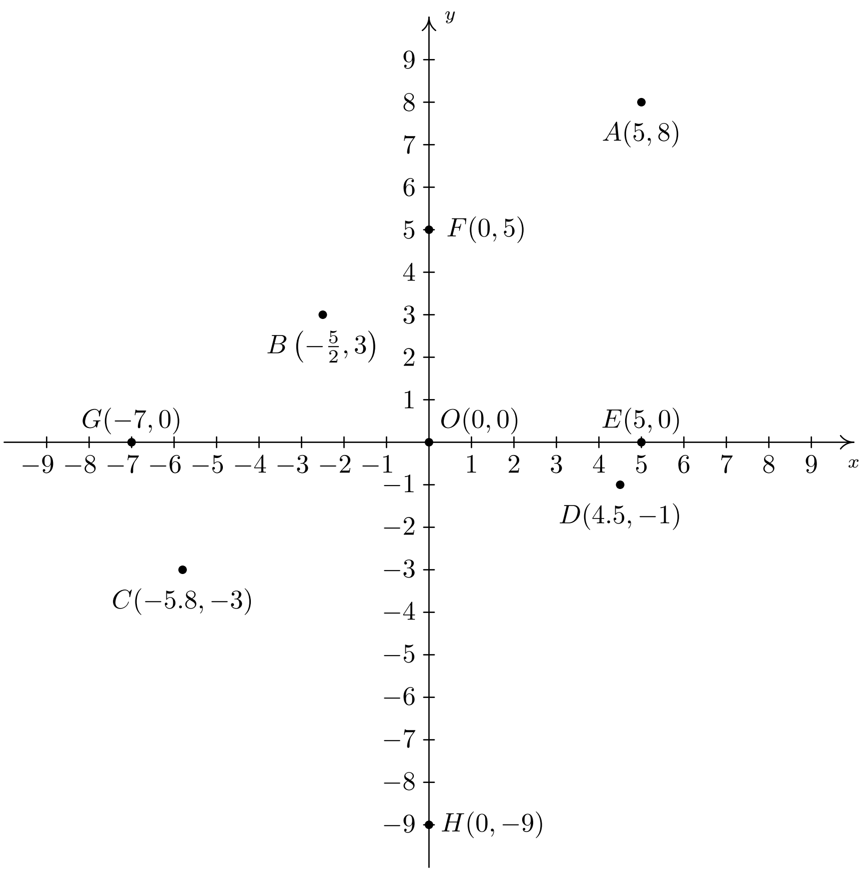
Example 2 Plotting points in the Cartesian Plane

Plot the following points: $A(5, 8)$, $B\left(-\frac{5}{2}, 3\right)$, $C(-5.8, -3)$, $D(4.5, -1)$, $E(5, 0)$, $F(0, 5)$, $G(-7, 0)$, $H(0, -9)$, $O(0, 0)$.

Cartesian coordinates are sometimes referred to as *rectangular coordinates*, to distinguish them from other coordinate systems such as *polar coordinates*: see Section 7.2.

Solution To plot these points, we start at the origin and move to the right if the x-coordinate is positive; to the left if it is negative. Next, we move up if the y-coordinate is positive or down if it is negative. If the x-coordinate is 0, we start at the origin and move along the y-axis only. If the y-coordinate is 0 we move along the x-axis only.

We will also see in Chapter 7 that the best way to visualize the set of *complex numbers* is by identifying complex numbers with points in the Cartesian plane.



The axes divide the plane into four regions called **quadrants**. They are labelled with Roman numerals and proceed counterclockwise around the plane: see Figure 1.4.

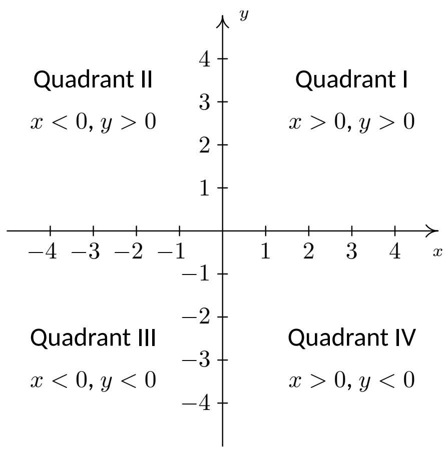


Figure 1.4: The four quadrants of the Cartesian plane

For example, $(1, 2)$ lies in Quadrant I, $(-1, 2)$ in Quadrant II, $(-1, -2)$ in Quadrant III and $(1, -2)$ in Quadrant IV. If a point other than the origin happens to lie on the axes, we typically refer to that point as lying on the positive or negative *x*-axis (if $y = 0$) or on the positive or negative *y*-axis (if $x = 0$). For example, $(0, 4)$ lies on the positive *y*-axis whereas $(-117, 0)$ lies on the negative *x*-axis. Such points do not belong to any of the four quadrants.

1.3.1 Distance in the Plane

Another important concept in Geometry is the notion of length. If we are going to unite Algebra and Geometry using the Cartesian Plane, then we need to develop an algebraic understanding of what distance in the plane means. Suppose we have two points, $P\,(x_0, y_0)$ and $Q\,(x_1, y_1)$, in the plane. By the **distance** d between P and Q, we mean the length of the line segment joining P with Q. (Remember, given any two distinct points in the plane, there is a unique line containing both points.) Our goal now is to create an algebraic formula to compute the distance between these two points. Consider the generic situation in Figure 1.5(a).

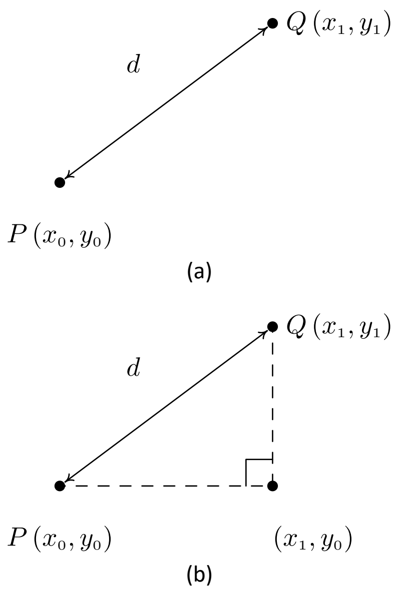


Figure 1.5: Distance between P and Q

With a little more imagination, we can envision a right triangle whose hypotenuse has length d as drawn in Figure 1.5(b). From the latter figure, we see that the lengths of the legs of the triangle are $|x_1 - x_0|$ and $|y_1 - y_0|$ so the Pythagorean Theorem gives us

$$
\begin{aligned}
|x_1 - x_0|^2 + |y_1 - y_0|^2 &= d^2 \\
(x_1 - x_0)^2 + (y_1 - y_0)^2 &= d^2
\end{aligned}
$$

Recall that the **absolute value** of a real number a, denoted by $|a|$, measures the distance of a from the origin. Since we never want a negative distance, this means $|a| = a$ if $a \geq 0$, while if $a < 0$, $|a| = -a$. For example, $|4| = 4$, and $|-7| = -(-7) = 7$. Given two real numbers a and b, $|a - b|$ measures the distance between them.

(Since the square of a number is always positive, we can drop the absolute value signs.) By extracting the square root of both sides of the second equation and using the fact that distance is never negative, we get

Key Idea 5 The Distance Formula

The distance d between the points $P\,(x_0, y_0)$ and $Q\,(x_1, y_1)$ is:

$$d = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$$

It is not always the case that the points P and Q lend themselves to constructing such a triangle. If the points P and Q are arranged vertically or horizontally, or describe the exact same point, we cannot use the above geometric argument to derive the distance formula. It is left to the reader in Exercise 15 to verify Equation 5 for these cases.

Example 3 Distance between two points

Find and simplify the distance between $P(-2,3)$ and $Q(1,-3)$.

SOLUTION

$$\begin{aligned} d &= \sqrt{(x_1-x_0)^2+(y_1-y_0)^2} \\ &= \sqrt{(1-(-2))^2+(-3-3)^2} \\ &= \sqrt{9+36} \\ &= 3\sqrt{5} \end{aligned}$$

So the distance is $3\sqrt{5}$.

Example 4 Finding points at a given distance

Find all of the points with x-coordinate 1 which are 4 units from the point $(3,2)$.

SOLUTION We shall soon see that the points we wish to find are on the line $x=1$, but for now we'll just view them as points of the form $(1,y)$.

We require that the distance from $(3,2)$ to $(1,y)$ be 4. The Distance Formula, Equation 5, yields

$$\begin{aligned} d &= \sqrt{(x_1-x_0)^2+(y_1-y_0)^2} \\ 4 &= \sqrt{(1-3)^2+(y-2)^2} \\ 4 &= \sqrt{4+(y-2)^2} \\ 4^2 &= \left(\sqrt{4+(y-2)^2}\right)^2 && \text{squaring both sides} \\ 16 &= 4+(y-2)^2 \\ 12 &= (y-2)^2 \\ (y-2)^2 &= 12 \\ y-2 &= \pm\sqrt{12} && \text{extracting the square root} \\ y-2 &= \pm 2\sqrt{3} \\ y &= 2\pm 2\sqrt{3} \end{aligned}$$

We obtain two answers: $(1,2+2\sqrt{3})$ and $(1,2-2\sqrt{3})$. The reader is encouraged to think about why there are two answers.

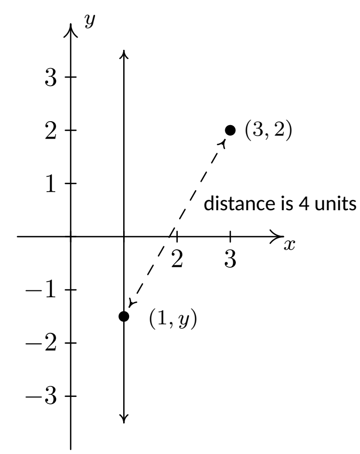


Figure 1.6: Diagram for Example 4

Related to finding the distance between two points is the problem of finding the **midpoint** of the line segment connecting two points. Given two points, $P(x_0,y_0)$ and $Q(x_1,y_1)$, the **midpoint** M of P and Q is defined to be the point on the line segment connecting P and Q whose distance from P is equal to its distance from Q.

If we think of reaching M by going 'halfway over' and 'halfway up' we get the following formula.

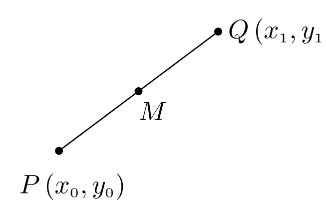


Figure 1.7: The midpoint of a line segment

Key Idea 6 The Midpoint Formula

The midpoint M of the line segment connecting $P(x_0,y_0)$ and $Q(x_1,y_1)$ is:

$$M = \left(\frac{x_0+x_1}{2}, \frac{y_0+y_1}{2}\right)$$

If we let d denote the distance between P and Q, we leave it as Exercise 16 to show that the distance between P and M is $d/2$ which is the same as the distance between M and Q. This suffices to show that Key Idea 6 gives the coordinates of the midpoint.

Example 5 **Finding the midpoint of a line segment**

Find the midpoint of the line segment connecting $P(-2, 3)$ and $Q(1, -3)$.

Solution

$$\begin{aligned} M &= \left(\frac{x_0 + x_1}{2}, \frac{y_0 + y_1}{2}\right) \\ &= \left(\frac{(-2) + 1}{2}, \frac{3 + (-3)}{2}\right) = \left(-\frac{1}{2}, \frac{0}{2}\right) \\ &= \left(-\frac{1}{2}, 0\right) \end{aligned}$$

The midpoint is $\left(-\frac{1}{2}, 0\right)$.

We close with a more abstract application of the Midpoint Formula.

Example 6 **An abstract midpoint problem**

If $a \neq b$, prove that the line $y = x$ equally divides the line segment with endpoints (a, b) and (b, a).

Solution To prove the claim, we use Equation 6 to find the midpoint

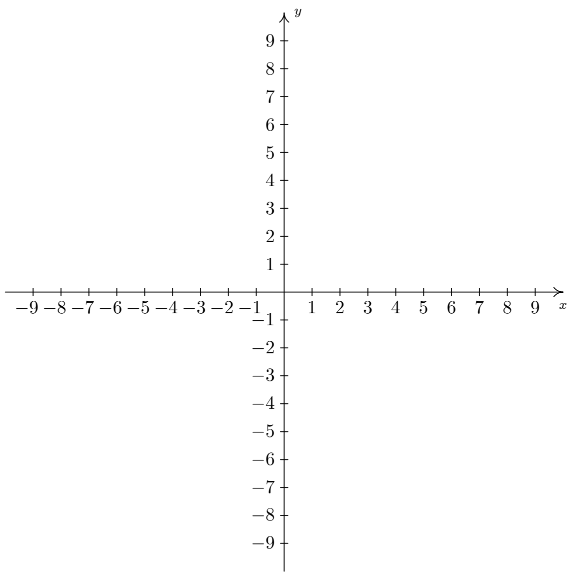
$$\begin{aligned} M &= \left(\frac{a + b}{2}, \frac{b + a}{2}\right) \\ &= \left(\frac{a + b}{2}, \frac{a + b}{2}\right) \end{aligned}$$

Since the x and y coordinates of this point are the same, we find that the midpoint lies on the line $y = x$, as required.

Exercises 1.3

Problems

1. Plot and label the points $A(-3,-7)$, $B(1.3,-2)$, $C(\pi,\sqrt{10})$, $D(0,8)$, $E(-5.5,0)$, $F(-8,4)$, $G(9.2,-7.8)$ and $H(7,5)$ in the Cartesian Coordinate Plane given below.



In Exercises 2 – 9, find the distance *d* between the points and the midpoint *M* of the line segment which connects them.

2. $(1,2)$, $(-3,5)$
3. $(3,-10)$, $(-1,2)$
4. $\left(\dfrac{1}{2},4\right)$, $\left(\dfrac{3}{2},-1\right)$
5. $\left(-\dfrac{2}{3},\dfrac{3}{2}\right)$, $\left(\dfrac{7}{3},2\right)$
6. $\left(\dfrac{24}{5},\dfrac{6}{5}\right)$, $\left(-\dfrac{11}{5},-\dfrac{19}{5}\right)$.
7. $\left(\sqrt{2},\sqrt{3}\right)$, $\left(-\sqrt{8},-\sqrt{12}\right)$
8. $\left(2\sqrt{45},\sqrt{12}\right)$, $\left(\sqrt{20},\sqrt{27}\right)$.
9. $(0,0)$, (x,y)
10. Find all of the points of the form $(x,-1)$ which are 4 units from the point $(3,2)$.
11. Find all of the points on the y-axis which are 5 units from the point $(-5,3)$.
12. Find all of the points on the x-axis which are 2 units from the point $(-1,1)$.
13. Find all of the points of the form $(x,-x)$ which are 1 unit from the origin.
14. Let's assume for a moment that we are standing at the origin and the positive y-axis points due North while the positive x-axis points due East. Our Sasquatch-o-meter tells us that Sasquatch is 3 miles West and 4 miles South of our current position. What are the coordinates of his position? How far away is he from us? If he runs 7 miles due East what would his new position be?
15. Verify the Distance Formula 5 for the cases when:
    (a) The points are arranged vertically. (Hint: Use $P(a,y_0)$ and $Q(a,y_1)$.)
    (b) The points are arranged horizontally. (Hint: Use $P(x_0,b)$ and $Q(x_1,b)$.)
    (c) The points are actually the same point. (You shouldn't need a hint for this one.)
16. Verify the Midpoint Formula by showing the distance between $P(x_1,y_1)$ and M and the distance between M and $Q(x_2,y_2)$ are both half of the distance between P and Q.
17. Show that the points A, B and C below are the vertices of a right triangle.
    (a) $A(-3,2)$, $B(-6,4)$, and $C(1,8)$
    (b) $A(-3,1)$, $B(4,0)$ and $C(0,-3)$
18. Find a point $D(x,y)$ such that the points $A(-3,1)$, $B(4,0)$, $C(0,-3)$ and D are the corners of a square. Justify your answer.
19. Discuss with your classmates how many numbers are in the interval $(0,1)$.
20. The world is not flat. (There are those who disagree with this statement. Look them up on the Internet some time when you're bored.) Thus the Cartesian Plane cannot possibly be the end of the story. Discuss with your classmates how you would extend Cartesian Coordinates to represent the three dimensional world. What would the Distance and Midpoint formulas look like, assuming those concepts make sense at all?

2: Vector Geometry

This chapter introduces a new mathematical object, the **vector**. Defined in Section 2.2, we will see that vectors provide a powerful language for describing quantities that have magnitude and direction aspects. A simple example of such a quantity is force: when applying a force, one is generally interested in how much force is applied (i.e., the magnitude of the force) and the direction in which the force was applied. Vectors will play an important role in many of the subsequent chapters in this text.

This chapter begins with moving our mathematics out of the plane and into "space." That is, we begin to think mathematically not only in two dimensions, but in three. With this foundation, we can explore vectors both in the plane and in space.

In this chapter we restrict ourselves to vectors in two and three dimensions so that we're able to understand things visually. However, we'll also see that the *algebraic* behaviour of vectors is the same in *any* dimension, including dimension four or greater. The only thing that changes is the number of coordinates involved. This is one of the great powers of mathematics: we are aided by our visual imagination, but not limited by it.

2.1 Introduction to Cartesian Coordinates in Space

We reviewed the two-dimensional Cartesian Plane in Section 1.3. This is the familiar background on which much of your high school mathematics played out, and it provides the setting for the Calculus of one variable encountered in courses like Math 1010 and Math 1560.

While there is wonderful mathematics to explore in "2D," we live in a "3D" world and eventually we will want to apply mathematics involving this third dimension. In this section we introduce Cartesian coordinates in space and explore basic surfaces. This will lay a foundation for much of what we do in the remainder of the text.

Each point P in space can be represented with an ordered triple, $P = (a, b, c)$, where a, b and c represent the relative position of P along the x-, y- and z-axes, respectively. Each axis is perpendicular to the other two.

Visualizing points in space on paper can be problematic, as we are trying to represent a 3-dimensional concept on a 2–dimensional medium. We cannot draw three lines representing the three axes in which each line is perpendicular to the other two. Despite this issue, standard conventions exist for plotting shapes in space that we will discuss that are more than adequate.

One convention is that the axes must conform to the **right hand rule**. This rule states that when the index finger of the right hand is extended in the direction of the positive x-axis, and the middle finger (bent "inward" so it is perpendicular to the palm) points along the positive y-axis, then the extended thumb will point in the direction of the positive z-axis. (It may take some thought to verify this, but this system is inherently different from the one created by using the "left hand rule.")

As long as the coordinate axes are positioned so that they follow this rule, it does not matter how the axes are drawn on paper. There are two popular methods that we briefly discuss.

In Figure 2.1 we see the point $P = (2, 1, 3)$ plotted on a set of axes. The basic convention here is that the x-y plane is drawn in its standard way, with the z-axis down to the left. The perspective is that the paper represents the x-y plane and the positive z axis is coming up, off the page. This method is preferred by many engineers. Because it can be hard to tell where a single point lies in relation to all the axes, dashed lines have been added to let one see how far along each axis the point lies.

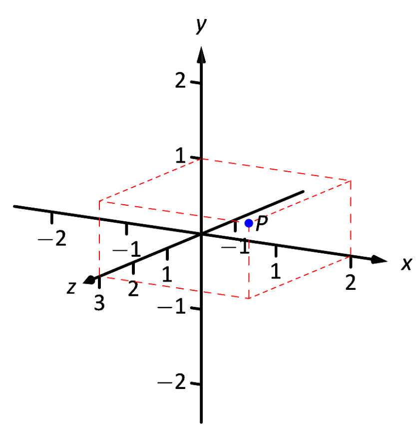


Figure 2.1: Plotting the point $P = (2, 1, 3)$ in space.

One can also consider the x-y plane as being a horizontal plane in, say, a

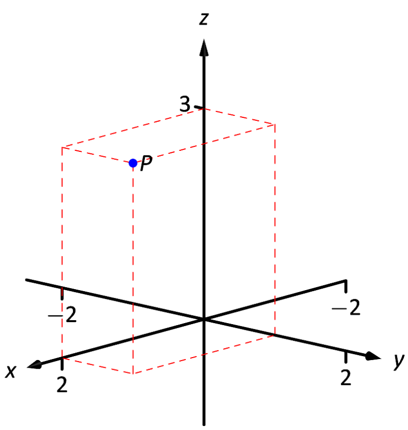


Figure 2.2: Plotting the point $P = (2, 1, 3)$ in space with a perspective used in this text.

room, where the positive z-axis is pointing up. When one steps back and looks at this room, one might draw the axes as shown in Figure 2.2. The same point P is drawn, again with dashed lines. This point of view is preferred by most mathematicians, and is the convention adopted by this text.

Measuring Distances

It is of critical importance to know how to measure distances between points in space. The formula for doing so is based on measuring distance in the plane (see Key Idea 5 on Page 14), and is known (in both contexts) as the Euclidean measure of distance.

Definition 9 Distance In Space

Let $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$ be points in space. The distance D between P and Q is

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

We refer to the line segment that connects points P and Q in space as $\overline{PQ}$, and refer to the length of this segment as $\left\|\overline{PQ}\right\|$. The above distance formula allows us to compute the length of this segment.

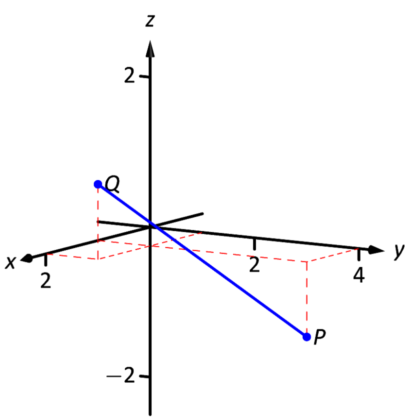


Figure 2.3: Plotting points P and Q in Example 7.

Example 7 Length of a line segment

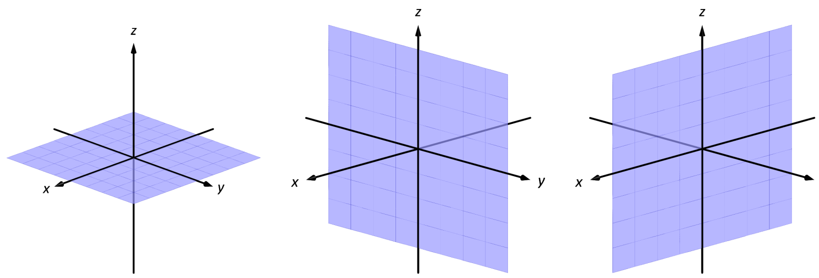
Let $P = (1, 4, -1)$ and let $Q = (2, 1, 1)$. Draw the line segment $\overline{PQ}$ and find its length.

Solution The points P and Q are plotted in Figure 2.3; no special consideration need be made to draw the line segment connecting these two points; simply connect them with a straight line. One *cannot* actually measure this line on the page and deduce anything meaningful; its true length must be measured analytically. Applying Definition 9, we have

$$\left\|\overline{PQ}\right\| = \sqrt{(2-1)^2 + (1-4)^2 + (1-(-1))^2} = \sqrt{14} \approx 3.74.$$

Introduction to Planes in Space

The coordinate axes naturally define three planes (shown in Figure 2.4), the **coordinate planes**: the x-y plane, the y-z plane and the x-z plane. The x-y plane is characterized as the set of all points in space where the z-value is 0. This, in fact, gives us an equation that describes this plane: $z = 0$. Likewise, the x-z plane is all points where the y-value is 0, characterized by $y = 0$.



the x-y plane   the y-z plane   the x-z plane

Figure 2.4: The coordinate planes.

The equation $x = 2$ describes all points in space where the x-value is 2. This is a plane, parallel to the y-z coordinate plane, shown in Figure 2.5.

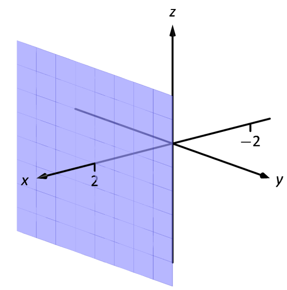


Figure 2.5: The plane $x = 2$.

Example 8 **Regions defined by planes**
Sketch the region defined by the inequalities $-1 \leq y \leq 2$.

Solution The region is all points between the planes $y = -1$ and $y = 2$. These planes are sketched in Figure 2.6, which are parallel to the x-z plane. Thus the region extends infinitely in the x and z directions, and is bounded by planes in the y direction.

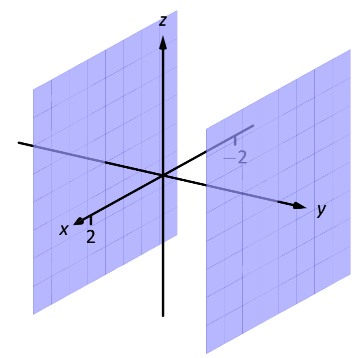


Figure 2.6: Sketching the boundaries of a region in Example 8.

This section has introduced points and distance in space and introduced equations of basic planes in space. We'll reconsider planes in more detail in Section 2.6, but first, we need to introduce the language of *vectors*.

Exercises 2.1

Terms and Concepts

1. Axes drawn in space must conform to the ________ ________ rule.

2. In the plane, the equation $x = 2$ defines a ________; in space, $x = 2$ defines a ________.

3. In the plane, the equation $y = x^2$ defines a ________; in space, $y = x^2$ defines a ________.

4. Which quadric surface looks like a Pringles® chip?

5. Consider the hyperbola $x^2 - y^2 = 1$ in the plane. If this hyperbola is rotated about the x-axis, what quadric surface is formed?

6. Consider the hyperbola $x^2 - y^2 = 1$ in the plane. If this hyperbola is rotated about the y-axis, what quadric surface is formed?

Problems

7. The points $A = (1, 4, 2)$, $B = (2, 6, 3)$ and $C = (4, 3, 1)$ form a triangle in space. Find the distances between each pair of points and determine if the triangle is a right triangle.

8. The points $A = (1, 1, 3)$, $B = (3, 2, 7)$, $C = (2, 0, 8)$ and $D = (0, -1, 4)$ form a quadrilateral $ABCD$ in space. Is this a parallelogram?

9. Find the center and radius of the sphere defined by $x^2 - 8x + y^2 + 2y + z^2 + 8 = 0$.

10. Find the center and radius of the sphere defined by $x^2 + y^2 + z^2 + 4x - 2y - 4z + 4 = 0$.

In Exercises 11 – 14, describe the region in space defined by the inequalities.

11. $x^2 + y^2 + z^2 < 1$

12. $0 \leq x \leq 3$

13. $x \geq 0,\ y \geq 0,\ z \geq 0$

14. $y \geq 3$

2.2 An Introduction to Vectors

Many quantities we think about daily can be described by a single number: temperature, speed, cost, weight and height. There are also many other concepts we encounter daily that cannot be described with just one number. For instance, a weather forecaster often describes wind with its speed and its direction ("... with winds from the south-east gusting up to 30 mph ..."). When applying a force, we are concerned with both the magnitude and direction of that force. In both of these examples, *direction* is important. Because of this, we study *vectors*, mathematical objects that convey both magnitude and direction information.

One "bare–bones" definition of a vector is based on what we wrote above: "a vector is a mathematical object with magnitude and direction parameters." This definition leaves much to be desired, as it gives no indication as to how such an object is to be used. Several other definitions exist; we choose here a definition rooted in a geometric visualization of vectors. (In later chapters we will instead focus on algebraic aspects of vectors.) It is very simplistic but readily permits further investigation.

Definition 10 Vector

A **vector** is a directed line segment.

Given points P and Q (either in the plane or in space), we denote with $\overrightarrow{PQ}$ the vector *from P to Q*. The point P is said to be the **initial point** of the vector, and the point Q is the **terminal point**.

The **magnitude**, **length** or **norm** of $\overrightarrow{PQ}$ is the length of the line segment $\overline{PQ}$: $\| \overrightarrow{PQ} \| = \| \overline{PQ} \|$.

Two vectors are **equal** if they have the same magnitude and direction.

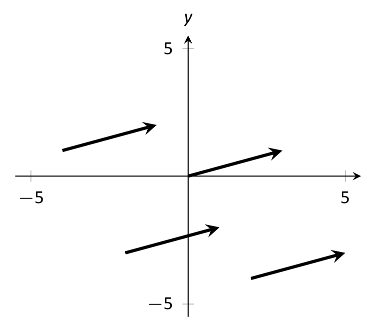


Figure 2.7: Drawing the same vector with different initial points.

Figure 2.7 shows multiple instances of the same vector. Each directed line segment has the same direction and length (magnitude), hence each is the same vector.

Following typical (but potentially confusing) mathematical convention, we use $\mathbb{R}^2$ (pronounced "r two") to represent all the vectors in the plane (as well as the plane itself), and use $\mathbb{R}^3$ (pronounced "r three") to represent all the vectors in space, as well as three-dimensional space itself.

Consider the vectors $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ as shown in Figure 2.8. The vectors look to be equal; that is, they seem to have the same length and direction. Indeed, they are. Both vectors move 2 units to the right and 1 unit up from the initial point to reach the terminal point. One can analyze this movement to measure the magnitude of the vector, and the movement itself gives direction information (one could also measure the slope of the line passing through P and Q or R and S). Since they have the same length and direction, these two vectors are equal.

This demonstrates that inherently all we care about is *displacement*; that is, how far in the x, y and possibly z directions the terminal point is from the initial point. Both the vectors $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ in Figure 2.8 have an x-displacement of 2 and a y-displacement of 1. This suggests a standard way of describing vectors in the plane. A vector whose x-displacement is a and whose y-displacement is b will have terminal point (a, b) when the initial point is the origin, $(0, 0)$. This leads

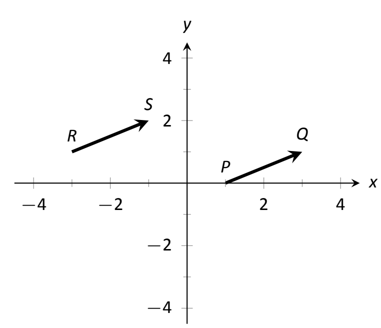


Figure 2.8: Illustrating how equal vectors have the same displacement.

The component form of a vector allows us to identify a point (a, b) (or (a, b, c)) with the corresponding vector $\langle a, b\rangle$ (or $\langle a, b, c\rangle$), so that vectors and points contain essentially the same information, presented in different contexts. This is why mathematicians don't mind using the notation $\mathbb{R}^n$ to refer to both a set of vectors and the set of points containing those vectors.

Caution: The notation $\langle a, b, c\rangle$ used in this chapter for a vector is common in geometry, physics, and calculus, but in later chapters we will use *column vector* notation $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ to represent the same vector. This notation is more natural in the context of matrix algebra.

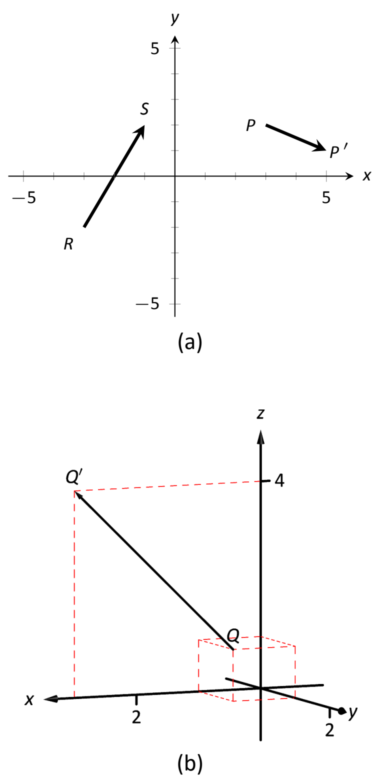


Figure 2.9: Graphing vectors in Example 9.

us to a definition of a standard and concise way of referring to vectors.

Definition 11 Component Form of a Vector

1. The **component form** of a vector $\vec{v}$ in $\mathbb{R}^2$, whose terminal point is (a, b) when its initial point is $(0, 0)$, is $\langle a, b\rangle$.
2. The **component form** of a vector $\vec{v}$ in $\mathbb{R}^3$, whose terminal point is (a, b, c) when its initial point is $(0, 0, 0)$, is $\langle a, b, c\rangle$.

The numbers a, b (and c, respectively) are the **components** of $\vec{v}$.

It follows from the definition that the component form of the vector $\overrightarrow{PQ}$, where $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ is

$$\overrightarrow{PQ} = \langle x_2 - x_1, y_2 - y_1\rangle ;$$

in space, where $P = (x_1, y_1, z_1)$ and $Q = (x_2, y_2, z_2)$, the component form of $\overrightarrow{PQ}$ is

$$\overrightarrow{PQ} = \langle x_2 - x_1, y_2 - y_1, z_2 - z_1\rangle .$$

We practice using this notation in the following example.

Example 9 Using component form notation for vectors

1. Sketch the vector $\vec{v} = \langle 2, -1\rangle$ starting at $P = (3, 2)$ and find its magnitude.
2. Find the component form of the vector $\vec{w}$ whose initial point is $R = (-3, -2)$ and whose terminal point is $S = (-1, 2)$.
3. Sketch the vector $\vec{u} = \langle 2, -1, 3\rangle$ starting at the point $Q = (1, 1, 1)$ and find its magnitude.

Solution

1. Using P as the initial point, we move 2 units in the positive x-direction and -1 units in the positive y-direction to arrive at the terminal point $P' = (5, 1)$, as drawn in Figure 2.9(a).

   The magnitude of $\vec{v}$ is determined directly from the component form:

   $$\| \vec{v} \| = \sqrt{2^2 + (-1)^2} = \sqrt{5}.$$

2. Using the note following Definition 11, we have

   $$\overrightarrow{RS} = \langle -1 - (-3), 2 - (-2)\rangle = \langle 2, 4\rangle .$$

   One can readily see from Figure 2.9(a) that the x- and y-displacement of $\overrightarrow{RS}$ is 2 and 4, respectively, as the component form suggests.

3. Using Q as the initial point, we move 2 units in the positive x-direction, -1 unit in the positive y-direction, and 3 units in the positive z-direction to arrive at the terminal point $Q' = (3, 0, 4)$, illustrated in Figure 2.9(b).

   The magnitude of $\vec{u}$ is:

   $$\| \vec{u} \| = \sqrt{2^2 + (-1)^2 + 3^2} = \sqrt{14}.$$

Now that we have defined vectors, and have created a nice notation by which to describe them, we start considering how vectors interact with each other. That is, we define an *algebra* on vectors.

Definition 12 Vector Algebra

1. Let $\vec{u} = \langle u_1, u_2 \rangle$ and $\vec{v} = \langle v_1, v_2 \rangle$ be vectors in $\mathbb{R}^2$, and let c be a scalar.
   (a) The addition, or sum, of the vectors $\vec{u}$ and $\vec{v}$ is the vector
   $$\vec{u} + \vec{v} = \langle u_1 + v_1, u_2 + v_2 \rangle .$$
   (b) The scalar product of c and $\vec{v}$ is the vector
   $$c\vec{v} = c \langle v_1, v_2 \rangle = \langle cv_1, cv_2 \rangle .$$
2. Let $\vec{u} = \langle u_1, u_2, u_3 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3 \rangle$ be vectors in $\mathbb{R}^3$, and let c be a scalar.
   (a) The addition, or sum, of the vectors $\vec{u}$ and $\vec{v}$ is the vector
   $$\vec{u} + \vec{v} = \langle u_1 + v_1, u_2 + v_2, u_3 + v_3 \rangle .$$
   (b) The scalar product of c and $\vec{v}$ is the vector
   $$c\vec{v} = c \langle v_1, v_2, v_3 \rangle = \langle cv_1, cv_2, cv_3 \rangle .$$

In short, we say addition and scalar multiplication are computed "component–wise."

Example 10 Adding vectors

Sketch the vectors $\vec{u} = \langle 1, 3 \rangle$, $\vec{v} = \langle 2, 1 \rangle$ and $\vec{u} + \vec{v}$ all with initial point at the origin.

Solution We first compute $\vec{u} + \vec{v}$.

$$\begin{aligned}\vec{u} + \vec{v} &= \langle 1, 3 \rangle + \langle 2, 1 \rangle \\ &= \langle 3, 4 \rangle .\end{aligned}$$

These are all sketched in Figure 2.10.

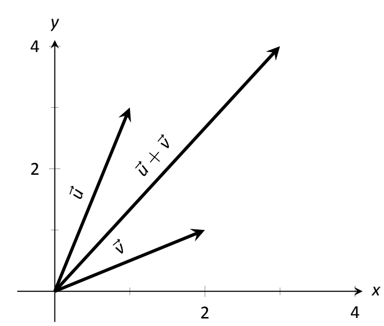


Figure 2.10: Graphing the sum of vectors in Example 10.

As vectors convey magnitude and direction information, the sum of vectors also convey length and magnitude information. Adding $\vec{u} + \vec{v}$ suggests the following idea:

"Starting at an initial point, go out $\vec{u}$, then go out $\vec{v}$."

This idea is sketched in Figure 2.11, where the initial point of $\vec{v}$ is the terminal point of $\vec{u}$. This is known as the "Head to Tail Rule" of adding vectors. Vector addition is very important. For instance, if the vectors $\vec{u}$ and $\vec{v}$ represent forces acting on a body, the sum $\vec{u} + \vec{v}$ gives the resulting force. Because of various physical applications of vector addition, the sum $\vec{u} + \vec{v}$ is often referred to as the **resultant vector**, or just the "resultant."

Analytically, it is easy to see that $\vec{u} + \vec{v} = \vec{v} + \vec{u}$. Figure 2.11 also gives a graphical representation of this, using gray vectors. Note that the vectors $\vec{u}$ and

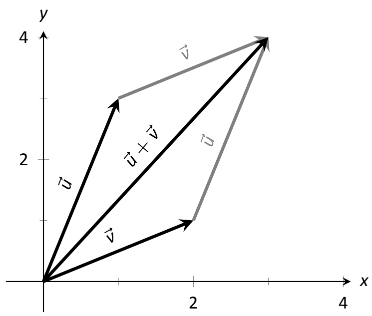


Figure 2.11: Illustrating how to add vectors using the Head to Tail Rule and Parallelogram Law.

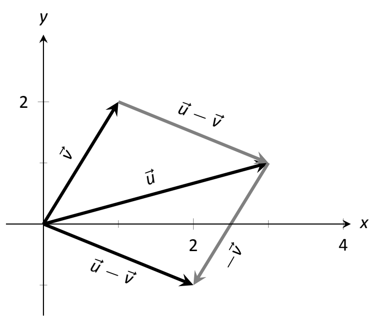


Figure 2.12: Illustrating how to subtract vectors graphically.

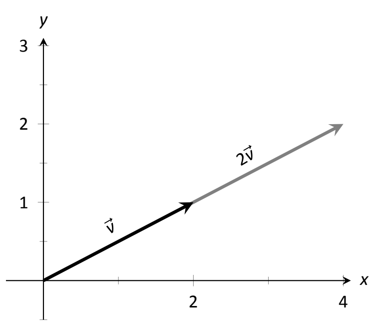


Figure 2.13: Graphing vectors $\vec{v}$ and $2\vec{v}$ in Example 12.

$\vec{v}$, when arranged as in the figure, form a parallelogram. Because of this, the Head to Tail Rule is also known as the Parallelogram Law: the vector $\vec{u}+\vec{v}$ is defined by forming the parallelogram defined by the vectors $\vec{u}$ and $\vec{v}$; the initial point of $\vec{u}+\vec{v}$ is the common initial point of parallelogram, and the terminal point of the sum is the common terminal point of the parallelogram.

While not illustrated here, the Head to Tail Rule and Parallelogram Law hold for vectors in $\mathbb{R}^3$ as well.

It follows from the properties of the real numbers and Definition 12 that

$$\vec{u}-\vec{v}=\vec{u}+(-1)\vec{v}.$$

The Parallelogram Law gives us a good way to visualize this subtraction. We demonstrate this in the following example.

Example 11 Vector Subtraction

Let $\vec{u}=\langle 3,1\rangle$ and $\vec{v}=\langle 1,2\rangle$. Compute and sketch $\vec{u}-\vec{v}$.

Solution The computation of $\vec{u}-\vec{v}$ is straightforward, and we show all steps below. Usually the formal step of multiplying by (-1) is omitted and we "just subtract."

$$\begin{aligned}\vec{u}-\vec{v}&=\vec{u}+(-1)\vec{v}\\&=\langle 3,1\rangle+\langle -1,-2\rangle\\&=\langle 2,-1\rangle.\end{aligned}$$

Figure 2.12 illustrates, using the Head to Tail Rule, how the subtraction can be viewed as the sum $\vec{u}+(-\vec{v})$. The figure also illustrates how $\vec{u}-\vec{v}$ can be obtained by looking only at the terminal points of $\vec{u}$ and $\vec{v}$ (when their initial points are the same).

Example 12 Scaling vectors

1. Sketch the vectors $\vec{v}=\langle 2,1\rangle$ and $2\vec{v}$ with initial point at the origin.
2. Compute the magnitudes of $\vec{v}$ and $2\vec{v}$.

Solution

1. We compute $2\vec{v}$:

$$\begin{aligned}2\vec{v}&=2\langle 2,1\rangle\\&=\langle 4,2\rangle.\end{aligned}$$

Both $\vec{v}$ and $2\vec{v}$ are sketched in Figure 2.13. Make note that $2\vec{v}$ does not start at the terminal point of $\vec{v}$; rather, its initial point is also the origin.

2. The figure suggests that $2\vec{v}$ is twice as long as $\vec{v}$. We compute their magnitudes to confirm this.

$$\begin{aligned}\|\,\vec{v}\,\|&=\sqrt{2^2+1^2}\\&=\sqrt{5}.\\\|\,2\vec{v}\,\|&=\sqrt{4^2+2^2}\\&=\sqrt{20}\\&=\sqrt{4\cdot 5}=2\sqrt{5}.\end{aligned}$$

As we suspected, $2\vec{v}$ is twice as long as $\vec{v}$.

In Example 12 above, we saw that $\|2\vec{v}\| = 2\|\vec{v}\|$, which makes sense geometrically: $2\vec{v} = \vec{v} + \vec{v}$, and adding a vector to itself should produce a vector twice as long with the same direction. The following theorem tells us that this is true in general.

Theorem 2 Magnitude and scalar multiplication

For any vector $\vec{v}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, and any real number c, we have

$$\|c\vec{v}\| = |c|\,\|\vec{v}\|\,.$$

To prove Theorem 2, let $\vec{v} = \langle a, b\rangle$ be any vector in $\mathbb{R}^2$ (the proof for $\mathbb{R}^3$ is similar), and let c be any scalar. Then

$$\begin{aligned}\|c\vec{v}\| &= \|c\langle a, b\rangle\| \\ &= \|\langle ca, cb\rangle\| \\ &= \sqrt{(ca)^2 + (cb)^2} \\ &= \sqrt{c^2a^2 + c^2b^2} \\ &= \sqrt{c^2(a^2 + b^2)} \\ &= \sqrt{c^2}\sqrt{a^2 + b^2} \\ &= |c|\,\|\vec{v}\|\,,\end{aligned}$$

as required.
(Recall that $\sqrt{c^2} = |c|$, the absolute value of c, since c might be negative, but the square root is always positive.)

In particular, Theorem 2 tells us that if $c > 0$, then $\|c\vec{v}\| = c\|\vec{v}\|$, so that $c\vec{v}$ is a vector in the *same* direction as $\vec{v}$ whose magnitude has been stretched (if $c > 1$) or shrunk (if $c < 1$) by a factor of c relative to that of $\vec{v}$. On the other hand, if $c < 0$, then we have $\|c\vec{v}\| = -c\|\vec{v}\|$, so that $c\vec{v}$ points in the *opposite* direction to that of $\vec{v}$.

The **zero vector** is the vector whose initial point is also its terminal point. It is denoted by $\vec{0}$. Its component form, in $\mathbb{R}^2$, is $\langle 0, 0\rangle$; in $\mathbb{R}^3$, it is $\langle 0, 0, 0\rangle$. Usually the context makes is clear whether $\vec{0}$ is referring to a vector in the plane or in space.

Our examples have illustrated key principles in vector algebra: how to add and subtract vectors and how to multiply vectors by a scalar. The following theorem states formally the properties of these operations.

Theorem 3 Properties of Vector Operations

The following are true for all scalars c and d, and for all vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$, where $\vec{u}$, $\vec{v}$ and $\vec{w}$ are all in $\mathbb{R}^2$ or where $\vec{u}$, $\vec{v}$ and $\vec{w}$ are all in $\mathbb{R}^3$:

1. $\vec{u} + \vec{v} = \vec{v} + \vec{u}$ Commutative Property
2. $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$ Associative Property
3. $\vec{v} + \vec{0} = \vec{v}$ Additive Identity
4. $(cd)\vec{v} = c(d\vec{v})$
5. $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$ Distributive Property
6. $(c + d)\vec{v} = c\vec{v} + d\vec{v}$ Distributive Property
7. $0\vec{v} = \vec{0}$
8. $\|\, c\vec{v}\,\| = |c| \cdot \|\,\vec{v}\,\|$
9. $\|\,\vec{u}\,\| = 0$ if, and only if, $\vec{u} = \vec{0}$.

The verification of each of the properties in Theorem 3 is straightforward, and left as an exercises for the reader.

As stated before, each vector $\vec{v}$ conveys magnitude and direction information. We have a method of extracting the magnitude, which we write as $\|\,\vec{v}\,\|$. *Unit vectors* are a way of extracting just the direction information from a vector.

Definition 13 Unit Vector

A **unit vector** is a vector $\vec{v}$ with a magnitude of 1; that is,

$$\| \vec{v} \| = 1.$$

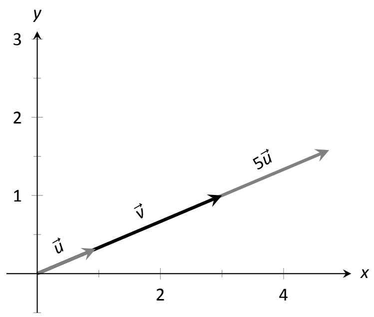


Figure 2.14: Graphing vectors in Example 13. All vectors shown have their initial point at the origin.

Consider this scenario: you are given a vector $\vec{v}$ and are told to create a vector of length 10 in the direction of $\vec{v}$. How does one do that? If we knew that $\vec{u}$ was the unit vector in the direction of $\vec{v}$, the answer would be easy: $10\vec{u}$. So how do we find $\vec{u}$?

Property 8 of Theorem 3 holds the key. If we divide $\vec{v}$ by its magnitude, it becomes a vector of length 1. Consider:

$$\left\| \frac{1}{\| \vec{v} \|}\vec{v} \right\| = \frac{1}{\| \vec{v} \|}\| \vec{v} \| \qquad \text{(we can pull out } \frac{1}{\| \vec{v} \|} \text{ as it is a scalar)}$$
$$= 1.$$

So the vector of length 10 in the direction of $\vec{v}$ is $10\dfrac{1}{\| \vec{v} \|}\vec{v}$. An example will make this more clear.

Example 13 Using Unit Vectors

Let $\vec{v} = \langle 3, 1\rangle$ and let $\vec{w} = \langle 1, 2, 2\rangle$.

1. Find the unit vector in the direction of $\vec{v}$.
2. Find the unit vector in the direction of $\vec{w}$.
3. Find the vector in the direction of $\vec{v}$ with magnitude 5.

Solution

1. We find $\| \vec{v} \| = \sqrt{10}$. So the unit vector $\vec{u}$ in the direction of $\vec{v}$ is
$$\vec{u} = \frac{1}{\sqrt{10}}\vec{v} = \left\langle \frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}} \right\rangle.$$
2. We find $\| \vec{w} \| = 3$, so the unit vector $\vec{z}$ in the direction of $\vec{w}$ is
$$\vec{u} = \frac{1}{3}\vec{w} = \left\langle \frac{1}{3}, \frac{2}{3}, \frac{2}{3} \right\rangle.$$
3. To create a vector with magnitude 5 in the direction of $\vec{v}$, we multiply the unit vector $\vec{u}$ by 5. Thus $5\vec{u} = \langle 15/\sqrt{10}, 5/\sqrt{10}\rangle$ is the vector we seek. This is sketched in Figure 2.14.

The basic formation of the unit vector $\vec{u}$ in the direction of a vector $\vec{v}$ leads to a interesting equation. It is:

$$\vec{v} = \| \vec{v} \| \frac{1}{\| \vec{v} \|}\vec{v}.$$

We rewrite the equation with parentheses to make a point:

$$\vec{v} = \underbrace{\| \vec{v} \|}_{\text{magnitude}} \cdot \underbrace{\left(\frac{1}{\| \vec{v} \|}\vec{v}\right)}_{\text{direction}}.$$

This equation illustrates the fact that a vector has both magnitude and direction, where we view a unit vector as supplying *only* direction information. Identifying unit vectors with direction allows us to define **parallel vectors**.

Definition 14 Parallel Vectors

1. Unit vectors $\vec{u}_1$ and $\vec{u}_2$ are **parallel** if $\vec{u}_1 = \pm\vec{u}_2$.
2. Nonzero vectors $\vec{v}_1$ and $\vec{v}_2$ are **parallel** if their respective unit vectors are parallel.

It is equivalent to say that vectors $\vec{v}_1$ and $\vec{v}_2$ are parallel if there is a scalar $c \neq 0$ such that $\vec{v}_1 = c\vec{v}_2$ (see marginal note).

If one graphed all unit vectors in $\mathbb{R}^2$ with the initial point at the origin, then the terminal points would all lie on the unit circle $x^2 + y^2 = 1$. Based on what we know from trigonometry, we can then say that the component form of all unit vectors in $\mathbb{R}^2$ is $\langle \cos\theta, \sin\theta\rangle$ for some angle θ.

A similar construction in $\mathbb{R}^3$ shows that the terminal points all lie on the unit sphere $x^2 + y^2 + z^2 = 1$. These vectors also have a particular component form, but its derivation is not as straightforward as the one for unit vectors in $\mathbb{R}^2$. Important concepts about unit vectors are given in the following Key Idea.

Note: $\vec{0}$ is directionless; because $\| \vec{0} \| = 0$, there is no unit vector in the "direction" of $\vec{0}$.
Some texts define two vectors as being parallel if one is a scalar multiple of the other. By this definition, $\vec{0}$ is parallel to all vectors as $\vec{0} = 0\vec{v}$ for all $\vec{v}$.
For this chapter on vector *geometry*, we will use Definition 14 of parallel as it is grounded in the fact that unit vectors provide direction information. One may adopt the convention that $\vec{0}$ is parallel to all vectors if they desire. (See also the marginal note on page 49.) In later chapters (once we move on to vector *algebra*) it will be more convenient to say that a vector $\vec{w}$ is parallel to a given vector $\vec{v}$ if there exists a scalar c such that $\vec{w} = c\vec{v}$.

Key Idea 7 Unit Vectors

1. The unit vector in the direction of $\vec{v}$ is
$$\vec{u} = \frac{1}{\| \vec{v} \|}\vec{v}.$$
2. A vector $\vec{u}$ in $\mathbb{R}^2$ is a unit vector if, and only if, its component form is $\langle \cos\theta, \sin\theta\rangle$ for some angle θ.
3. A vector $\vec{u}$ in $\mathbb{R}^3$ is a unit vector if, and only if, its component form is $\langle \sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta\rangle$ for some angles θ and φ.

Note: the component form given in Key Idea 7 for a unit vector in $\mathbb{R}^3$ is derived from the spherical coordinate system for $\mathbb{R}^3$.

These formulas can come in handy in a variety of situations, especially the formula for unit vectors in the plane.

Example 14 Finding Component Forces
Consider a weight of 50lb hanging from two chains, as shown in Figure 2.15. One chain makes an angle of 30° with the vertical, and the other an angle of 45°. Find the force applied to each chain.

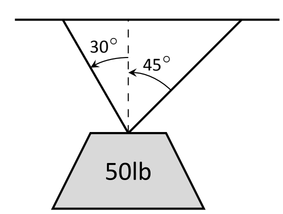


Figure 2.15: A diagram of a weight hanging from 2 chains in Example 14.

Solution Knowing that gravity is pulling the 50lb weight straight down, we can create a vector $\vec{F}$ to represent this force.

$$\vec{F} = 50\,\langle 0, -1\rangle = \langle 0, -50\rangle\,.$$

We can view each chain as "pulling" the weight up, preventing it from falling. We can represent the force from each chain with a vector. Let $\vec{F}_1$ represent the force from the chain making an angle of 30° with the vertical, and let $\vec{F}_2$ represent the force form the other chain. Convert all angles to be measured from

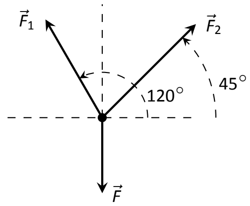


Figure 2.16: A diagram of the force vectors from Example 14.

the horizontal (as shown in Figure 2.16), and apply Key Idea 7. As we do not yet know the magnitudes of these vectors, (that is the problem at hand), we use m_1 and m_2 to represent them.

$$\vec{F}_1 = m_1 \langle \cos 120°, \sin 120° \rangle$$
$$\vec{F}_2 = m_2 \langle \cos 45°, \sin 45° \rangle$$

As the weight is not moving, we know the sum of the forces is $\vec{0}$. This gives:

$$\begin{aligned}\vec{0} &= \vec{F} + \vec{F}_1 + \vec{F}_2 \\ &= \langle 0, -50 \rangle + m_1 \langle \cos 120°, \sin 120° \rangle + m_2 \langle \cos 45°, \sin 45° \rangle\end{aligned}$$

The sum of the entries in the first component is 0, and the sum of the entries in the second component is also 0. This leads us to the following two equations:

$$\begin{aligned} m_1 \cos 120° + m_2 \cos 45° &= 0 \\ m_1 \sin 120° + m_2 \sin 45° &= 50 \end{aligned}$$

This is a simple 2-equation, 2-unknown system of linear equations. We leave it to the reader to verify that the solution is

$$m_1 = 50(\sqrt{3} - 1) \approx 36.6; \qquad m_2 = \frac{50\sqrt{2}}{1 + \sqrt{3}} \approx 25.88.$$

It might seem odd that the sum of the forces applied to the chains is more than 50lb. We leave it to a physics class to discuss the full details, but offer this short explanation. Our equations were established so that the *vertical* components of each force sums to 50lb, thus supporting the weight. Since the chains are at an angle, they also pull against each other, creating an "additional" horizontal force while holding the weight in place.

Unit vectors were very important in the previous calculation; they allowed us to define a vector in the proper direction but with an unknown magnitude. Our computations were then computed component–wise. Because such calculations are often necessary, the *standard unit vectors* can be useful.

Definition 15 Standard Unit Vectors

1. In $\mathbb{R}^2$, the standard unit vectors are

$$\vec{i} = \langle 1, 0 \rangle \quad \text{and} \quad \vec{j} = \langle 0, 1 \rangle .$$

2. In $\mathbb{R}^3$, the standard unit vectors are

$$\vec{i} = \langle 1, 0, 0 \rangle \quad \text{and} \quad \vec{j} = \langle 0, 1, 0 \rangle \quad \text{and} \quad \vec{k} = \langle 0, 0, 1 \rangle .$$

Example 15 Using standard unit vectors

1. Rewrite $\vec{v} = \langle 2, -3 \rangle$ using the standard unit vectors.
2. Rewrite $\vec{w} = 4\vec{i} - 5\vec{j} + 2\vec{k}$ in component form.

Solution

1. $$\begin{aligned}\vec{v} &= \langle 2,-3\rangle \\ &= \langle 2,0\rangle + \langle 0,-3\rangle \\ &= 2\langle 1,0\rangle - 3\langle 0,1\rangle \\ &= 2\vec{i} - 3\vec{j}\end{aligned}$$

2. $$\begin{aligned}\vec{w} &= 4\vec{i} - 5\vec{j} + 2\vec{k} \\ &= \langle 4,0,0\rangle + \langle 0,-5,0\rangle + \langle 0,0,2\rangle \\ &= \langle 4,-5,2\rangle\end{aligned}$$

These two examples demonstrate that converting between component form and the standard unit vectors is rather straightforward. Many mathematicians prefer component form, and it is the preferred notation in this text. Many engineers prefer using the standard unit vectors, and many engineering texts use that notation.

Example 16 Finding Component Force

A weight of 25lb is suspended from a chain of length 2ft while a wind pushes the weight to the right with constant force of 5lb as shown in Figure 2.17. What angle will the chain make with the vertical as a result of the wind's pushing? How much higher will the weight be?

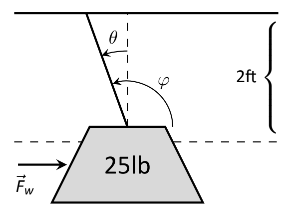


Figure 2.17: A figure of a weight being pushed by the wind in Example 16.

Solution The force of the wind is represented by the vector $\vec{F}_w = 5\vec{i}$. The force of gravity on the weight is represented by $\vec{F}_g = -25\vec{j}$. The direction and magnitude of the vector representing the force on the chain are both unknown. We represent this force with

$$\vec{F}_c = m\langle\cos\varphi, \sin\varphi\rangle = m\cos\varphi\,\vec{i} + m\sin\varphi\,\vec{j}$$

for some magnitude m and some angle with the horizontal φ. (Note: θ is the angle the chain makes with the *vertical*; φ is the angle with the *horizontal*.)

As the weight is at equilibrium, the sum of the forces is $\vec{0}$:

$$\begin{aligned}\vec{F}_c + \vec{F}_w + \vec{F}_g &= \vec{0} \\ m\cos\varphi\,\vec{i} + m\sin\varphi\,\vec{j} + 5\vec{i} - 25\vec{j} &= \vec{0}\end{aligned}$$

Thus the sum of the $\vec{i}$ and $\vec{j}$ components are 0, leading us to the following system of equations:

$$\begin{aligned}5 + m\cos\varphi &= 0 \\ -25 + m\sin\varphi &= 0\end{aligned} \tag{2.1}$$

This is enough to determine $\vec{F}_c$ already, as we know $m\cos\varphi = -5$ and $m\sin\varphi = 25$. Thus $F_c = \langle -5, 25\rangle$. We can use this to find the magnitude m:

$$m = \sqrt{(-5)^2 + 25^2} = 5\sqrt{26} \approx 25.5\text{lb}.$$

We can then use either equality from Equation (2.1) to solve for φ. We choose the first equality as using arccosine will return an angle in the 2nd quadrant:

$$5 + 5\sqrt{26}\cos\varphi = 0 \quad\Rightarrow\quad \varphi = \cos^{-1}\left(\frac{-5}{5\sqrt{26}}\right) \approx 1.7682 \approx 101.31^\circ.$$

Subtracting 90° from this angle gives us an angle of 11.31° with the vertical.

We can now use trigonometry to find out how high the weight is lifted. The diagram shows that a right triangle is formed with the 2ft chain as the hypotenuse with an interior angle of 11.31°. The length of the adjacent side (in the diagram, the dashed vertical line) is $2\cos 11.31^\circ \approx 1.96$ft. Thus the weight is lifted by about 0.04ft, almost 1/2in.

The algebra we have applied to vectors is already demonstrating itself to be very useful. There are two more fundamental operations we can perform with vectors, the *dot product* and the *cross product*. The next two sections explore each in turn.

Exercises 2.2

Terms and Concepts

1. Name two different things that cannot be described with just one number, but rather need 2 or more numbers to fully describe them.

2. What is the difference between $(1, 2)$ and $\langle 1, 2 \rangle$?

3. What is a unit vector?

4. What does it mean for two vectors to be parallel?

5. What effect does multiplying a vector by -2 have?

Problems

In Exercises 6 – 9, points P and Q are given. Write the vector $\vec{v}PQ$ in component form and using the standard unit vectors.

6. $P = (2, -1), \quad Q = (3, 5)$

7. $P = (3, 2), \quad Q = (7, -2)$

8. $P = (0, 3, -1), \quad Q = (6, 2, 5)$

9. $P = (2, 1, 2), \quad Q = (4, 3, 2)$

10. Let $\vec{u} = \langle 1, -2 \rangle$ and $\vec{v} = \langle 1, 1 \rangle$.

    (a) Find $\vec{u} + \vec{v}$, $\vec{u} - \vec{v}$, $2\vec{u} - 3\vec{v}$.

    (b) Sketch the above vectors on the same axes, along with $\vec{u}$ and $\vec{v}$.

    (c) Find $\vec{x}$ where $\vec{u} + \vec{x} = 2\vec{v} - \vec{x}$.

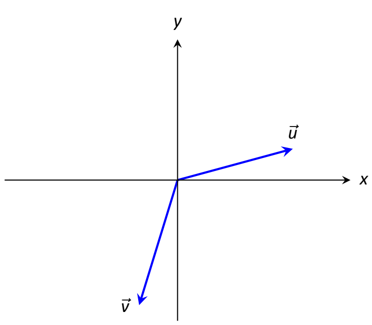
11. Let $\vec{u} = \langle 1, 1, -1 \rangle$ and $\vec{v} = \langle 2, 1, 2 \rangle$.

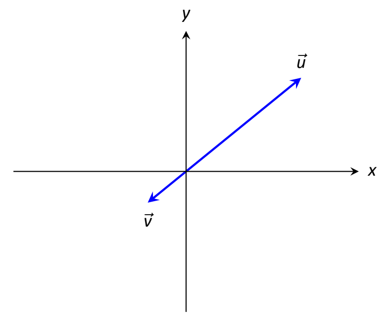
    (a) Find $\vec{u} + \vec{v}$, $\vec{u} - \vec{v}$, $\pi\vec{u} - \sqrt{2}\vec{v}$.

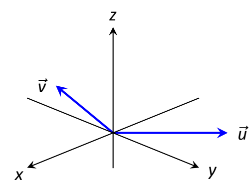
    (b) Sketch the above vectors on the same axes, along with $\vec{u}$ and $\vec{v}$.

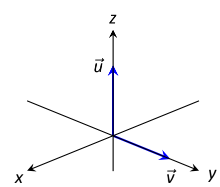
    (c) Find $\vec{x}$ where $\vec{u} + \vec{x} = \vec{v} + 2\vec{x}$.

In Exercises 12 – 15, sketch $\vec{u}$, $\vec{v}$, $\vec{u} + \vec{v}$ and $\vec{u} - \vec{v}$ on the same axes.

12. 

13. 

14. 

15. 

In Exercises 16 – 19, find $\| \vec{u} \|$, $\| \vec{v} \|$, $\| \vec{u} + \vec{v} \|$ and $\| \vec{u} - \vec{v} \|$.

16. $\vec{u} = \langle 2, 1 \rangle, \quad \vec{v} = \langle 3, -2 \rangle$

17. $\vec{u} = \langle -3, 2, 2 \rangle, \quad \vec{v} = \langle 1, -1, 1 \rangle$

18. $\vec{u} = \langle 1, 2 \rangle, \quad \vec{v} = \langle -3, -6 \rangle$

19. $\vec{u} = \langle 2, -3, 6 \rangle, \quad \vec{v} = \langle 10, -15, 30 \rangle$

20. Under what conditions is $\| \vec{u} \| + \| \vec{v} \| = \| \vec{u} + \vec{v} \|$?

In Exercises 21 – 24, find the unit vector $\vec{u}$ in the direction of $\vec{v}$.

21. $\vec{v} = \langle 3, 7 \rangle$

22. $\vec{v} = \langle 6, 8 \rangle$

23. $\vec{v} = \langle 1, -2, 2 \rangle$

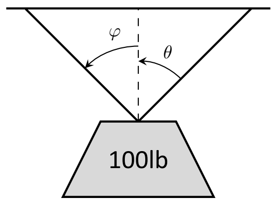
24. $\vec{v} = \langle 2, -2, 2 \rangle$

25. Find the unit vector in the first quadrant of $\mathbb{R}^2$ that makes a 50° angle with the x-axis.

26. Find the unit vector in the second quadrant of $\mathbb{R}^2$ that makes a 30° angle with the y-axis.

27. Verify, from Key Idea 7, that $\vec{u} = \langle \sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta\rangle$ is a unit vector for all angles θ and φ.

A weight of 100lb is suspended from two chains, making angles with the vertical of θ and φ as shown in the figure below.



In Exercises 28 – 31, angles θ and φ are given. Find the force applied to each chain.

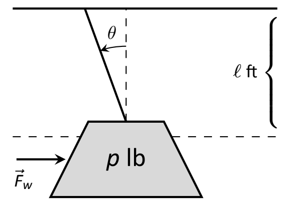
28. $\theta = 30^\circ$, $\varphi = 30^\circ$

29. $\theta = 60^\circ$, $\varphi = 60^\circ$

30. $\theta = 20^\circ$, $\varphi = 15^\circ$

31. $\theta = 0^\circ$, $\varphi = 0^\circ$

A weight of $p$lb is suspended from a chain of length ℓ while a constant force of $\vec{F}_w$ pushes the weight to the right, making an angle of θ with the vertical, as shown in the figure below.



In Exercises 32 – 35, a force $\vec{F}_w$ and length ℓ are given. Find the angle θ and the height the weight is lifted as it moves to the right.

32. $\vec{F}_w = 1\text{lb}$, $\ell = 1\text{ft}$, $p = 1\text{lb}$

33. $\vec{F}_w = 1\text{lb}$, $\ell = 1\text{ft}$, $p = 10\text{lb}$

34. $\vec{F}_w = 1\text{lb}$, $\ell = 10\text{ft}$, $p = 1\text{lb}$

35. $\vec{F}_w = 10\text{lb}$, $\ell = 10\text{ft}$, $p = 1\text{lb}$

2.3 The Dot Product

The previous section introduced vectors and described how to add them together and how to multiply them by scalars. This section introduces *a* multiplication on vectors called the **dot product**.

Definition 16 Dot Product

1. Let $\vec{u} = \langle u_1, u_2 \rangle$ and $\vec{v} = \langle v_1, v_2 \rangle$ in $\mathbb{R}^2$. The **dot product** of $\vec{u}$ and $\vec{v}$, denoted $\vec{u} \cdot \vec{v}$, is
$$\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2.$$
2. Let $\vec{u} = \langle u_1, u_2, u_3 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3 \rangle$ in $\mathbb{R}^3$. The **dot product** of $\vec{u}$ and $\vec{v}$, denoted $\vec{u} \cdot \vec{v}$, is
$$\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2 + u_3v_3.$$

Note how this product of vectors returns a *scalar*, not another vector. We practice evaluating a dot product in the following example, then we will discuss why this product is useful.

Example 17 Evaluating dot products

1. Let $\vec{u} = \langle 1, 2 \rangle$, $\vec{v} = \langle 3, -1 \rangle$ in $\mathbb{R}^2$. Find $\vec{u} \cdot \vec{v}$.
2. Let $\vec{x} = \langle 2, -2, 5 \rangle$ and $\vec{y} = \langle -1, 0, 3 \rangle$ in $\mathbb{R}^3$. Find $\vec{x} \cdot \vec{y}$.

Solution

1. Using Definition 16, we have
$$\vec{u} \cdot \vec{v} = 1(3) + 2(-1) = 1.$$
2. Using the definition, we have
$$\vec{x} \cdot \vec{y} = 2(-1) - 2(0) + 5(3) = 13.$$

The dot product, as shown by the preceding example, is very simple to evaluate. It is only the sum of products. While the definition gives no hint as to why we would care about this operation, there is an amazing connection between the dot product and angles formed by the vectors. Before stating this connection, we give a theorem stating some of the properties of the dot product.

Note: proving Theorem 4 is straightforward and left to the reader. The reader is cautioned, however, that proofs must be *general*: choosing particular numbers for the vectors $\vec{u}$, $\vec{v}$, etc. only shows that the properties hold for those particular numbers. Instead, one should write $\vec{u} = \langle u_1, u_2, u_3\rangle$, $\vec{v} = \langle v_1, v_2, v_3\rangle$, etc. and then proceed using the rules of algebra for real numbers in Section 1.2. For example, $\vec{u}\cdot\vec{v} = \vec{v}\cdot\vec{u}$ since

$$\begin{aligned}\vec{u}\cdot\vec{v} &= u_1v_1 + u_2v_2 + u_3v_3\\ &= v_1u_1 + v_2u_2 + v_3u_3\\ &= \vec{v}\cdot\vec{u},\end{aligned}$$

and this argument is valid no matter what values are substituted for the components of the two vectors.

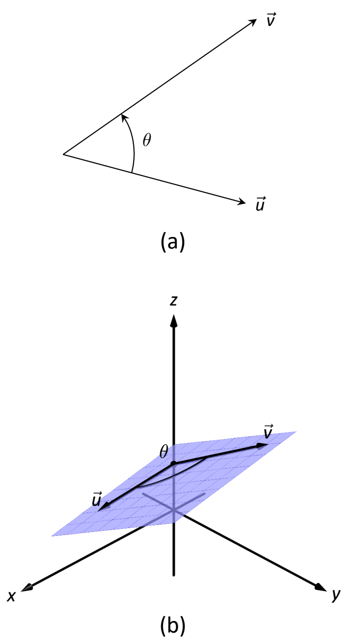


Figure 2.18: Illustrating the angle formed by two vectors with the same initial point.

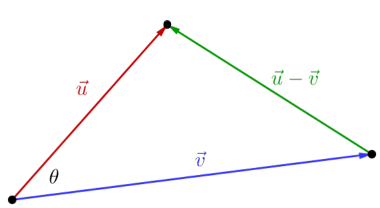


Figure 2.19: Proving Theorem 5

Theorem 4 Properties of the Dot Product

Let $\vec{u}$, $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ and let c be a scalar.

1. $\vec{u}\cdot\vec{v} = \vec{v}\cdot\vec{u}$ — Commutative Property
2. $\vec{u}\cdot(\vec{v}+\vec{w}) = \vec{u}\cdot\vec{v}+\vec{u}\cdot\vec{w}$ — Distributive Property
3. $c(\vec{u}\cdot\vec{v}) = (c\vec{u})\cdot\vec{v} = \vec{u}\cdot(c\vec{v})$
4. $\vec{0}\cdot\vec{v} = 0$
5. $\vec{v}\cdot\vec{v} = \|\,\vec{v}\,\|^2$

The last statement of the theorem makes a handy connection between the magnitude of a vector and the dot product with itself. Our definition and theorem give properties of the dot product, but we are still likely wondering "What does the dot product *mean*?" It is helpful to understand that the dot product of a vector with itself is connected to its magnitude.

The next theorem extends this understanding by connecting the dot product to magnitudes and angles. Given vectors $\vec{u}$ and $\vec{v}$ in the plane, an angle θ is clearly formed when $\vec{u}$ and $\vec{v}$ are drawn with the same initial point as illustrated in Figure 2.18(a). (We always take θ to be the angle in $[0,\pi]$ as two angles are actually created.)

The same is also true of 2 vectors in space: given $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$ with the same initial point, there is a plane that contains both $\vec{u}$ and $\vec{v}$. (When $\vec{u}$ and $\vec{v}$ are co-linear, there are infinite planes that contain both vectors.) In that plane, we can again find an angle θ between them (and again, $0\le\theta\le\pi$). This is illustrated in Figure 2.18(b).

The following theorem connects this angle θ to the dot product of $\vec{u}$ and $\vec{v}$.

Theorem 5 The Dot Product and Angles

Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$. Then

$$\vec{u}\cdot\vec{v} = \|\,\vec{u}\,\|\,\|\,\vec{v}\,\|\cos\theta,$$

where θ, $0\le\theta\le\pi$, is the angle between $\vec{u}$ and $\vec{v}$.

The proof of Theorem 5 is an application of the Law of Cosines, using the properties in Theorem 4. Referring to Figure 2.19, if we let $a = \|\vec{u}\|$, $b = \|\vec{v}\|$, and $c = \|\vec{u}-\vec{v}\|$, then the Law of Cosines tells us that

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

Thus, we have

$$\begin{aligned}\|\vec{u}-\vec{v}\|^2 &= \|\vec{u}\|^2+\|\vec{v}\|^2-2\,\|\vec{u}\|\,\|\vec{v}\|\cos\theta\\ (\vec{u}-\vec{v})\cdot(\vec{u}-\vec{v}) &= \vec{u}\cdot\vec{u}+\vec{v}\cdot\vec{v}-2\,\|\vec{u}\|\,\|\vec{v}\|\cos\theta\\ \vec{u}\cdot\vec{u}-\vec{u}\cdot\vec{v}-\vec{v}\cdot\vec{u}+\vec{v}\cdot\vec{v} &= \vec{u}\cdot\vec{u}+\vec{v}\cdot\vec{v}-2\,\|\vec{u}\|\,\|\vec{v}\|\cos\theta\\ -2\vec{u}\cdot\vec{v} &= -2\,\|\vec{u}\|\,\|\vec{v}\|\cos\theta\\ \vec{u}\cdot\vec{v} &= \|\vec{u}\|\,\|\vec{v}\|\cos\theta,\end{aligned}$$

as required.

When θ is an acute angle (i.e., $0 \leq \theta < \pi/2$), $\cos\theta$ is positive; when $\theta = \pi/2$, $\cos\theta = 0$; when θ is an obtuse angle ($\pi/2 < \theta \leq \pi$), $\cos\theta$ is negative. Thus the sign of the dot product gives a general indication of the angle between the vectors, illustrated in Figure 2.20.

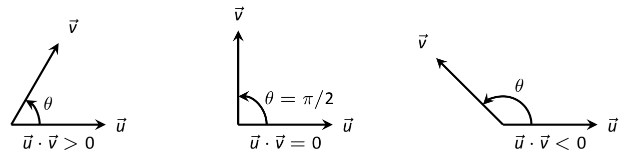


Figure 2.20: Illustrating the relationship between the angle between vectors and the sign of their dot product.

We *can* use Theorem 5 to compute the dot product, but generally this theorem is used to find the angle between known vectors (since the dot product is generally easy to compute). To this end, we rewrite the theorem's equation as

$$\cos\theta = \frac{\vec u \cdot \vec v}{\|\,\vec u\,\|\|\,\vec v\,\|} \quad \Leftrightarrow \quad \theta = \cos^{-1}\left(\frac{\vec u \cdot \vec v}{\|\,\vec u\,\|\|\,\vec v\,\|}\right).$$

We practice using this theorem in the following example.

Example 18 Using the dot product to find angles

Let $\vec u = \langle 3, 1\rangle$, $\vec v = \langle -2, 6\rangle$ and $\vec w = \langle -4, 3\rangle$, as shown in Figure 2.21. Find the angles α, β and θ.

Solution We start by computing the magnitude of each vector.

$$\|\,\vec u\,\| = \sqrt{10}; \quad \|\,\vec v\,\| = 2\sqrt{10}; \quad \|\,\vec w\,\| = 5.$$

We now apply Theorem 5 to find the angles.

$$\begin{aligned}\alpha &= \cos^{-1}\left(\frac{\vec u \cdot \vec v}{(\sqrt{10})(2\sqrt{10})}\right)\\ &= \cos^{-1}(0) = \frac{\pi}{2} = 90^\circ.\end{aligned}$$

$$\begin{aligned}\beta &= \cos^{-1}\left(\frac{\vec v \cdot \vec w}{(2\sqrt{10})(5)}\right)\\ &= \cos^{-1}\left(\frac{26}{10\sqrt{10}}\right)\\ &\approx 0.6055 \approx 34.7^\circ.\end{aligned}$$

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{\vec u \cdot \vec w}{(\sqrt{10})(5)}\right)\\ &= \cos^{-1}\left(\frac{-9}{5\sqrt{10}}\right)\\ &\approx 2.1763 \approx 124.7^\circ\end{aligned}$$

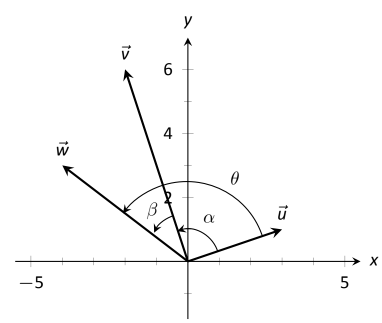


Figure 2.21: Vectors used in Example 18.

We see from our computation that $\alpha + \beta = \theta$, as indicated by Figure 2.21. While we knew this should be the case, it is nice to see that this non-intuitive formula indeed returns the results we expected.

We do a similar example next in the context of vectors in space.

Example 19 Using the dot product to find angles
Let $\vec{u} = \langle 1, 1, 1\rangle$, $\vec{v} = \langle -1, 3, -2\rangle$ and $\vec{w} = \langle -5, 1, 4\rangle$, as illustrated in Figure 2.22. Find the angle between each pair of vectors.

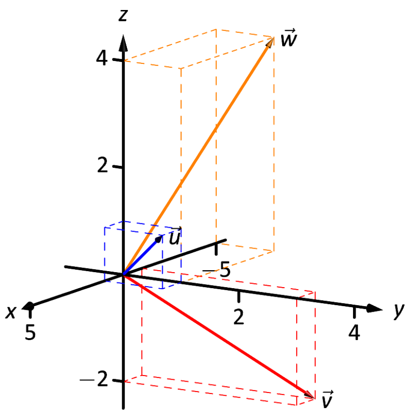


Figure 2.22: Vectors used in Example 19.

Solution

1. Between $\vec{u}$ and $\vec{v}$:

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{\vec{u}\cdot\vec{v}}{\parallel\vec{u}\parallel\parallel\vec{v}\parallel}\right)\\ &= \cos^{-1}\left(\frac{0}{\sqrt{3}\sqrt{14}}\right)\\ &= \frac{\pi}{2}.\end{aligned}$$

2. Between $\vec{u}$ and $\vec{w}$:

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{\vec{u}\cdot\vec{w}}{\parallel\vec{u}\parallel\parallel\vec{w}\parallel}\right)\\ &= \cos^{-1}\left(\frac{0}{\sqrt{3}\sqrt{42}}\right)\\ &= \frac{\pi}{2}.\end{aligned}$$

3. Between $\vec{v}$ and $\vec{w}$:

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{\vec{v}\cdot\vec{w}}{\parallel\vec{v}\parallel\parallel\vec{w}\parallel}\right)\\ &= \cos^{-1}\left(\frac{0}{\sqrt{14}\sqrt{42}}\right)\\ &= \frac{\pi}{2}.\end{aligned}$$

While our work shows that each angle is $\pi/2$, i.e., 90°, none of these angles looks to be a right angle in Figure 2.22. Such is the case when drawing three–dimensional objects on the page.

All three angles between these vectors was $\pi/2$, or 90°. We know from geometry and everyday life that 90° angles are "nice" for a variety of reasons, so it should seem significant that these angles are all $\pi/2$. Notice the common feature in each calculation (and also the calculation of α in Example 18): the dot products of each pair of angles was 0. We use this as a basis for a definition of the term **orthogonal**, which is essentially synonymous to *perpendicular*.

Definition 17 Orthogonal

Vectors $\vec{u}$ and $\vec{v}$ are **orthogonal** if their dot product is 0.

Note: The term *perpendicular* originally referred to lines. As mathematics progressed, the concept of "being at right angles to" was applied to other objects, such as vectors and planes, and the term *orthogonal* was introduced. It is especially used when discussing objects that are hard, or impossible, to visualize: two vectors in 5-dimensional space are orthogonal if their dot product is 0. It is not wrong to say they are *perpendicular*, but common convention gives preference to the word *orthogonal*.
Note also that Definition 17 makes sense if either $\vec{u}$ or $\vec{v}$ is the zero vector, but this is not the case for the conventional understanding of the word perpendicular.

Example 20 **Finding orthogonal vectors**

Let $\vec{u} = \langle 3, 5 \rangle$ and $\vec{v} = \langle 1, 2, 3 \rangle$.

1. Find two vectors in $\mathbb{R}^2$ that are orthogonal to $\vec{u}$.
2. Find two non–parallel vectors in $\mathbb{R}^3$ that are orthogonal to $\vec{v}$.

Solution

1. Recall that a line perpendicular to a line with slope m has slope $-1/m$, the "opposite reciprocal slope." We can think of the slope of $\vec{u}$ as $5/3$, its "rise over run." A vector orthogonal to $\vec{u}$ will have slope $-3/5$. There are many such choices, though all parallel:
$$\langle -5, 3 \rangle \quad \text{or} \quad \langle 5, -3 \rangle \quad \text{or} \quad \langle -10, 6 \rangle \quad \text{or} \quad \langle 15, -9 \rangle \text{, etc.}$$
2. There are infinite directions in space orthogonal to any given direction, so there are an infinite number of non–parallel vectors orthogonal to $\vec{v}$. Since there are so many, we have great leeway in finding some.

   One way is to arbitrarily pick values for the first two components, leaving the third unknown. For instance, let $\vec{v}_1 = \langle 2, 7, z \rangle$. If $\vec{v}_1$ is to be orthogonal to $\vec{v}$, then $\vec{v}_1 \cdot \vec{v} = 0$, so
$$2 + 14 + 3z = 0 \quad \Rightarrow z = \frac{-16}{3}.$$
   So $\vec{v}_1 = \langle 2, 7, -16/3 \rangle$ is orthogonal to $\vec{v}$. We can apply a similar technique by leaving the first or second component unknown.

   Another method of finding a vector orthogonal to $\vec{v}$ mirrors what we did in part 1. Let $\vec{v}_2 = \langle -2, 1, 0 \rangle$. Here we switched the first two components of $\vec{v}$, changing the sign of one of them (similar to the "opposite reciprocal" concept before). Letting the third component be 0 effectively ignores the third component of $\vec{v}$, and it is easy to see that
$$\vec{v}_2 \cdot \vec{v} = \langle -2, 1, 0 \rangle \cdot \langle 1, 2, 3 \rangle = 0.$$
   Clearly $\vec{v}_1$ and $\vec{v}_2$ are not parallel.

An important construction is illustrated in Figure 2.23, where vectors $\vec{u}$ and $\vec{v}$ are sketched. In part (a), a dotted line is drawn from the tip of $\vec{u}$ to the line containing $\vec{v}$, where the dotted line is orthogonal to $\vec{v}$. In part (b), the dotted line is replaced with the vector $\vec{z}$ and $\vec{w}$ is formed, parallel to $\vec{v}$. It is clear by the diagram that $\vec{u} = \vec{w} + \vec{z}$. What is important about this construction is this: $\vec{u}$ is *decomposed* as the sum of two vectors, one of which is parallel to $\vec{v}$ and one that is perpendicular to $\vec{v}$. It is hard to overstate the importance of this construction (as we'll see in upcoming examples).

The vectors $\vec{w}$, $\vec{z}$ and $\vec{u}$ as shown in Figure 2.23 (b) form a right triangle, where the angle between $\vec{v}$ and $\vec{u}$ is labeled θ. We can find $\vec{w}$ in terms of $\vec{v}$ and $\vec{u}$.

Using trigonometry, we can state that
$$\| \vec{w} \| = \| \vec{u} \| \cos\theta. \tag{2.2}$$

We also know that $\vec{w}$ is parallel to to $\vec{v}$; that is, the direction of $\vec{w}$ is the direction of $\vec{v}$, described by the unit vector $\frac{1}{\| \vec{v} \|}\vec{v}$. The vector $\vec{w}$ is the vector in the direction $\frac{1}{\| \vec{v} \|}\vec{v}$ with magnitude $\| \vec{u} \| \cos\theta$:

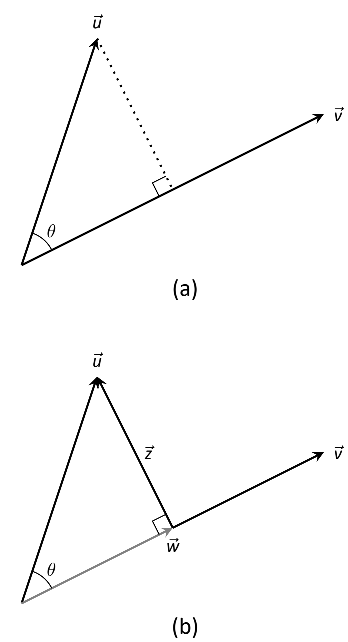


Figure 2.23: Developing the construction of the *orthogonal projection*.

$$\begin{aligned}\vec w &= \Big(\parallel \vec u\parallel \cos\theta\Big)\frac{1}{\parallel \vec v\parallel}\vec v.\\ &= \left(\parallel \vec u\parallel \frac{\vec u\cdot\vec v}{\parallel \vec u\parallel\parallel \vec v\parallel}\right)\frac{1}{\parallel \vec v\parallel}\vec v && \text{Replace } \cos\theta \text{ using Theorem 5}\\ &= \frac{\vec u\cdot\vec v}{\parallel \vec v\parallel^2}\vec v = \frac{\vec u\cdot\vec v}{\vec v\cdot\vec v}\vec v && \text{Using Theorem 4.}\end{aligned}$$

Since this construction is so important, it is given a special name.

Definition 18 Orthogonal Projection

Let $\vec u$ and $\vec v$ be given. The **orthogonal projection of $\vec u$ onto $\vec v$**, denoted $\text{proj}_{\,\vec v}\,\vec u$, is

$$\text{proj}_{\,\vec v}\,\vec u = \left(\frac{\vec u\cdot\vec v}{\vec v\cdot\vec v}\right)\vec v.$$

Example 21 Computing the orthogonal projection

1. Let $\vec u = \langle -2, 1\rangle$ and $\vec v = \langle 3, 1\rangle$. Find $\text{proj}_{\,\vec v}\,\vec u$, and sketch all three vectors with initial points at the origin.
2. Let $\vec w = \langle 2, 1, 3\rangle$ and $\vec x = \langle 1, 1, 1\rangle$. Find $\text{proj}_{\,\vec x}\,\vec w$, and sketch all three vectors with initial points at the origin.

Solution

1. Applying Definition 18, we have

$$\begin{aligned}\text{proj}_{\,\vec v}\,\vec u &= \left(\frac{\vec u\cdot\vec v}{\vec v\cdot\vec v}\right)\vec v\\ &= \frac{-5}{10}\langle 3, 1\rangle\\ &= \left\langle -\frac{3}{2}, -\frac{1}{2}\right\rangle.\end{aligned}$$

Vectors $\vec u$, $\vec v$ and $\text{proj}_{\,\vec v}\,\vec u$ are sketched in Figure 2.24(a). Note how the projection is parallel to $\vec v$; that is, it lies on the same line through the origin as $\vec v$, although it points in the opposite direction. That is because the angle between $\vec u$ and $\vec v$ is obtuse (i.e., greater than 90°).

2. Apply the definition:

$$\begin{aligned}\text{proj}_{\,\vec x}\,\vec w &= \frac{\vec w\cdot\vec x}{\vec x\cdot\vec x}\vec x\\ &= \frac{6}{3}\langle 1, 1, 1\rangle\\ &= \langle 2, 2, 2\rangle.\end{aligned}$$

These vectors are sketched in Figure 2.24(b), and again in part (c) from a different perspective. Because of the nature of graphing these vectors, the sketch in part (b) makes it difficult to recognize that the drawn projection has the geometric properties it should. The graph shown in part (c) illustrates these properties better.

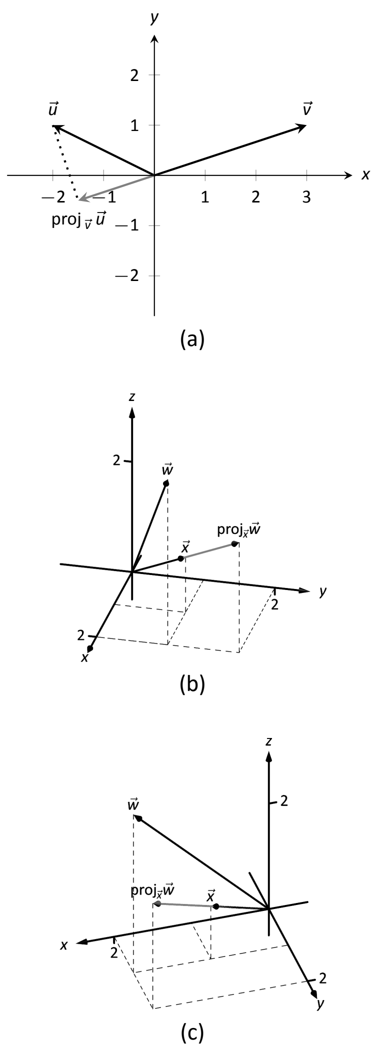


Figure 2.24: Graphing the vectors used in Example 21.

Consider Figure 2.25 where the concept of the orthogonal projection is again illustrated. It is clear that

$$\vec{u} = \text{proj}_{\,\vec{v}}\,\vec{u} + \vec{z}. \tag{2.3}$$

As we know what $\vec{u}$ and $\text{proj}_{\,\vec{v}}\,\vec{u}$ are, we can solve for $\vec{z}$ and state that

$$\vec{z} = \vec{u} - \text{proj}_{\,\vec{v}}\,\vec{u}.$$

This leads us to rewrite Equation (2.3) in a seemingly silly way:

$$\vec{u} = \text{proj}_{\,\vec{v}}\,\vec{u} + (\vec{u} - \text{proj}_{\,\vec{v}}\,\vec{u}).$$

This is not nonsense, as pointed out in the following Key Idea. (Notation note: the expression "$\parallel \vec{y}$" means "is parallel to $\vec{y}$." We can use this notation to state "$\vec{x} \parallel \vec{y}$" which means "$\vec{x}$ is parallel to $\vec{y}$." The expression "$\perp \vec{y}$" means "is orthogonal to $\vec{y}$," and is used similarly.)

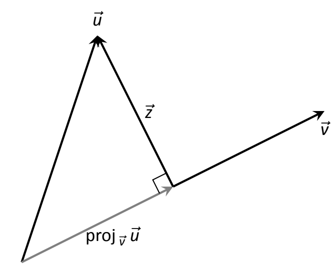


Figure 2.25: Illustrating the orthogonal projection.

Key Idea 8 Orthogonal Decomposition of Vectors

Let $\vec{u}$ and $\vec{v}$ be given. Then $\vec{u}$ can be written as the sum of two vectors, one of which is parallel to $\vec{v}$, and one of which is orthogonal to $\vec{v}$:

$$\vec{u} = \underbrace{\text{proj}_{\,\vec{v}}\,\vec{u}}_{\parallel\,\vec{v}} + (\underbrace{\vec{u} - \text{proj}_{\,\vec{v}}\,\vec{u}}_{\perp\,\vec{v}}).$$

We illustrate the use of this equality in the following example.

Example 22 Orthogonal decomposition of vectors

1. Let $\vec{u} = \langle -2, 1\rangle$ and $\vec{v} = \langle 3, 1\rangle$ as in Example 21. Decompose $\vec{u}$ as the sum of a vector parallel to $\vec{v}$ and a vector orthogonal to $\vec{v}$.

2. Let $\vec{w} = \langle 2, 1, 3\rangle$ and $\vec{x} = \langle 1, 1, 1\rangle$ as in Example 21. Decompose $\vec{w}$ as the sum of a vector parallel to $\vec{x}$ and a vector orthogonal to $\vec{x}$.

Solution

1. In Example 21, we found that $\text{proj}_{\,\vec{v}}\,\vec{u} = \langle -1.5, -0.5\rangle$. Let

$$\vec{z} = \vec{u} - \text{proj}_{\,\vec{v}}\,\vec{u} = \langle -2, 1\rangle - \langle -1.5, -0.5\rangle = \langle -0.5, 1.5\rangle.$$

Is $\vec{z}$ orthogonal to $\vec{v}$? (I.e, is $\vec{z} \perp \vec{v}$?) We check for orthogonality with the dot product:

$$\vec{z} \cdot \vec{v} = \langle -0.5, 1.5\rangle \cdot \langle 3, 1\rangle = 0.$$

Since the dot product is 0, we know $\vec{z} \perp \vec{v}$. Thus:

$$\begin{aligned} \vec{u} &= \text{proj}_{\,\vec{v}}\,\vec{u} \ + \ (\vec{u} - \text{proj}_{\,\vec{v}}\,\vec{u}) \\ \langle -2, 1\rangle &= \underbrace{\langle -1.5, -0.5\rangle}_{\parallel\,\vec{v}} \ + \ \underbrace{\langle -0.5, 1.5\rangle}_{\perp\,\vec{v}}. \end{aligned}$$

2. We found in Example 21 that $\text{proj}_{\,\vec{x}}\,\vec{w} = \langle 2, 2, 2\rangle$. Applying the Key Idea, we have:

$$\vec{z} = \vec{w} - \text{proj}_{\,\vec{x}}\,\vec{w} = \langle 2, 1, 3\rangle - \langle 2, 2, 2\rangle = \langle 0, -1, 1\rangle.$$

Note: The argument leading to Definition 18 is not quite a proof, since it depended on choices made in forming the diagram in Figure 2.23. However, we can easily verify that the result in Key Idea 8 is always valid: since

$$\begin{aligned} \vec{v} \cdot (\vec{u} - \text{proj}_{\,\vec{v}}\,\vec{u}) &= \vec{v} \cdot \vec{u} - \vec{v} \cdot \left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{v}\|^2}\vec{v}\right) \\ &= \vec{v} \cdot \vec{u} - \frac{\vec{u} \cdot \vec{v}}{\vec{v} \cdot \vec{v}}(\vec{v} \cdot \vec{v}) \\ &= \vec{v} \cdot \vec{u} - \vec{u} \cdot \vec{v} = 0 \end{aligned}$$

for **any** vectors $\vec{u}$ and $\vec{v} \neq \vec{0}$, we are guaranteed that the vector $u - \text{proj}_{\,\vec{v}}\,\vec{u}$ will always be orthogonal to $\vec{v}$.

We check to see if $\vec{z} \perp \vec{x}$:

$$\vec{z} \cdot \vec{x} = \langle 0, -1, 1 \rangle \cdot \langle 1, 1, 1 \rangle = 0.$$

Since the dot product is 0, we know the two vectors are orthogonal. We now write $\vec{w}$ as the sum of two vectors, one parallel and one orthogonal to $\vec{x}$:

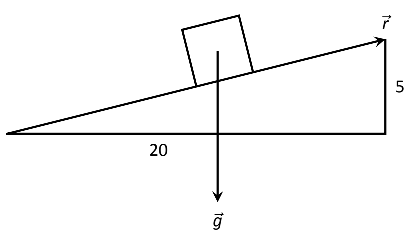
$$\begin{aligned}\vec{w} &= \text{proj}_{\vec{x}}\,\vec{w} \;+\; (\vec{w} - \text{proj}_{\vec{x}}\,\vec{w})\\ \langle 2,1,3\rangle &= \underbrace{\langle 2,2,2\rangle}_{\parallel\, \vec{x}} \;+\; \underbrace{\langle 0,-1,1\rangle}_{\perp\, \vec{x}}\end{aligned}$$

We give an example of where this decomposition is useful.

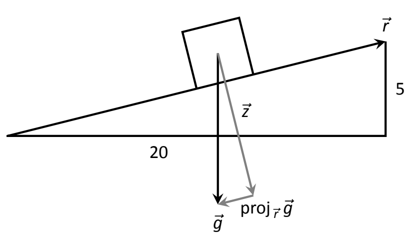
Example 23 Orthogonally decomposing a force vector

Consider Figure 2.26(a), showing a box weighing 50lb on a ramp that rises 5ft over a span of 20ft. Find the components of force, and their magnitudes, acting on the box (as sketched in part (b) of the figure):

1. in the direction of the ramp, and
2. orthogonal to the ramp.



(a)



(b)

Figure 2.26: Sketching the ramp and box in Example 23. Note: *The vectors are not drawn to scale.*

Solution As the ramp rises 5ft over a horizontal distance of 20ft, we can represent the direction of the ramp with the vector $\vec{r} = \langle 20, 5\rangle$. Gravity pulls down with a force of 50lb, which we represent with $\vec{g} = \langle 0, -50\rangle$.

1. To find the force of gravity in the direction of the ramp, we compute $\text{proj}_{\vec{r}}\,\vec{g}$:

$$\begin{aligned}\text{proj}_{\vec{r}}\,\vec{g} &= \frac{\vec{g}\cdot\vec{r}}{\vec{r}\cdot\vec{r}}\vec{r}\\ &= \frac{-250}{425}\langle 20, 5\rangle\\ &= \left\langle -\frac{200}{17}, -\frac{50}{17}\right\rangle \approx \langle -11.76, -2.94\rangle\,.\end{aligned}$$

The magnitude of $\text{proj}_{\vec{r}}\,\vec{g}$ is $\|\,\text{proj}_{\vec{r}}\,\vec{g}\,\| = 50/\sqrt{17} \approx 12.13$lb. Though the box weighs 50lb, a force of about 12lb is enough to keep the box from sliding down the ramp.

2. To find the component $\vec{z}$ of gravity orthogonal to the ramp, we use Key Idea 8.

$$\begin{aligned}\vec{z} &= \vec{g} - \text{proj}_{\vec{r}}\,\vec{g}\\ &= \left\langle \frac{200}{17}, -\frac{800}{17}\right\rangle \approx \langle 11.76, -47.06\rangle\,.\end{aligned}$$

The magnitude of this force is $\|\,\vec{z}\,\| \approx 48.51$lb. In physics and engineering, knowing this force is important when computing things like static frictional force. (For instance, we could easily compute if the static frictional force alone was enough to keep the box from sliding down the ramp.)

Application to Work

In physics, the application of a force F to move an object in a straight line a distance d produces *work*; the amount of work W is $W = Fd$, (where F is in the direction of travel). The orthogonal projection allows us to compute work when the force is not in the direction of travel.

Consider Figure 2.27, where a force $\vec{F}$ is being applied to an object moving in the direction of $\vec{d}$. (The distance the object travels is the magnitude of $\vec{d}$.) The work done is the amount of force in the direction of $\vec{d}$, $\| \operatorname{proj}_{\vec{d}} \vec{F} \|$, times $\| \vec{d} \|$:

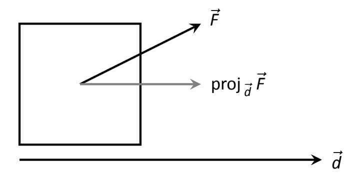


Figure 2.27: Finding work when the force and direction of travel are given as vectors.

$$\begin{aligned}
\| \operatorname{proj}_{\vec{d}} \vec{F} \| \cdot \| \vec{d} \| &= \left\| \frac{\vec{F} \cdot \vec{d}}{\vec{d} \cdot \vec{d}} \vec{d} \right\| \cdot \| \vec{d} \| \\
&= \left| \frac{\vec{F} \cdot \vec{d}}{\| \vec{d} \|^2} \right| \cdot \| \vec{d} \| \cdot \| \vec{d} \| \\
&= \frac{\left| \vec{F} \cdot \vec{d} \right|}{\| \vec{d} \|^2} \| \vec{d} \|^2 \\
&= \left| \vec{F} \cdot \vec{d} \right|.
\end{aligned}$$

The expression $\vec{F} \cdot \vec{d}$ will be positive if the angle between $\vec{F}$ and $\vec{d}$ is acute; when the angle is obtuse (hence $\vec{F} \cdot \vec{d}$ is negative), the force is causing motion in the opposite direction of $\vec{d}$, resulting in "negative work." We want to capture this sign, so we drop the absolute value and find that $W = \vec{F} \cdot \vec{d}$.

Definition 19 Work

Let $\vec{F}$ be a constant force that moves an object in a straight line from point P to point Q. Let $\vec{d} = \vec{v}PQ$. The **work** W done by $\vec{F}$ along $\vec{d}$ is $W = \vec{F} \cdot \vec{d}$.

Example 24 Computing work

A man slides a box along a ramp that rises 3ft over a distance of 15ft by applying 50lb of force as shown in Figure 2.28. Compute the work done.

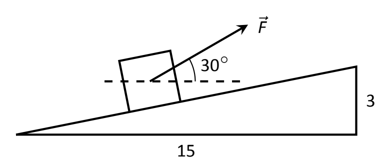


Figure 2.28: Computing work when sliding a box up a ramp in Example 24.

Solution The figure indicates that the force applied makes a 30° angle with the horizontal, so $\vec{F} = 50 \langle \cos 30^\circ, \sin 30^\circ \rangle \approx \langle 43.3, 25 \rangle$. The ramp is represented by $\vec{d} = \langle 15, 3 \rangle$. The work done is simply

$$\vec{F} \cdot \vec{d} = 50 \langle \cos 30^\circ, \sin 30^\circ \rangle \cdot \langle 15, 3 \rangle \approx 724.5\text{ft–lb}.$$

Note how we did not actually compute the distance the object traveled, nor the magnitude of the force in the direction of travel; this is all inherently computed by the dot product!

The dot product is a powerful way of evaluating computations that depend on angles without actually using angles. The next section explores another "product" on vectors, the *cross product*. Once again, angles play an important role, though in a much different way.

Exercises 2.3

Terms and Concepts

1. The dot product of two vectors is a ________, not a vector.

2. How are the concepts of the dot product and vector magnitude related?

3. How can one quickly tell if the angle between two vectors is acute or obtuse?

4. Give a synonym for "orthogonal."

Problems

In Exercises 5 – 10, find the dot product of the given vectors.

5. $\vec{u} = \langle 2, -4 \rangle$, $\vec{v} = \langle 3, 7 \rangle$

6. $\vec{u} = \langle 5, 3 \rangle$, $\vec{v} = \langle 6, 1 \rangle$

7. $\vec{u} = \langle 1, -1, 2 \rangle$, $\vec{v} = \langle 2, 5, 3 \rangle$

8. $\vec{u} = \langle 3, 5, -1 \rangle$, $\vec{v} = \langle 4, -1, 7 \rangle$

9. $\vec{u} = \langle 1, 1 \rangle$, $\vec{v} = \langle 1, 2, 3 \rangle$

10. $\vec{u} = \langle 1, 2, 3 \rangle$, $\vec{v} = \langle 0, 0, 0 \rangle$

11. Create an example of vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^3$ to show that $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$. Then prove that the result holds in general.

12. Create an example of vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$ and scalar c to show that $c(\vec{u} \cdot \vec{v}) = \vec{u} \cdot (c\vec{v})$. Then prove that the result holds in general.

In Exercises 13 – 16, find the measure of the angle between the two vectors in both radians and degrees.

13. $\vec{u} = \langle 1, 1 \rangle$, $\vec{v} = \langle 1, 2 \rangle$

14. $\vec{u} = \langle -2, 1 \rangle$, $\vec{v} = \langle 3, 5 \rangle$

15. $\vec{u} = \langle 8, 1, -4 \rangle$, $\vec{v} = \langle 2, 2, 0 \rangle$

16. $\vec{u} = \langle 1, 7, 2 \rangle$, $\vec{v} = \langle 4, -2, 5 \rangle$

In Exercises 17 – 20, a vector $\vec{v}$ is given. Give two vectors that are orthogonal to $\vec{v}$.

17. $\vec{v} = \langle 4, 7 \rangle$

18. $\vec{v} = \langle -3, 5 \rangle$

19. $\vec{v} = \langle 1, 1, 1 \rangle$

20. $\vec{v} = \langle 1, -2, 3 \rangle$

In Exercises 21 – 26, vectors $\vec{u}$ and $\vec{v}$ are given. Find $\text{proj}_{\vec{v}}\,\vec{u}$, the orthogonal projection of $\vec{u}$ onto $\vec{v}$, and sketch all three vectors on the same axes.

21. $\vec{u} = \langle 1, 2 \rangle$, $\vec{v} = \langle -1, 3 \rangle$

22. $\vec{u} = \langle 5, 5 \rangle$, $\vec{v} = \langle 1, 3 \rangle$

23. $\vec{u} = \langle -3, 2 \rangle$, $\vec{v} = \langle 1, 1 \rangle$

24. $\vec{u} = \langle -3, 2 \rangle$, $\vec{v} = \langle 2, 3 \rangle$

25. $\vec{u} = \langle 1, 5, 1 \rangle$, $\vec{v} = \langle 1, 2, 3 \rangle$

26. $\vec{u} = \langle 3, -1, 2 \rangle$, $\vec{v} = \langle 2, 2, 1 \rangle$

In Exercises 27 – 32, vectors $\vec{u}$ and $\vec{v}$ are given. Write $\vec{u}$ as the sum of two vectors, one of which is parallel to $\vec{v}$ and one of which is perpendicular to $\vec{v}$. Note: these are the same pairs of vectors as found in Exercises 21 – 26.

27. $\vec{u} = \langle 1, 2 \rangle$, $\vec{v} = \langle -1, 3 \rangle$

28. $\vec{u} = \langle 5, 5 \rangle$, $\vec{v} = \langle 1, 3 \rangle$

29. $\vec{u} = \langle -3, 2 \rangle$, $\vec{v} = \langle 1, 1 \rangle$

30. $\vec{u} = \langle -3, 2 \rangle$, $\vec{v} = \langle 2, 3 \rangle$

31. $\vec{u} = \langle 1, 5, 1 \rangle$, $\vec{v} = \langle 1, 2, 3 \rangle$

32. $\vec{u} = \langle 3, -1, 2 \rangle$, $\vec{v} = \langle 2, 2, 1 \rangle$

33. A 10lb box sits on a ramp that rises 4ft over a distance of 20ft. How much force is required to keep the box from sliding down the ramp?

34. A 10lb box sits on a 15ft ramp that makes a 30° angle with the horizontal. How much force is required to keep the box from sliding down the ramp?

35. How much work is performed in moving a box horizontally 10ft with a force of 20lb applied at an angle of 45° to the horizontal?

36. How much work is performed in moving a box horizontally 10ft with a force of 20lb applied at an angle of 10° to the horizontal?

37. How much work is performed in moving a box up the length of a ramp that rises 2ft over a distance of 10ft, with a force of 50lb applied horizontally?

38. How much work is performed in moving a box up the length of a ramp that rises 2ft over a distance of 10ft, with a force of 50lb applied at an angle of 45° to the horizontal?

39. How much work is performed in moving a box up the length of a 10ft ramp that makes a 5° angle with the horizontal, with 50lb of force applied in the direction of the ramp?

2.4 The Cross Product

"Orthogonality" is immensely important. A quick scan of your current environment will undoubtedly reveal numerous surfaces and edges that are perpendicular to each other (including the edges of this page). The dot product provides a quick test for orthogonality: vectors $\vec{u}$ and $\vec{v}$ are perpendicular if, and only if, $\vec{u} \cdot \vec{v} = 0$.

Given two non–parallel, nonzero vectors $\vec{u}$ and $\vec{v}$ in space, it is very useful to find a vector $\vec{w}$ that is perpendicular to both $\vec{u}$ and $\vec{v}$. There is a operation, called the **cross product**, that creates such a vector. This section defines the cross product, then explores its properties and applications.

Definition 20 Cross Product

Let $\vec{u} = \langle u_1, u_2, u_3\rangle$ and $\vec{v} = \langle v_1, v_2, v_3\rangle$ be vectors in $\mathbb{R}^3$. The **cross product of $\vec{u}$ and $\vec{v}$**, denoted $\vec{u} \times \vec{v}$, is the vector

$$\vec{u} \times \vec{v} = \langle u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1\rangle .$$

The definition of the cross product may look strange (and complicated) at first, but it's more or less forced by the requirement that it be orthogonal to both $\vec{u}$ and $\vec{v}$. To begin to see why, suppose $\vec{w} = \langle a, b, c\rangle$ is an arbitrary vector such that $\vec{w} \cdot \vec{u} = 0$ and $\vec{w} \cdot \vec{v} = 0$. This gives us the pair of equations

$$u_1a + u_2b + u_3c = 0$$
$$v_1a + v_2b + v_3c = 0.$$

This is a *system of linear equations* in the variables a, b, and c. We'll learn the techniques for solving any such system in Chapter 4, at which point we'll be able to see that (up to a scalar multiple) the solution is given by Definition 20.

This definition can be a bit cumbersome to remember. After an example we will give a convenient method for computing the cross product. For now, careful examination of the products and differences given in the definition should reveal a pattern that is not too difficult to remember. (For instance, in the first component only 2 and 3 appear as subscripts; in the second component, only 1 and 3 appear as subscripts. Further study reveals the order in which they appear.)

Let's practice using this definition by computing a cross product.

Example 25 Computing a cross product

Let $\vec{u} = \langle 2, -1, 4\rangle$ and $\vec{v} = \langle 3, 2, 5\rangle$. Find $\vec{u} \times \vec{v}$, and verify that it is orthogonal to both $\vec{u}$ and $\vec{v}$.

Solution Using Definition 20, we have

$$\begin{aligned} \vec{u} \times \vec{v} &= \langle u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1\rangle \\ &= \langle (-1)5 - (4)2, (4)3 - (2)5, (2)2 - (-1)3\rangle = \langle -13, 2, 7\rangle . \end{aligned}$$

(We encourage the reader to compute this product on their own, then verify their result.)

We test whether or not $\vec{u} \times \vec{v}$ is orthogonal to $\vec{u}$ and $\vec{v}$ using the dot product:

$$\begin{aligned} (\vec{u} \times \vec{v}) \cdot \vec{u} &= \langle -13, 2, 7\rangle \cdot \langle 2, -1, 4\rangle = 0, \\ (\vec{u} \times \vec{v}) \cdot \vec{v} &= \langle -13, 2, 7\rangle \cdot \langle 3, 2, 5\rangle = 0. \end{aligned}$$

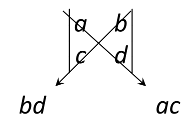
Since both dot products are zero, $\vec{u} \times \vec{v}$ is indeed orthogonal to both $\vec{u}$ and $\vec{v}$.

We now introduce a method for computing the cross-product that is easier to remember, and has the added benefit of allowing us to preview **determinants**, which we will return to in earnest in Section 6.3.

Consider a rectangular array $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ of four real numbers a, b, c, and d. A 2×2 determinant takes any such array and assigns the number $ad - bc$. This is commonly denoted as follows:

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

Most people find it easiest to remember this in terms of the two *diagonals* of the array: we take the product of the two numbers on the *main diagonal* (top-left to bottom-right), and subtract the product of the two numbers on the other diagonal:



For example, we have $\begin{vmatrix} 4 & -2 \\ 6 & 3 \end{vmatrix} = 4(3) - (-2)(6) = 24$. Once we get comfortable with 2×2 determinants, we can write the cross product in terms of them, as follows:

$$\begin{aligned} \vec{u} \times \vec{v} &= \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} \vec{i} - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} \vec{j} + \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \vec{k} \\ &= (u_2 v_3 - u_3 v_2)\vec{i} - (u_3 v_1 - u_1 v_3)\vec{j} + (u_1 v_2 - u_2 v_1)\vec{k}, \end{aligned} \tag{2.4}$$

as before. Now, this might not seem like much of an improvement over the previous formula, so we take things one step further. First, we form a 3×3 array as shown below.

$$\begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}.$$

The first row comprises the standard unit vectors $\vec{i}, \vec{j}$, and $\vec{k}$. The second and third rows are the vectors $\vec{u}$ and $\vec{v}$, respectively. Next, we *expand* our 3×3 array as a vector, where the coefficient of each standard unit vector is given by the 2×2 determinant that's left over when we delete the row and column containing that unit vector.

For example, if we use $\vec{u}$ and $\vec{v}$ from Example 25, we obtain the array

$$\begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 2 & -1 & 4 \\ 3 & 2 & 5 \end{vmatrix}.$$

The expansion process used to obtain the coefficients of $\vec{i}, \vec{j}\vec{k}$ looks like the following:

$$\begin{vmatrix} \boxed{\vec{i}} & \vec{j} & \vec{k} \\ 2 & -1 & 4 \\ 3 & 2 & 5 \end{vmatrix} \longrightarrow \begin{vmatrix} -1 & 4 \\ 2 & 5 \end{vmatrix} \vec{i} = -13\vec{i}$$

$$\begin{vmatrix} \vec{i} & \boxed{\vec{j}} & \vec{k} \\ 2 & -1 & 4 \\ 3 & 2 & 5 \end{vmatrix} \longrightarrow \begin{vmatrix} 2 & 4 \\ 3 & 5 \end{vmatrix} \vec{j} = -2\vec{j}$$

$$\begin{vmatrix} \vec{i} & \vec{j} & \boxed{\vec{k}} \\ 2 & -1 & 4 \\ 3 & 2 & 5 \end{vmatrix} \longrightarrow \begin{vmatrix} 2 & -1 \\ 3 & 2 \end{vmatrix} \vec{k} = 7\vec{k}$$

There is one more important detail to note: notice in Equation (2.4) that there is a **minus sign** in front of the coefficient of the unit vector $\vec{j}$. We need to make sure that the signs in front of each 2×2 determinant follow this $+, -, +$ pattern when we expand our array as a vector. For the vectors $\vec{u}$ and $\vec{v}$ in Example 25, we end up with the following:

$$\begin{aligned}\vec{u} \times \vec{v} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 2 & -1 & 4 \\ 3 & 2 & 5 \end{vmatrix} = \begin{vmatrix} -1 & 4 \\ 2 & 5 \end{vmatrix} \vec{i} - \begin{vmatrix} 2 & 4 \\ 3 & 5 \end{vmatrix} \vec{j} + \begin{vmatrix} 2 & -1 \\ 3 & 2 \end{vmatrix} \vec{k} \\ &= -13\vec{i} - (-2)\vec{j} + 7\vec{k} = \langle -13, 2, 7 \rangle ,\end{aligned}$$

as before. The method will become more clear with a bit of practice.

Note: If the minus sign in front of the $\vec{j}$ coefficient seems out of place to you, it might help to imagine wrapping our 3×3 array around a cylinder (like the label on a tin can). If we read from left to right, *beginning in the $\vec{j}$ column*, then we should place the $\vec{k}$ column first, followed by the $\vec{i}$ column. For the vectors $\vec{u}$ and $\vec{v}$ in Example 25, this would result in the coefficient $\begin{vmatrix} 4 & 2 \\ 5 & 2 \end{vmatrix} = 2$ for the $\vec{j}$ component, which has the correct sign. However, since our habit is to read starting from the far left, we tend to write the $\vec{i}$ column first, and then introduce the minus sign to compensate.

Example 26 Computing a cross product

Let $\vec{u} = \langle 1, 3, 6 \rangle$ and $\vec{v} = \langle -1, 2, 1 \rangle$. Compute both $\vec{u} \times \vec{v}$ and $\vec{v} \times \vec{u}$.

Solution To compute $\vec{u} \times \vec{v}$, we form our 3×3 array as prescribed above, and expand it into a vector:

$$\begin{aligned}\vec{u} \times \vec{v} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & 3 & 6 \\ -1 & 2 & 1 \end{vmatrix} = \begin{vmatrix} 3 & 6 \\ 2 & 1 \end{vmatrix} \vec{i} - \begin{vmatrix} 1 & 6 \\ -1 & 1 \end{vmatrix} \vec{j} + \begin{vmatrix} 1 & 3 \\ -1 & 2 \end{vmatrix} \vec{k} \\ &= (3(1) - 6(2))\vec{i} - (1(1) - 6(-1))\vec{j} + (1(2) - 3(-1))\vec{k} \\ &= -9\vec{i} - 7\vec{j} + 5\vec{k} = \langle -9, -7, 5 \rangle .\end{aligned}$$

To compute $\vec{v} \times \vec{u}$, we switch the second and third rows of the above matrix, then expand as before:

$$\begin{aligned}\vec{v} \times \vec{u} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ -1 & 2 & 1 \\ 1 & 3 & 6 \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 3 & 6 \end{vmatrix} \vec{i} - \begin{vmatrix} -1 & 1 \\ 1 & 6 \end{vmatrix} \vec{j} + \begin{vmatrix} -1 & 2 \\ 1 & 3 \end{vmatrix} \vec{k} \\ &= (2(6) - 1(3))\vec{i} - ((-1)(6) - 1(1))\vec{j} + ((-1)(3) - 2(1))\vec{k} \\ &= 9\vec{i} + 7\vec{j} - 5\vec{k} = \langle 9, 7, -5 \rangle = -\vec{u} \times \vec{v}.\end{aligned}$$

Note how with the rows being switched, the products that once appeared on the right now appear on the left, and vice–versa, so that the result is the opposite of $\vec{u} \times \vec{v}$. We leave it to the reader to verify that each of these vectors is orthogonal to $\vec{u}$ and $\vec{v}$.

Properties of the Cross Product

It is not coincidence that $\vec{v} \times \vec{u} = -(\vec{u} \times \vec{v})$ in the preceding example; one can show using Definition 20 that this will always be the case. The following theorem states several useful properties of the cross product, each of which can be verified by referring to the definition.

Theorem 6 Properties of the Cross Product

Let $\vec{u}$, $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^3$ and let c be a scalar. The following identities hold:

1. $\vec{u} \times \vec{v} = -(\vec{v} \times \vec{u})$ — Anticommutative Property
2. (a) $(\vec{u} + \vec{v}) \times \vec{w} = \vec{u} \times \vec{w} + \vec{v} \times \vec{w}$ — Distributive Properties
   (b) $\vec{u} \times (\vec{v} + \vec{w}) = \vec{u} \times \vec{v} + \vec{u} \times \vec{w}$
3. $c(\vec{u} \times \vec{v}) = (c\vec{u}) \times \vec{v} = \vec{u} \times (c\vec{v})$
4. (a) $(\vec{u} \times \vec{v}) \cdot \vec{u} = 0$ — Orthogonality Properties
   (b) $(\vec{u} \times \vec{v}) \cdot \vec{v} = 0$
5. $\vec{u} \times \vec{u} = \vec{0}$
6. $\vec{u} \times \vec{0} = \vec{0}$
7. $\vec{u} \cdot (\vec{v} \times \vec{w}) = (\vec{u} \times \vec{v}) \cdot \vec{w}$ — Scalar Triple Product

We introduced the cross product as a way to find a vector orthogonal to two given vectors, but we did not give a proof that the construction given in Definition 20 satisfies this property. Theorem 6 asserts this property holds; we leave it as a problem in the Exercise section to verify this.

The algebraic properties of the cross product in Theorem 6 also give us an additional method for computing the cross product in terms of the unit vectors $\vec{i}, \vec{j}, \vec{k}$. We know from Property 5 that

$$\vec{i} \times \vec{i} = \vec{0}, \vec{j} \times \vec{j} = \vec{0}, \vec{k} \times \vec{k} = \vec{0},$$

and it's easy to check that

$$\vec{i} \times \vec{j} = \vec{k}, \vec{j} \times \vec{k} = \vec{i}, \vec{k} \times \vec{i} = \vec{j},$$

and then Property 1 guarantees that

$$\vec{j} \times \vec{i} = -\vec{k}, \vec{k} \times \vec{j} = -\vec{i}, \vec{i} \times \vec{k} = -\vec{j}.$$

Using Properties 2 and 3, we can then compute, for example,

$$\begin{aligned}\langle 2, 0, 3\rangle \times \langle -1, 4, 2\rangle &= (2\vec{i} + 3\vec{k}) \times (-\vec{i} + 4\vec{j} + 2\vec{k}) \\ &= -2(\vec{i} \times \vec{i}) + 8(\vec{i} \times \vec{j}) + 4(\vec{i} \times \vec{k}) \\ &\quad - 3(\vec{k} \times \vec{i}) + 12(\vec{k} \times \vec{j}) + 6(\vec{k} \times \vec{k}) \\ &= \vec{0} + 8\vec{k} - 4\vec{j} - 3\vec{j} - 12\vec{i} + \vec{0} = \langle -12, -7, 8\rangle .\end{aligned}$$

Property 5 from the theorem is also left to the reader to prove in the Exercise section, but it reveals something more interesting than "the cross product of a vector with itself is $\vec{0}$." Let $\vec{u}$ and $\vec{v}$ be parallel vectors; that is, let there be a scalar c such that $\vec{v} = c\vec{u}$. Consider their cross product:

$$\begin{aligned}\vec{u} \times \vec{v} &= \vec{u} \times (c\vec{u}) \\ &= c(\vec{u} \times \vec{u}) && \text{(by Property 3 of Theorem 6)} \\ &= \vec{0}. && \text{(by Property 5 of Theorem 6)}\end{aligned}$$

We have just shown that the cross product of parallel vectors is $\vec{0}$. This hints at something deeper. Theorem 5 related the angle between two vectors and their dot product; there is a similar relationship relating the cross product of two vectors and the angle between them, given by the following theorem.

Theorem 7 The Cross Product and Angles

Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^3$. Then

$$\parallel \vec{u} \times \vec{v} \parallel = \parallel \vec{u} \parallel \parallel \vec{v} \parallel \sin\theta,$$

where θ, $0 \leq \theta \leq \pi$, is the angle between $\vec{u}$ and $\vec{v}$.

Note that this theorem makes a statement about the *magnitude* of the cross product. When the angle between $\vec{u}$ and $\vec{v}$ is 0 or π (i.e., the vectors are parallel), the magnitude of the cross product is 0. The only vector with a magnitude of 0 is $\vec{0}$ (see Property 9 of Theorem 3), hence the cross product of parallel vectors is $\vec{0}$.

We provide some anecdotal evidence of the truth of this theorem in the following example.

Note: Definition 17 (through Theorem 5) defines $\vec{u}$ and $\vec{v}$ to be orthogonal if $\vec{u} \cdot \vec{v} = 0$. We could use Theorem 7 to define $\vec{u}$ and $\vec{v}$ are parallel if $\vec{u} \times \vec{v} = 0$. By such a definition, $\vec{0}$ would be both orthogonal and parallel to every vector. Apparent paradoxes such as this are not uncommon in mathematics and can be very useful. (See also the first marginal note on page 29.)

Example 27 The cross product and angles

Let $\vec{u} = \langle 1, 3, 6 \rangle$ and $\vec{v} = \langle -1, 2, 1 \rangle$ as in Example 26. Verify Theorem 7 by finding θ, the angle between $\vec{u}$ and $\vec{v}$, and the magnitude of $\vec{u} \times \vec{v}$.

SOLUTION We use Theorem 5 to find the angle between $\vec{u}$ and $\vec{v}$.

$$\begin{aligned} \theta &= \cos^{-1}\left(\frac{\vec{u} \cdot \vec{v}}{\parallel \vec{u} \parallel \parallel \vec{v} \parallel}\right) \\ &= \cos^{-1}\left(\frac{11}{\sqrt{46}\sqrt{6}}\right) \\ &\approx 0.8471 = 48.54^\circ. \end{aligned}$$

Our work in Example 26 showed that $\vec{u} \times \vec{v} = \langle -9, -7, 5 \rangle$, hence $\parallel \vec{u} \times \vec{v} \parallel = \sqrt{155}$. Is $\parallel \vec{u} \times \vec{v} \parallel = \parallel \vec{u} \parallel \parallel \vec{v} \parallel \sin\theta$? Using numerical approximations, we find:

$$\begin{aligned} \parallel \vec{u} \times \vec{v} \parallel &= \sqrt{155} \\ &\approx 12.45. \end{aligned} \qquad \begin{aligned} \parallel \vec{u} \parallel \parallel \vec{v} \parallel \sin\theta &= \sqrt{46}\sqrt{6}\sin 0.8471 \\ &\approx 12.45. \end{aligned}$$

Numerically, they seem equal. Using a right triangle, one can show that

$$\sin\left(\cos^{-1}\left(\frac{11}{\sqrt{46}\sqrt{6}}\right)\right) = \frac{\sqrt{155}}{\sqrt{46}\sqrt{6}},$$

which allows us to verify the theorem exactly.

To see that Theorem 7 holds in general, let $\vec{u} = \langle u_1, u_2, u_3 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3 \rangle$ be two arbitrary vectors in $\mathbb{R}^3$. Since the angle between $\vec{u}$ and $\vec{v}$ is defined to lie between 0 and π, we know that $\sin\theta \geq 0$, so that both sides of the equation $\|\vec{u} \times \vec{v}\| = \|\vec{u}\| \, \|\vec{v}\| \sin\theta$ are positive. Thus, we can show that both sides are equal if we can show that their squares are equal. We have

$$\begin{aligned}
(\|\vec u\|\,\|\vec v\|\sin\theta)^2 &= \|\vec u\|^2\,\|\vec v\|^2\sin^2\theta \\
&= \|\vec u\|^2\,\|\vec v\|^2\,(1-\cos^2\theta) && \text{since } \sin^2\theta+\cos^2\theta=1 \\
&= \|\vec u\|^2\,\|\vec v\|^2 - (\|\vec u\|\,\|\vec v\|\cos\theta)^2 \\
&= \|\vec u\|^2\,\|\vec v\|^2 - (\vec u\cdot\vec v)^2 && \text{by Theorem 5} \\
&= (u_1^2+u_2^2+u_3^2)(v_1^2+v_2^2+v_3^2) - (u_1v_1+u_2v_2+u_3v_3)^2 \\
&= u_2^2v_3^2 - 2u_2u_3v_2v_3 + u_3^2v_2^2 + u_1v_3^2 - 2u_1u_3v_1v_3 \\
&\qquad + u_3^2v_1^2 + u_1^2v_2^2 - 2u_1u_2v_1v_2 + u_2^2v_1^2 \\
&= (u_2v_3-u_3v_2)^2 + (u_3v_1-u_1v_3)^2 + (u_1v_2-u_2v_2)^2 \\
&= \|\vec u\times\vec v\|^2,
\end{aligned}$$

as required.

Right Hand Rule

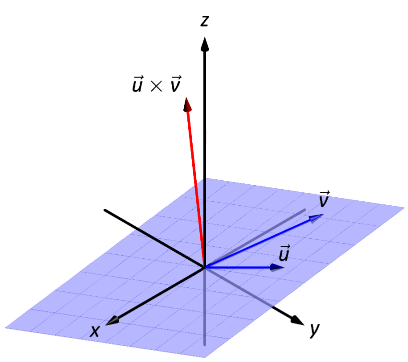


Figure 2.29: Illustrating the Right Hand Rule of the cross product.

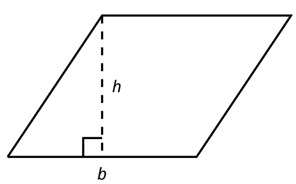
The anticommutative property of the cross product demonstrates that $\vec u\times\vec v$ and $\vec v\times\vec u$ differ only by a sign – these vectors have the same magnitude but point in the opposite direction. When seeking a vector perpendicular to $\vec u$ and $\vec v$, we essentially have two directions to choose from, one in the direction of $\vec u\times\vec v$ and one in the direction of $\vec v\times\vec u$. Does it matter which we choose? How can we tell which one we will get without graphing, etc.?

Another wonderful property of the cross product, as defined, is that it follows the **right hand rule.** Given $\vec u$ and $\vec v$ in $\mathbb{R}^3$ with the same initial point, point the index finger of your right hand in the direction of $\vec u$ and let your middle finger point in the direction of $\vec v$ (much as we did when establishing the right hand rule for the 3-dimensional coordinate system). Your thumb will naturally extend in the direction of $\vec u\times\vec v$. One can "practice" this using Figure 2.29. If you switch, and point the index finder in the direction of $\vec v$ and the middle finger in the direction of $\vec u$, your thumb will now point in the opposite direction, allowing you to "visualize" the anticommutative property of the cross product.

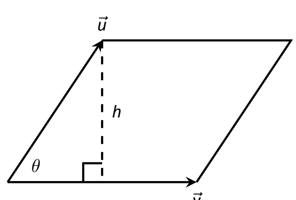
Applications of the Cross Product

There are a number of ways in which the cross product is useful in mathematics, physics and other areas of science beyond "just" finding a vector perpendicular to two others. We highlight a few here.

Area of a Parallelogram



(a)



(b)

Figure 2.30: Using the cross product to find the area of a parallelogram.

It is a standard geometry fact that the area of a parallelogram is $A = bh$, where b is the length of the base and h is the height of the parallelogram, as illustrated in Figure 2.30(a). As shown when defining the Parallelogram Law of vector addition, two vectors $\vec u$ and $\vec v$ define a parallelogram when drawn from the same initial point, as illustrated in Figure 2.30(b). Trigonometry tells us that $h = \|\,\vec u\,\|\sin\theta$, hence the area of the parallelogram is

$$A = \|\,\vec u\,\|\,\|\,\vec v\,\|\sin\theta = \|\,\vec u\times\vec v\,\|, \tag{2.5}$$

where the second equality comes from Theorem 7. We illustrate using Equation (2.5) in the following example.

Example 28 **Finding the area of a parallelogram**

1. Find the area of the parallelogram defined by the vectors $\vec{u} = \langle 2, 1\rangle$ and $\vec{v} = \langle 1, 3\rangle$.

2. Verify that the points $A = (1, 1, 1)$, $B = (2, 3, 2)$, $C = (4, 5, 3)$ and $D = (3, 3, 2)$ are the vertices of a parallelogram. Find the area of the parallelogram.

Solution

1. Figure 2.31(a) sketches the parallelogram defined by the vectors $\vec{u}$ and $\vec{v}$. We have a slight problem in that our vectors exist in $\mathbb{R}^2$, not $\mathbb{R}^3$, and the cross product is only defined on vectors in $\mathbb{R}^3$. We skirt this issue by viewing $\vec{u}$ and $\vec{v}$ as vectors in the $x-y$ plane of $\mathbb{R}^3$, and rewrite them as $\vec{u} = \langle 2, 1, 0\rangle$ and $\vec{v} = \langle 1, 3, 0\rangle$. We can now compute the cross product. It is easy to show that $\vec{u}\times\vec{v} = \langle 0, 0, 5\rangle$; therefore the area of the parallelogram is $A = \parallel \vec{u}\times\vec{v} \parallel = 5$.

2. To show that the quadrilateral $ABCD$ is a parallelogram (shown in Figure 2.31(b)), we need to show that the opposite sides are parallel. We can quickly show that $\overrightarrow{AB} = \overrightarrow{DC} = \langle 1, 2, 1\rangle$ and $\overrightarrow{BC} = \overrightarrow{AD} = \langle 2, 2, 1\rangle$. We find the area by computing the magnitude of the cross product of $\overrightarrow{AB}$ and $\overrightarrow{BC}$:

$$\overrightarrow{AB}\times\overrightarrow{BC} = \langle 0, 1, -2\rangle \quad \Rightarrow \quad \parallel \overrightarrow{AB}\times\overrightarrow{BC} \parallel = \sqrt{5} \approx 2.236.$$

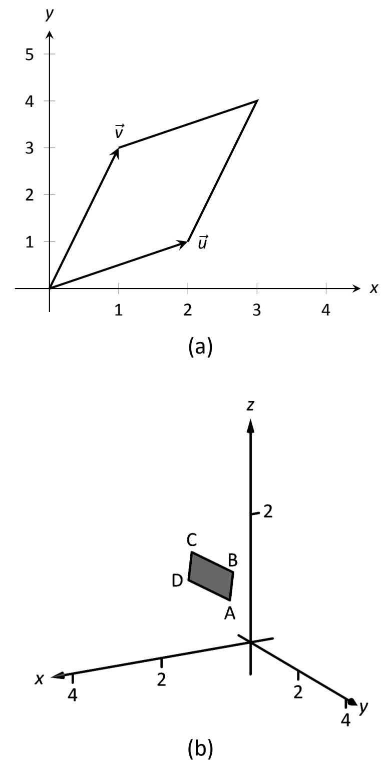


Figure 2.31: Sketching the parallelograms in Example 28.

This application is perhaps more useful in finding the area of a triangle (in short, triangles are used more often than parallelograms). We illustrate this in the following example.

Example 29 **Area of a triangle**

Find the area of the triangle with vertices $A = (1, 2)$, $B = (2, 3)$ and $C = (3, 1)$, as pictured in Figure 2.32.

Solution We found the area of this triangle in Example 202 to be 1.5 using integration. There we discussed the fact that finding the area of a triangle can be inconvenient using the "$\frac{1}{2}bh$" formula as one has to compute the height, which generally involves finding angles, etc. Using a cross product is much more direct.

We can choose any two sides of the triangle to use to form vectors; we choose $\overrightarrow{AB} = \langle 1, 1\rangle$ and $\overrightarrow{AC} = \langle 2, -1\rangle$. As in the previous example, we will rewrite these vectors with a third component of 0 so that we can apply the cross product. The area of the triangle is

$$\frac{1}{2}\parallel \overrightarrow{AB}\times\overrightarrow{AC} \parallel = \frac{1}{2}\parallel \langle 1, 1, 0\rangle \times \langle 2, -1, 0\rangle \parallel = \frac{1}{2}\parallel \langle 0, 0, -3\rangle \parallel = \frac{3}{2}.$$

We arrive at the same answer as before with less work.

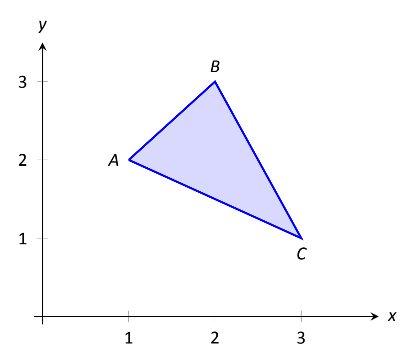


Figure 2.32: Finding the area of a triangle in Example 29.

Volume of a Parallelepiped

The three dimensional analogue to the parallelogram is the **parallelepiped**. Each face is parallel to the face opposite face, as illustrated in Figure 2.33. The volume of any three-dimensional solid whose cross-sectional area is a constant is given by $V = B \cdot h$, where B is the area of the base (the constant cross-sectional area), and h is the height. To determine a formula for the volume, we refer to Figure 2.34. By crossing $\vec{v}$ and $\vec{w}$, one gets a vector whose magnitude is the area of the base, and whose direction is perpendicular to the parallelogram forming the base of the solid. We can then see that the height of the parallelepiped is equal to the length of the projection of the vector $\vec{u}$ onto $\vec{v}\times\vec{w}$. Thus, our volume is given by:

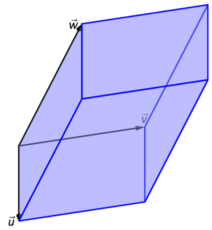


Figure 2.33: A parallelepiped is the three dimensional analogue to the parallelogram.

$$
\begin{aligned}
V &= B \cdot h \\
&= \|\vec{v}\times\vec{w}\| \cdot \left\|\text{proj}_{\vec{v}\times\vec{w}}\,\vec{u}\right\| \\
&= \|\vec{v}\times\vec{w}\| \cdot \left\| \left(\frac{\vec{u}\cdot(\vec{v}\times\vec{w})}{\|\vec{v}\times\vec{w}\|^2}\right)(\vec{v}\times\vec{w}) \right\| \\
&= \|\vec{v}\times\vec{w}\| \,\frac{|\vec{u}\cdot(\vec{v}\times\vec{w})|}{\|\vec{v}\times\vec{w}\|^2}\, \|\vec{v}\times\vec{w}\| \\
&= |\vec{u}\cdot(\vec{v}\times\vec{w})| .
\end{aligned}
$$

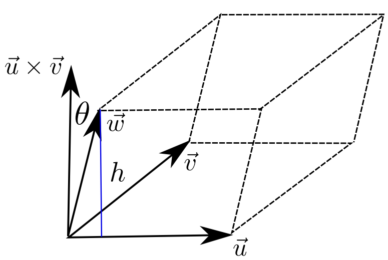


Figure 2.34: Determining the volume of a parallelepiped

Thus the volume of a parallelepiped defined by vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ is

$$V = |\vec{u}\cdot(\vec{v}\times\vec{w})| . \tag{2.6}$$

Note how this is the Scalar Triple Product, first seen in Theorem 6. Applying the identities given in the theorem shows that we can apply the Scalar Triple Product in any "order" we choose to find the volume. That is,

$$V = |\vec{u}\cdot(\vec{v}\times\vec{w})| = |\vec{u}\cdot(\vec{w}\times\vec{v})| = |(\vec{u}\times\vec{v})\cdot\vec{w}|, \quad \text{etc.}$$

Example 30 Finding the volume of parallelepiped

Find the volume of the parallelepiped defined by the vectors $\vec{u} = \langle 1,1,0\rangle$, $\vec{v} = \langle -1,1,0\rangle$ and $\vec{w} = \langle 0,1,1\rangle$.

Solution We apply Equation (2.6). We first find $\vec{v}\times\vec{w} = \langle 1,1,-1\rangle$. Then

$$|\vec{u}\cdot(\vec{v}\times\vec{w})| = |\,\langle 1,1,0\rangle\cdot\langle 1,1,-1\rangle\,| = 2.$$

So the volume of the parallelepiped is 2 cubic units.

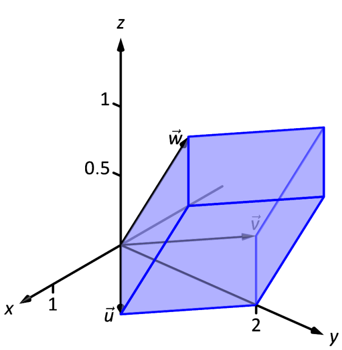


Figure 2.35: A parallelepiped in Example 30.

Let's take another look at how Equation (2.6) is computed in terms of our formulas for the dot and cross products. With $\vec{u} = \langle u_1,u_2,u_3\rangle$, $\vec{v} = \langle v_1,v_2,v_3\rangle$, and $\vec{w} = \langle w_1,w_2,w_3\rangle$, we have

$$
\begin{aligned}
\vec{u}\cdot(\vec{v}\times\vec{w}) &= \langle u_1,u_2,u_3\rangle\cdot\left\langle \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3\end{vmatrix}, -\begin{vmatrix} v_1 & v_3 \\ w_1 & w_3\end{vmatrix}, \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2\end{vmatrix}\right\rangle \\
&= u_1\begin{vmatrix} v_2 & v_3 \\ w_2 & w_3\end{vmatrix} - u_2\begin{vmatrix} v_1 & v_3 \\ w_1 & w_3\end{vmatrix} + u_3\begin{vmatrix} v_1 & v_2 \\ w_1 & w_2\end{vmatrix}.
\end{aligned}
$$

Compare this with our determinant formula for computing the cross product,

$$\vec{v}\times\vec{w} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3\end{vmatrix} = \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3\end{vmatrix}\vec{i} - \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3\end{vmatrix}\vec{j} + \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2\end{vmatrix}\vec{k}.$$

If we replace the unit vectors $\vec{i}, \vec{j}, veck$ in the above equation with the components of $\vec{u}$, we arrive at our first instance of a 3×3 **determinant**, along with a method for computing such an object:

$$\begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = u_1 \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} - u_2 \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} + u_3 \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix} = \vec{u} \cdot (\vec{v} \times \vec{w}).$$

We will return to our study of determinants in Section 6.3, where we will learn techniques for efficiently computing determinants of any size.

While this application of the Scalar Triple Product is interesting, it is not used all that often: parallelepipeds are not a common shape in physics and engineering. (It is, however, essential to understanding the change of variables formula for multiple integrals in Calculus.) The last application of the cross product is very applicable in engineering.

Torque

Torque is a measure of the turning force applied to an object. A classic scenario involving torque is the application of a wrench to a bolt. When a force is applied to the wrench, the bolt turns. When we represent the force and wrench with vectors $\vec{F}$ and $\vec{\ell}$, we see that the bolt moves (because of the threads) in a direction orthogonal to $\vec{F}$ and $\vec{\ell}$. Torque is usually represented by the Greek letter τ, or tau, and has units of N·m, a Newton–meter, or ft·lb, a foot–pound.

While a full understanding of torque is beyond the purposes of this book, when a force $\vec{F}$ is applied to a lever arm $\vec{\ell}$, the resulting torque is

$$\vec{\tau} = \vec{\ell} \times \vec{F}. \tag{2.7}$$

Example 31 **Computing torque**

A lever of length 2ft makes an angle with the horizontal of 45°. Find the resulting torque when a force of 10lb is applied to the end of the level where:

1. the force is perpendicular to the lever, and
2. the force makes an angle of 60° with the lever, as shown in Figure 2.36.

Solution

1. We start by determining vectors for the force and lever arm. Since the lever arm makes a 45° angle with the horizontal and is 2ft long, we can state that $\vec{\ell} = 2\langle \cos 45^\circ, \sin 45^\circ \rangle = \left\langle \sqrt{2}, \sqrt{2} \right\rangle$.

   Since the force vector is perpendicular to the lever arm (as seen in the left hand side of Figure 2.36), we can conclude it is making an angle of $-45°$ with the horizontal. As it has a magnitude of 10lb, we can state $\vec{F} = 10\langle \cos(-45^\circ), \sin(-45^\circ) \rangle = \left\langle 5\sqrt{2}, -5\sqrt{2} \right\rangle$.

   Using Equation (2.7) to find the torque requires a cross product. We again let the third component of each vector be 0 and compute the cross product:

$$\begin{aligned} \vec{\tau} &= \vec{\ell} \times \vec{F} \\ &= \left\langle \sqrt{2}, \sqrt{2}, 0 \right\rangle \times \left\langle 5\sqrt{2}, -5\sqrt{2}, 0 \right\rangle \\ &= \langle 0, 0, -20 \rangle \end{aligned}$$

This clearly has a magnitude of 20 ft-lb.

We can view the force and lever arm vectors as lying "on the page"; our computation of $\vec{\tau}$ shows that the torque goes "into the page." This follows the Right Hand Rule of the cross product, and it also matches well with the example of the wrench turning the bolt. Turning a bolt clockwise moves it in.

2. Our lever arm can still be represented by $\vec{\ell} = \left\langle \sqrt{2}, \sqrt{2} \right\rangle$. As our force vector makes a 60° angle with $\vec{\ell}$, we can see (referencing the right hand side of the figure) that $\vec{F}$ makes a -15° angle with the horizontal. Thus

$$\begin{aligned}\vec{F} = 10 \left\langle \cos -15^\circ, \sin -15^\circ \right\rangle &= \left\langle \frac{5(1+\sqrt{3})}{\sqrt{2}}, -\frac{5(1+\sqrt{3})}{\sqrt{2}} \right\rangle \\ &\approx \left\langle 9.659, -2.588 \right\rangle.\end{aligned}$$

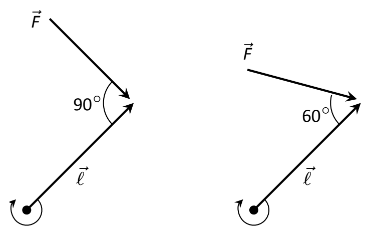


Figure 2.36: Showing a force being applied to a lever in Example 31.

We again make the third component 0 and take the cross product to find the torque:

$$\begin{aligned}\vec{\tau} &= \vec{\ell} \times \vec{F} \\ &= \left\langle \sqrt{2}, \sqrt{2}, 0 \right\rangle \times \left\langle \frac{5(1+\sqrt{3})}{\sqrt{2}}, -\frac{5(1+\sqrt{3})}{\sqrt{2}}, 0 \right\rangle \\ &= \left\langle 0, 0, -10\sqrt{3} \right\rangle \\ &\approx \left\langle 0, 0, -17.321 \right\rangle.\end{aligned}$$

As one might expect, when the force and lever arm vectors *are* orthogonal, the magnitude of force is greater than when the vectors *are not* orthogonal.

While the cross product has a variety of applications (as noted in this chapter), its fundamental use is finding a vector perpendicular to two others. Knowing a vector is orthogonal to two others is of incredible importance, as it allows us to find the equations of lines and planes in a variety of contexts. The importance of the cross product, in some sense, relies on the importance of lines and planes, which see widespread use throughout engineering, physics and mathematics. We study lines and planes in the next two sections.

Exercises 2.4

Terms and Concepts

1. The cross product of two vectors is a ________, not a scalar.

2. One can visualize the direction of $\vec{u} \times \vec{v}$ using the ________ ________ ________.

3. Give a synonym for "orthogonal."

4. T/F: A fundamental principle of the cross product is that $\vec{u} \times \vec{v}$ is orthogonal to $\vec{u}$ and $\vec{v}$.

5. ________ is a measure of the turning force applied to an object.

Problems

In Exercises 6 – 14, vectors $\vec{u}$ and $\vec{v}$ are given. Compute $\vec{u} \times \vec{v}$ and show this is orthogonal to both $\vec{u}$ and $\vec{v}$.

6. $\vec{u} = \langle 3, 2, -2 \rangle, \quad \vec{v} = \langle 0, 1, 5 \rangle$

7. $\vec{u} = \langle 5, -4, 3 \rangle, \quad \vec{v} = \langle 2, -5, 1 \rangle$

8. $\vec{u} = \langle 4, -5, -5 \rangle, \quad \vec{v} = \langle 3, 3, 4 \rangle$

9. $\vec{u} = \langle -4, 7, -10 \rangle, \quad \vec{v} = \langle 4, 4, 1 \rangle$

10. $\vec{u} = \langle 1, 0, 1 \rangle, \quad \vec{v} = \langle 5, 0, 7 \rangle$

11. $\vec{u} = \langle 1, 5, -4 \rangle, \quad \vec{v} = \langle -2, -10, 8 \rangle$

12. $\vec{u} = \vec{i}, \quad \vec{v} = \vec{j}$

13. $\vec{u} = \vec{i}, \quad \vec{v} = \vec{k}$

14. $\vec{u} = \vec{j}, \quad \vec{v} = \vec{k}$

15. Pick any vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^3$ and show that $\vec{u} \times (\vec{v} + \vec{w}) = \vec{u} \times \vec{v} + \vec{u} \times \vec{w}$.

16. Pick any vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^3$ and show that $\vec{u} \cdot (\vec{v} \times \vec{w}) = (\vec{u} \times \vec{v}) \cdot \vec{w}$.

In Exercises 17 – 20, the magnitudes of vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^3$ are given, along with the angle θ between them. Use this information to find the magnitude of $\vec{u} \times \vec{v}$.

17. $\| \vec{u} \| = 2, \quad \| \vec{v} \| = 5, \quad \theta = 30^\circ$

18. $\| \vec{u} \| = 3, \quad \| \vec{v} \| = 7, \quad \theta = \pi/2$

19. $\| \vec{u} \| = 3, \quad \| \vec{v} \| = 4, \quad \theta = \pi$

20. $\| \vec{u} \| = 2, \quad \| \vec{v} \| = 5, \quad \theta = 5\pi/6$

In Exercises 21 – 24, find the area of the parallelogram defined by the given vectors.

21. $\vec{u} = \langle 1, 1, 2 \rangle, \quad \vec{v} = \langle 2, 0, 3 \rangle$

22. $\vec{u} = \langle -2, 1, 5 \rangle, \quad \vec{v} = \langle -1, 3, 1 \rangle$

23. $\vec{u} = \langle 1, 2 \rangle, \quad \vec{v} = \langle 2, 1 \rangle$

24. $\vec{u} = \langle 2, 0 \rangle, \quad \vec{v} = \langle 0, 3 \rangle$

In Exercises 25 – 28, find the area of the triangle with the given vertices.

25. Vertices: $(0, 0, 0)$, $(1, 3, -1)$ and $(2, 1, 1)$.

26. Vertices: $(5, 2, -1)$, $(3, 6, 2)$ and $(1, 0, 4)$.

27. Vertices: $(1, 1)$, $(1, 3)$ and $(2, 2)$.

28. Vertices: $(3, 1)$, $(1, 2)$ and $(4, 3)$.

In Exercises 29 – 30, find the area of the quadrilateral with the given vertices. (Hint: break the quadrilateral into 2 triangles.)

29. Vertices: $(0, 0)$, $(1, 2)$, $(3, 0)$ and $(4, 3)$.

30. Vertices: $(0, 0, 0)$, $(2, 1, 1)$, $(-1, 2, -8)$ and $(1, -1, 5)$.

In Exercises 31 – 32, find the volume of the parallelepiped defined by the given vectors.

31. $\vec{u} = \langle 1, 1, 1 \rangle, \quad \vec{v} = \langle 1, 2, 3 \rangle, \quad \vec{w} = \langle 1, 0, 1 \rangle$

32. $\vec{u} = \langle -1, 2, 1 \rangle, \quad \vec{v} = \langle 2, 2, 1 \rangle, \quad \vec{w} = \langle 3, 1, 3 \rangle$

In Exercises 33 – 36, find a unit vector orthogonal to both $\vec{u}$ and $\vec{v}$.

33. $\vec{u} = \langle 1, 1, 1 \rangle, \quad \vec{v} = \langle 2, 0, 1 \rangle$

34. $\vec{u} = \langle 1, -2, 1 \rangle, \quad \vec{v} = \langle 3, 2, 1 \rangle$

35. $\vec{u} = \langle 5, 0, 2 \rangle, \quad \vec{v} = \langle -3, 0, 7 \rangle$

36. $\vec{u} = \langle 1, -2, 1 \rangle, \quad \vec{v} = \langle -2, 4, -2 \rangle$

37. A bicycle rider applies 150lb of force, straight down, onto a pedal that extends 7in horizontally from the crankshaft. Find the magnitude of the torque applied to the crankshaft.

38. A bicycle rider applies 150lb of force, straight down, onto a pedal that extends 7in from the crankshaft, making a 30° angle with the horizontal. Find the magnitude of the torque applied to the crankshaft.

39. To turn a stubborn bolt, 80lb of force is applied to a 10in wrench. What is the maximum amount of torque that can be applied to the bolt?

40. To turn a stubborn bolt, 80lb of force is applied to a 10in wrench in a confined space, where the direction of applied force makes a 10° angle with the wrench. How much torque is subsequently applied to the wrench?

41. Show, using the definition of the Cross Product, that $\vec{u}\cdot(\vec{u}\times\vec{v}) = 0$; that is, that $\vec{u}$ is orthogonal to the cross product of $\vec{u}$ and $\vec{v}$.

42. Show, using the definition of the Cross Product, that $\vec{u}\times\vec{u} = \vec{0}$.

2.5 Lines

To find the equation of a line in the *x*-*y* plane, we need two pieces of information: a point and the slope. The slope conveys *direction* information. As vertical lines have an undefined slope, the following statement is more accurate:

> To define a line, one needs a point on the line and the direction of the line.

This holds true for lines in space.

Let P be a point in space, let $\vec{p}$ be the vector with initial point at the origin and terminal point at P (i.e., $\vec{p}$ "points" to P), and let $\vec{d}$ be a vector. Consider the points on the line through P in the direction of $\vec{d}$.

Clearly one point on the line is P; we can say that the *vector* $\vec{p}$ lies at this point on the line. To find another point on the line, we can start at $\vec{p}$ and move in a direction parallel to $\vec{d}$. For instance, starting at $\vec{p}$ and traveling one length of $\vec{d}$ places one at another point on the line. Consider Figure 2.38 where certain points along the line are indicated.

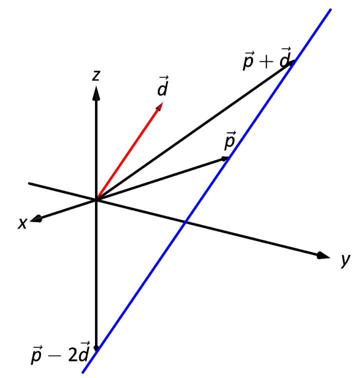


Figure 2.38: Defining a line in space.

The figure illustrates how every point on the line can be obtained by starting with $\vec{p}$ and moving a certain distance in the direction of $\vec{d}$. That is, we can define the line as a function of t:

$$\vec{\ell}(t) = \vec{p} + t\,\vec{d}. \tag{2.8}$$

In many ways, this is *not* a new concept. Compare Equation (2.8) to the familiar "$y = mx + b$" equation of a line:

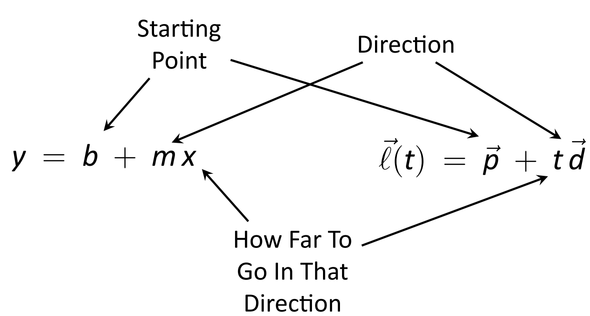


Figure 2.37: Understanding the vector equation of a line.

The equations exhibit the same structure: they give a starting point, define a direction, and state how far in that direction to travel.

Equation (2.8) is an example of a **vector–valued function**; the input of the function is a real number and the output is a vector. We will cover vector–valued functions extensively in the next chapter.

There are other ways to represent a line. Let $\vec{p} = \langle x_0, y_0, z_0\rangle$ and let $\vec{d} = \langle a, b, c\rangle$. Then the equation of the line through $\vec{p}$ in the direction of $\vec{d}$ is:

$$\begin{aligned}\vec{\ell}(t) &= \vec{p} + t\vec{d}\\ &= \langle x_0, y_0, z_0\rangle + t\,\langle a, b, c\rangle\\ &= \langle x_0 + at, y_0 + bt, z_0 + ct\rangle\,.\end{aligned}$$

The last line states the the x values of the line are given by $x = x_0 + at$, the y values are given by $y = y_0 + bt$, and the z values are given by $z = z_0 + ct$. These three equations, taken together, are the **parametric equations of the line** through $\vec{p}$ in the direction of $\vec{d}$.

Finally, each of the equations for x, y and z above contain the variable t. We can solve for t in each equation:

$$\begin{aligned} x = x_0 + at \quad &\Rightarrow \quad t = \frac{x - x_0}{a}, \\ y = y_0 + bt \quad &\Rightarrow \quad t = \frac{y - y_0}{b}, \\ z = z_0 + ct \quad &\Rightarrow \quad t = \frac{z - z_0}{c}, \end{aligned}$$

assuming $a, b, c \neq 0$. Since t is equal to each expression on the right, we can set these equal to each other, forming the **symmetric equations of the line** through $\vec{p}$ in the direction of $\vec{d}$:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$

Each representation has its own advantages, depending on the context. We summarize these three forms in the following definition, then give examples of their use.

Definition 21 Equations of Lines in Space

Consider the line in space that passes through $\vec{p} = \langle x_0, y_0, z_0 \rangle$ in the direction of $\vec{d} = \langle a, b, c \rangle$.

1. The **vector equation** of the line is
$$\vec{\ell}(t) = \vec{p} + t\vec{d}.$$
2. The **parametric equations** of the line are
$$x = x_0 + at, \quad y = y_0 + bt, \quad z = z_0 + ct.$$
3. The **symmetric equations** of the line are
$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$

Example 32 Finding the equation of a line

Give all three equations, as given in Definition 21, of the line through $P = (2, 3, 1)$ in the direction of $\vec{d} = \langle -1, 1, 2 \rangle$. Does the point $Q = (-1, 6, 6)$ lie on this line?

Solution We identify the point $P = (2, 3, 1)$ with the vector $\vec{p} = \langle 2, 3, 1 \rangle$. Following the definition, we have

- the vector equation of the line is $\vec{\ell}(t) = \langle 2, 3, 1 \rangle + t \langle -1, 1, 2 \rangle$;
- the parametric equations of the line are
$$x = 2 - t, \quad y = 3 + t, \quad z = 1 + 2t; \text{ and}$$
- the symmetric equations of the line are
$$\frac{x - 2}{-1} = \frac{y - 3}{1} = \frac{z - 1}{2}.$$

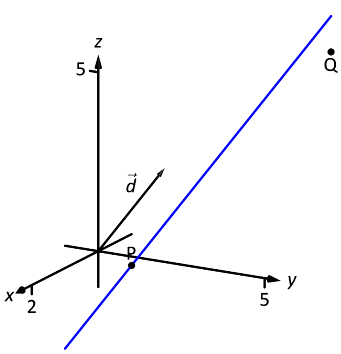


Figure 2.39: Graphing a line in Example 32.

The first two equations of the line are useful when a t value is given: one can immediately find the corresponding point on the line. These forms are good when calculating with a computer; most software programs easily handle equations in these formats. (For instance, to make Figure 2.39, a certain graphics program was given the input `(2-x,3+x,1+2*x)`. This particular program requires the variable always be "x" instead of "t").

Does the point $Q = (-1, 6, 6)$ lie on the line? The graph in Figure 2.39 makes it clear that it does not. We can answer this question without the graph using any of the three equation forms. Of the three, the symmetric equations are probably best suited for this task. Simply plug in the values of x, y and z and see if equality is maintained:

$$\frac{-1-2}{-1} \stackrel{?}{=} \frac{6-3}{1} \stackrel{?}{=} \frac{6-1}{2} \quad \Rightarrow \quad 3 = 3 \neq 2.5.$$

We see that Q does not lie on the line as it did not satisfy the symmetric equations.

Example 33 Finding the equation of a line through two points

Find the parametric equations of the line through the points $P = (2, -1, 2)$ and $Q = (1, 3, -1)$.

Solution Recall the statement made at the beginning of this section: to find the equation of a line, we need a point and a direction. We have *two* points; either one will suffice. The direction of the line can be found by the vector with initial point P and terminal point Q: $\overrightarrow{PQ} = \langle -1, 4, -3 \rangle$.

The parametric equations of the line ℓ through P in the direction of $\overrightarrow{PQ}$ are:

$$\ell : \quad x = 2 - t \quad y = -1 + 4t \quad z = 2 - 3t.$$

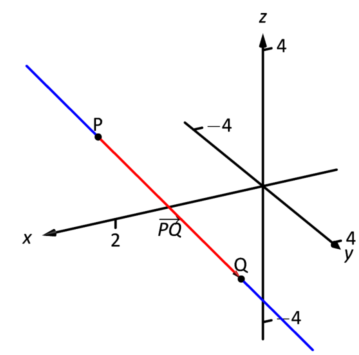


Figure 2.40: A graph of the line in Example 33.

A graph of the points and line are given in Figure 2.40. Note how in the given parametrization of the line, $t = 0$ corresponds to the point P, and $t = 1$ corresponds to the point Q. This relates to the understanding of the vector equation of a line described in Figure 2.37. The parametric equations "start" at the point P, and t determines how far in the direction of $\overrightarrow{PQ}$ to travel. When $t = 0$, we travel 0 lengths of $\overrightarrow{PQ}$; when $t = 1$, we travel one length of $\overrightarrow{PQ}$, resulting in the point Q.

Parallel, Intersecting and Skew Lines

In the plane, two *distinct* lines can either be parallel or they will intersect at exactly one point. In space, given equations of two lines, it can sometimes be difficult to tell whether the lines are distinct or not (i.e., the same line can be represented in different ways). Given lines $\vec{\ell}_1(t) = \vec{p}_1 + t\vec{d}_1$ and $\vec{\ell}_2(t) = \vec{p}_2 + t\vec{d}_2$, we have four possibilities: $\vec{\ell}_1$ and $\vec{\ell}_2$ are

the same line	they share all points;
intersecting lines	share only 1 point;
parallel lines	$\vec{d}_1 \parallel \vec{d}_2$, no points in common; or
skew lines	$\vec{d}_1 \nparallel \vec{d}_2$, no points in common.

The next two examples investigate these possibilities.

Example 34 Comparing lines

Consider lines ℓ_1 and ℓ_2, given in parametric equation form:

$$\ell_1: \begin{array}{rcl} x &=& 1+3t \\ y &=& 2-t \\ z &=& t \end{array} \qquad \ell_2: \begin{array}{rcl} x &=& -2+4s \\ y &=& 3+s \\ z &=& 5+2s. \end{array}$$

Determine whether ℓ_1 and ℓ_2 are the same line, intersect, are parallel, or skew.

Solution We start by looking at the directions of each line. Line ℓ_1 has the direction given by $\vec{d}_1 = \langle 3, -1, 1\rangle$ and line ℓ_2 has the direction given by $\vec{d}_2 = \langle 4, 1, 2\rangle$. It should be clear that $\vec{d}_1$ and $\vec{d}_2$ are not parallel, hence ℓ_1 and ℓ_2 are not the same line, nor are they parallel. Figure 2.41 verifies this fact (where the points and directions indicated by the equations of each line are identified).

We next check to see if they intersect (if they do not, they are skew lines). To find if they intersect, we look for t and s values such that the respective x, y and z values are the same. That is, we want s and t such that:

$$\begin{array}{rcccl} 1+3t &=& x &=& -2+4s \\ 2-t &=& y &=& 3+s \\ t &=& z &=& 5+2s. \end{array}$$

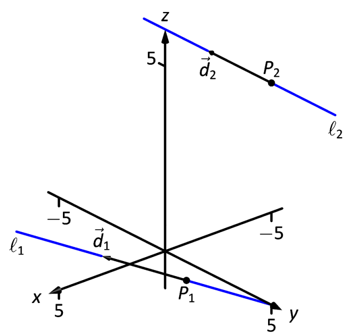


Figure 2.41: Sketching the lines from Example 34.

This is a relatively simple system of linear equations. Since the last equation is already solved for t, substitute that value of t into the equation above it:

$$2-(5+2s) = 3+s \quad \Rightarrow \quad s=-2,\ t=1.$$

A key to remember is that we have *three* equations; we need to check if $s = -2,\ t = 1$ satisfies the first equation as well:

$$1+3(1) \neq -2+4(-2).$$

It does not. Therefore, we conclude that the lines ℓ_1 and ℓ_2 are skew.

We say that a system of equations with no solution, such as the one in Example 34, is *inconsistent*. Although it is possible to find values that work for any two of the three equations, there is no set of values for s and t that work for all three equations simultaneously. We'll develop general techniques for studying systems of linear equations in Chapter 4.

Example 35 Comparing lines

Consider the lines ℓ_1 and ℓ_2 given by the vector equations

$$\vec{\ell_1}(s) = \langle 2,-1,4\rangle + s\langle 0,4,-8\rangle$$
$$\vec{\ell_2}(t) = \langle -3,4,-6\rangle + t\langle 2,-1,2\rangle.$$

Determine if the lines are parallel, skew, or intersecting.

Solution We can immediately see that the lines cannot be parallel, since the x-component of the direction vector for ℓ_1 is zero, but this is not the case for the direction vector of ℓ_2. (There is no scalar c such that $c(0) = 2$.) To determine if the lines intersect, we proceed as in the previous example. We must have

$$\begin{array}{rcccl} 2 &=& x &=& -3+2t \\ -1+4s &=& y &=& 4-t \\ 4-8s &=& z &=& -6+2t. \end{array}$$

The first equation immediately gives us $2t = 5$, so $t = \frac{5}{2}$. Plugging this into the second equation gives us

$$4s = 4 - \frac{5}{2} + 1 = \frac{5}{2} \quad \Rightarrow \quad s = \frac{5}{8}.$$

We now need to check to see if these values satisfy the third equation as well: we have

$$4 - 8s = 4 - 5 = -1,$$

and

$$-6 + 2t = -6 + 5 = -1,$$

so the values $s = \frac{5}{8}$, $t = \frac{5}{2}$ work for all three equations, and since

$$\vec{\ell}_1\left(\frac{5}{8}\right) = \langle 2, -1, 4\rangle + \frac{5}{8}\langle 0, 4, -8\rangle = \langle 2, \frac{3}{2}, -1\rangle \quad \text{and}$$

$$\vec{\ell}_2\left(\frac{5}{2}\right) = \langle -3, 4, -6\rangle + \frac{5}{2}\langle 2, -1, 2\rangle = \langle 2, \frac{3}{2}, -1\rangle,$$

our point of intersection is $(2, \frac{3}{2}, -1)$.

Example 36 Comparing lines

Consider lines ℓ_1 and ℓ_2, given in parametric equation form:

$$\ell_1: \begin{array}{rcl} x &=& -0.7 + 1.6t \\ y &=& 4.2 + 2.72t \\ z &=& 2.3 - 3.36t \end{array} \qquad \ell_2: \begin{array}{rcl} x &=& 2.8 - 2.9s \\ y &=& 10.15 - 4.93s \\ z &=& -5.05 + 6.09s. \end{array}$$

Determine whether ℓ_1 and ℓ_2 are the same line, intersect, are parallel, or skew.

Solution It is obviously very difficult to simply look at these equations and discern anything. This is done intentionally. In the "real world," most equations that are used do not have nice, integer coefficients. Rather, there are lots of digits after the decimal and the equations can look "messy."

We again start by deciding whether or not each line has the same direction. The direction of ℓ_1 is given by $\vec{d}_1 = \langle 1.6, 2.72, -3.36\rangle$ and the direction of ℓ_2 is given by $\vec{d}_2 = \langle -2.9, -4.93, 6.09\rangle$. When it is not clear through observation whether two vectors are parallel or not, the standard way of determining this is by comparing their respective unit vectors. Using a calculator, we find:

$$\vec{u}_1 = \frac{\vec{d}_1}{\parallel \vec{d}_1 \parallel} = \langle 0.3471, 0.5901, -0.7289\rangle$$

$$\vec{u}_2 = \frac{\vec{d}_2}{\parallel \vec{d}_2 \parallel} = \langle -0.3471, -0.5901, 0.7289\rangle\,.$$

The two vectors seem to be parallel (at least, their components are equal to 4 decimal places). In most situations, it would suffice to conclude that the lines are at least parallel, if not the same. One way to be sure is to rewrite $\vec{d}_1$ and $\vec{d}_2$ in terms of fractions, not decimals. We have

$$\vec{d}_1 = \left\langle \frac{16}{10}, \frac{272}{100}, -\frac{336}{100}\right\rangle \qquad \vec{d}_2 = \left\langle -\frac{29}{10}, -\frac{493}{100}, \frac{609}{100}\right\rangle.$$

One can then find the magnitudes of each vector in terms of fractions, then compute the unit vectors likewise. After a lot of manual arithmetic (or after briefly using a computer algebra system), one finds that

$$\vec{u}_1 = \left\langle \sqrt{\frac{10}{83}}, \frac{17}{\sqrt{830}}, -\frac{21}{\sqrt{830}}\right\rangle \qquad \vec{u}_2 = \left\langle -\sqrt{\frac{10}{83}}, -\frac{17}{\sqrt{830}}, \frac{21}{\sqrt{830}}\right\rangle.$$

We can now say without equivocation that these lines are parallel.

Are they the same line? The parametric equations for a line describe one point that lies on the line, so we know that the point $P_1 = (-0.7, 4.2, 2.3)$ lies on ℓ_1. To determine if this point also lies on ℓ_2, plug in the x, y and z values of P_1 into the symmetric equations for ℓ_2:

$$\begin{array}{ccccc} \frac{(-0.7)-2.8}{-2.9} & \stackrel{?}{=} & \frac{(4.2)-10.15}{-4.93} & \stackrel{?}{=} & \frac{(2.3)-(-5.05)}{6.09} \\ 1.2069 & = & 1.2069 & = & 1.2069. \end{array}$$

The point P_1 lies on both lines, so we conclude they are the same line, just parametrized differently. Figure 2.42 graphs this line along with the points and vectors described by the parametric equations. Note how $\vec{d}_1$ and $\vec{d}_2$ are parallel, though point in opposite directions (as indicated by their unit vectors above).

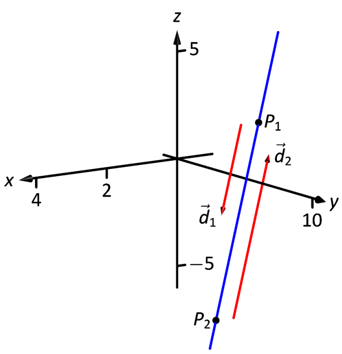


Figure 2.42: Graphing the lines in Example 36.

Distances

Given a point Q and a line $\vec{\ell}(t) = \vec{p} + t\vec{d}$ in space, it is often useful to know the distance from the point to the line. (Here we use the standard definition of "distance," i.e., the length of the shortest line segment from the point to the line.) Identifying $\vec{p}$ with the point P, Figure 2.43 will help establish a general method of computing this distance h.

From trigonometry, we know $h = \parallel \overrightarrow{PQ} \parallel \sin\theta$. We have a similar identity involving the cross product: $\parallel \overrightarrow{PQ} \times \vec{d} \parallel = \parallel \overrightarrow{PQ} \parallel \parallel \vec{d} \parallel \sin\theta$. Divide both sides of this latter equation by $\parallel \vec{d} \parallel$ to obtain h:

$$h = \frac{\parallel \overrightarrow{PQ} \times \vec{d} \parallel}{\parallel \vec{d} \parallel}. \tag{2.9}$$

We put Equation (2.9) to use in the following example.

Example 37 Finding the distance from a point to a line

Find the distance from the point $Q = (1, 1, 3)$ to the line $\vec{\ell}(t) = \langle 1, -1, 1 \rangle + t\,\langle 2, 3, 1 \rangle$.

Solution The equation of the line line gives us the point $P = (1, -1, 1)$ that lies on the line, hence $\overrightarrow{PQ} = \langle 0, 2, 2 \rangle$. The equation also gives $\vec{d} = \langle 2, 3, 1 \rangle$. Using Equation (2.9), we have the distance as

$$\begin{aligned} h &= \frac{\parallel \overrightarrow{PQ} \times \vec{d} \parallel}{\parallel \vec{d} \parallel} \\ &= \frac{\parallel \langle -4, 4, -4 \rangle \parallel}{\sqrt{14}} \\ &= \frac{4\sqrt{3}}{\sqrt{14}}. \end{aligned}$$

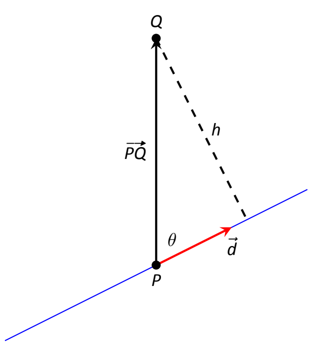


Figure 2.43: Establishing the distance from a point to a line.

While Equation (2.9) gives us a convenient formula for computing the distance, you are probably better off making sure you understand the argument used to obtain the formula. For one thing, a formula is easily forgotten. For another, understanding the method will allow you to adapt it to similar situations still to come, such as computing the distance between skew lines, or from a point to a plane. The general method for these types of problems can be outlined as follows.

Key Idea 9 Steps for solving shortest distance problems

Suppose you are asked to find the distance between two objects, or to determine an object (such as a point) that is closest to a given object (a line or plane). Your solution to the problem should always include the following steps:

1. Make a list of all the information provided in the problem.
2. Make a note of what quantities you're asked to determine.
3. **Draw a diagram**. Label all relevant points and vectors, including those you know, and those you want to find.
4. Using your diagram as a reference, compute any unknown points or vectors.

Note: We can't overemphasize the fact that the diagram referred to in Key Idea 9 **does not have to be accurate** with respect to the coordinates and directions involved. It simply has to be capable of representing the information in the problem. Note that in Figure 2.44 in Example 38 we've drawn a line, some points, and some vectors that represent the problem, without reference to a coordinate system. The goal is to provide enough detail to allow us to set up the problem.

We put the method in Key Idea 9 to use in the following example. Note that in this example we're asked not just for the distance from a point to a line, but also for the point on the line that is *closest* to the given point, so simply using Equation (2.9) is not enough.

Example 38 Finding the closest point on a line

Find the distance from the point $Q = (1, 3, -2)$ to the line $\vec{\ell}$ that passes through the point $P = (2, 0, -1)$ in the direction of $\vec{d} = \langle 1, -1, 0\rangle$, and find the point R on $\vec{\ell}$ that is closest to Q.

Solution We're given a point P on the line, along with a direction vector $\vec{d}$, and a point Q not on the line. We seek the point R on the line that is closest to Q, as well as the distance from Q to R. We begin by diagramming the information in Figure 2.44. From the given points P and Q we can immediately construct the vector

$$\overrightarrow{PQ} = \langle 1-2, 3-0, -2-(-1)\rangle = \langle -1, 3, -1\rangle .$$

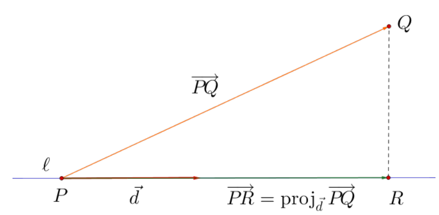


Figure 2.44: Setting up the solution in Example 38

Rather than use Formula (2.9) to find the distance, we begin instead by finding the point R on the line that is closest to Q. From our diagram, we can see that the vector $\overrightarrow{PR}$ from P to R is equal to the projection of $\overrightarrow{PQ}$ onto the distance vector $\vec{d}$:

$$\overrightarrow{PR} = \mathrm{proj}_{\vec{d}}\,\overrightarrow{PQ} = \left(\frac{\langle -1, 3, -1\rangle \cdot \langle 1, -1, 0\rangle}{\langle 1, -1, 0\rangle \cdot \langle 1, -1, 0\rangle}\right)\langle 1, -1, 0\rangle = \langle -2, 2, 0\rangle .$$

Now, we need to pause and take care that we don't make a very common mistake: the vector $\overrightarrow{PR}$ does **not** give the coordinates of the point R. Instead, $\overrightarrow{PR}$ tells us how to get *from* the point P *to* the point R. Letting O denote the origin, we can write $\overrightarrow{OP}$ and $\overrightarrow{OR}$ for the position vectors of P and R, respectively. Since $\overrightarrow{PR} = \overrightarrow{OR} - \overrightarrow{OP}$ using the "tip minus tail" rule for computing the vector between two points, we have

$$\overrightarrow{OR} = \overrightarrow{OP} + \overrightarrow{PR} = \langle 2, 0, -1\rangle + \langle -2, 2, 0\rangle = \langle 0, 2, -1\rangle .$$

Thus, we have $R = (0, 2, -1)$ as the point on the line closest to the point Q. We can now find the distance from Q to the line using the distance formula:

$$D = \sqrt{(1-0)^2 + (3-2)^2 + (-2-(-1))^2} = \sqrt{3}.$$

(You should verify that this agrees with the distance given by Formula (2.9).) An alternative way of computing the distance is to make use of the orthogonal decomposition in Key Idea 8. By definition of the distance from a point to a line, we know that the vector $\overrightarrow{RQ}$ must be orthogonal to the line, and thus to the direction vector $\vec{d}$. Using Key Idea 8, we have that

$$\overrightarrow{RQ} = \overrightarrow{PQ} - \overrightarrow{PR} = \langle -1, 3, -1 \rangle - \langle -2, 2, 0 \rangle = \langle -1, 1, 1 \rangle\,,$$

and the shortest distance is given by $\left\| \overrightarrow{RQ} \right\| = \sqrt{3}$, as before.

It is also useful to determine the distance between lines, which we define as the length of the shortest line segment that connects the two lines. This line segment is necessarily perpendicular to both lines. Let lines $\vec{\ell}_1(t) = \vec{p}_1 + t\vec{d}_1$ and $\vec{\ell}_2(t) = \vec{p}_2 + t\vec{d}_2$ be given, as shown in Figure 2.45. To find the direction orthogonal to both $\vec{d}_1$ and $\vec{d}_2$, we take the cross product: $\vec{c} = \vec{d}_1 \times \vec{d}_2$. The magnitude of the orthogonal projection of $\overrightarrow{P_1P_2}$ onto $\vec{c}$ is the distance h we seek:

$$\begin{aligned} h &= \left\| \operatorname{proj}_{\vec{c}} \overrightarrow{P_1P_2} \right\| = \left\| \frac{\overrightarrow{P_1P_2} \cdot \vec{c}}{\vec{c} \cdot \vec{c}} \vec{c} \right\| \\ &= \frac{|\overrightarrow{P_1P_2} \cdot \vec{c}|}{\|\vec{c}\|^2} \|\vec{c}\| = \frac{\left|\overrightarrow{P_1P_2} \cdot \vec{c}\right|}{\|\vec{c}\|}. \end{aligned} \tag{2.10}$$

A problem in the Exercise section is to show that this distance is 0 when the lines intersect. Note the use of the Triple Scalar Product: $\overrightarrow{P_1P_2} \cdot c = \overrightarrow{P_1P_2} \cdot (\vec{d}_1 \times \vec{d}_2)$.

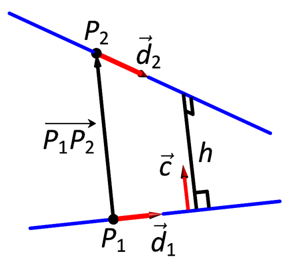


Figure 2.45: Establishing the distance between lines.

Note: Skew lines always lie in parallel *planes*. In the next section we'll see that a plane can be determined by two non-parallel direction vectors and a point on the plane. The distance between the two skew lines is then equal to the distance between the two parallel planes, which is given by the length of a line segment perpendicular to both planes, and therefore, to both lines.

Example 39 Finding the distance between lines

Find the distance between the lines

$$\ell_1 : \begin{array}{rcl} x &=& 1 + 3t \\ y &=& 2 - t \\ z &=& t \end{array} \qquad \ell_2 : \begin{array}{rcl} x &=& -2 + 4s \\ y &=& 3 + s \\ z &=& 5 + 2s. \end{array}$$

Solution These are the sames lines as given in Example 34, where we showed them to be skew. The equations allow us to identify the following points and vectors:

$$\begin{aligned} P_1 = (1, 2, 0) \quad P_2 = (-2, 3, 5) \quad &\Rightarrow \quad \overrightarrow{P_1P_2} = \langle -3, 1, 5 \rangle\,. \\ \vec{d}_1 = \langle 3, -1, 1 \rangle \quad \vec{d}_2 = \langle 4, 1, 2 \rangle \quad &\Rightarrow \quad \vec{c} = \vec{d}_1 \times \vec{d}_2 = \langle -3, -2, 7 \rangle\,. \end{aligned}$$

Using Equation (2.10) we have that the distance h between the two lines is

$$h = \frac{|\overrightarrow{P_1P_2} \cdot \vec{c}|}{\|\vec{c}\|} = \frac{42}{\sqrt{62}}.$$

Once again, we do not recommend attempting to memorize Equation (2.10). Unless you somehow find yourself at a point in your life where you need to find the distances between a whole lot of pairs of skew lines, you will be better served by learning the skills required to set up and think through a problem than you will be by memorizing a formula to plug numbers into. In the case of skew lines, the key observation is that if we take the vector between **any** pair of points, one on each line, and project it onto the vector $\vec{c} = \vec{d}_1 \times \vec{d}_2$, the length of the resulting vector is the distance we seek.

Somewhat more challenging is the problem of finding the points on each line that actually *realize* this shortest distance.

Example 40 Finding the closest points on skew lines

Find the points R_1 on $\vec{\ell}_1$ and R_2 on $\vec{\ell}_2$, where $\vec{\ell}_1$ and $\vec{\ell}_2$ are the lines from Example 39, such that the distance from R_1 to R_2 is a minimum.

Solution Since R_1 is a point on $\vec{\ell}_1$, we know that

$$R_1 = (1+3t, 2-t, t), \quad \text{for some real number } t, \tag{2.11}$$

and similarly,

$$R_2 = (-2+4s, 3+s, 5+2s), \quad \text{for some real number } s. \tag{2.12}$$

The vector $\overrightarrow{R_1R_2}$ is therefore given by

$$\overrightarrow{R_1R_2} = \langle -3+4s-3t, 1+s+t, 5+2s-t \rangle,$$

for some pair of real numbers s and t. We know that the line segment $\overline{R_1R_2}$ must be perpendicular to both $\vec{\ell}_1$ and $\vec{\ell}_2$ in order to minimize the distance, so the vector $\overrightarrow{R_1R_2}$ must be orthogonal to both $\vec{d}_1$ and $\vec{d}_2$. Thus,

$$\begin{aligned}
0 = \vec{d}_1 \cdot \overrightarrow{R_1R_2} &= 3(-3+4s-3t) - 1(1+s+t) + 1(5+2s-t) \\
&= 13s - 11t - 5, \text{ and} \\
0 = \vec{d}_2 \cdot \overrightarrow{R_1R_2} &= 4(-3+4s-3t) + 1(1+s+t) + 2(5+2s-t) \\
&= 21s - 13t - 1.
\end{aligned}$$

We end up having to solve a *system* of two linear equations in the two variables, s and t, given by

$$\begin{array}{rcrcl}
13s & - & 11t & = & 5, \\
21s & - & 13t & = & 1.
\end{array}$$

You probably had to solve such systems in high school. One option is to solve graphically, by plotting the lines given by each equation, and seeing where they intersect. However, this method has little hope of providing an accurate answer. Instead, we try a little algebra. Multiplying the first equation by 21 and the second by 13 gives us the equations $273s - 231t = 105$ and $273s - 169t = 13$, respectively. Subtracting the second equation from the first, we have $-62t = 92$, so $t = -\frac{92}{62} = -\frac{46}{31}$. Plugging this value back into any of the previous equations gives us $s = -\frac{351}{403} = -\frac{27}{31}$. (We didn't promise that the numbers would work out nicely!) Plugging these values back into equations (2.11) and (2.12), we find

$$R_1 = \left(-\frac{107}{31}, \frac{108}{31}, -\frac{46}{31}\right) \quad \text{and} \quad R_2 = \left(-\frac{170}{31}, \frac{66}{31}, \frac{101}{31}\right).$$

Our vector $\overrightarrow{R_1R_2}$ is then given by

$$\overrightarrow{R_1R_2} = \left\langle -\frac{63}{31}, -\frac{42}{31}, \frac{147}{31} \right\rangle = \frac{1}{31}\langle -63, -42, 147 \rangle,$$

and the distance between the two lines is given by

$$\left\| \overrightarrow{R_1R_2} \right\| = \frac{1}{31}\sqrt{63^2 + 42^2 + 147^2} = \frac{42}{\sqrt{31}},$$

as before.

Example 40 required us to solve a system of two linear equations in two unknowns s and t. Although this involved some messy fractions, the algebra involved was fairly straightforward. In many real life problems it is necessary to be able to solve systems involving hundreds or even thousands of equations and variables. We will begin our study of how to systematically solve such systems in the next chapter.

One of the key points to understand from this section is this: to describe a line, we need a point and a direction. Whenever a problem is posed concerning a line, one needs to take whatever information is offered and glean point and direction information. Many questions can be asked (and *are* asked in the Exercise section) whose answer immediately follows from this understanding.

Lines are one of two fundamental objects of study in space. The other fundamental object is the *plane*, which we study in detail in the next section. Many complex three dimensional objects are studied by approximating their surfaces with lines and planes.

Exercises 2.5

Terms and Concepts

1. To find an equation of a line, what two pieces of information are needed?

2. Two distinct lines in the plane can intersect or be ________.

3. Two distinct lines in space can intersect, be ________ or be ________.

4. Use your own words to describe what it means for two lines in space to be skew.

Problems

In Exercises 5 – 14, write the vector, parametric and symmetric equations of the lines described.

5. Passes through $P = (2, -4, 1)$, parallel to $\vec{d} = \langle 9, 2, 5 \rangle$.

6. Passes through $P = (6, 1, 7)$, parallel to $\vec{d} = \langle -3, 2, 5 \rangle$.

7. Passes through $P = (2, 1, 5)$ and $Q = (7, -2, 4)$.

8. Passes through $P = (1, -2, 3)$ and $Q = (5, 5, 5)$.

9. Passes through $P = (0, 1, 2)$ and orthogonal to both $\vec{d}_1 = \langle 2, -1, 7 \rangle$ and $\vec{d}_2 = \langle 7, 1, 3 \rangle$.

10. Passes through $P = (5, 1, 9)$ and orthogonal to both $\vec{d}_1 = \langle 1, 0, 1 \rangle$ and $\vec{d}_2 = \langle 2, 0, 3 \rangle$.

11. Passes through the point of intersection of $\vec{\ell}_1(t)$ and $\vec{\ell}_2(t)$ and orthogonal to both lines, where $\vec{\ell}_1(t) = \langle 2, 1, 1 \rangle + t\langle 5, 1, -2 \rangle$ and $\vec{\ell}_2(t) = \langle -2, -1, 2 \rangle + t\langle 3, 1, -1 \rangle$.

12. Passes through the point of intersection of $\ell_1(t)$ and $\ell_2(t)$ and orthogonal to both lines, where
$$\ell_1 = \begin{cases} x = t \\ y = -2 + 2t \\ z = 1 + t \end{cases} \quad \text{and} \quad \ell_2 = \begin{cases} x = 2 + t \\ y = 2 - t \\ z = 3 + 2t \end{cases}.$$

13. Passes through $P = (1, 1)$, parallel to $\vec{d} = \langle 2, 3 \rangle$.

14. Passes through $P = (-2, 5)$, parallel to $\vec{d} = \langle 0, 1 \rangle$.

In Exercises 15 – 22, determine if the described lines are the same line, parallel lines, intersecting or skew lines. If intersecting, give the point of intersection.

15. $\vec{\ell}_1(t) = \langle 1, 2, 1 \rangle + t\langle 2, -1, 1 \rangle$,
$\vec{\ell}_2(t) = \langle 3, 3, 3 \rangle + t\langle -4, 2, -2 \rangle$.

16. $\vec{\ell}_1(t) = \langle 2, 1, 1 \rangle + t\langle 5, 1, 3 \rangle$,
$\vec{\ell}_2(t) = \langle 14, 5, 9 \rangle + t\langle 1, 1, 1 \rangle$.

17. $\vec{\ell}_1(t) = \langle 3, 4, 1 \rangle + t\langle 2, -3, 4 \rangle$,
$\vec{\ell}_2(t) = \langle -3, 3, -3 \rangle + t\langle 3, -2, 4 \rangle$.

18. $\vec{\ell}_1(t) = \langle 1, 1, 1 \rangle + t\langle 3, 1, 3 \rangle$,
$\vec{\ell}_2(t) = \langle 7, 3, 7 \rangle + t\langle 6, 2, 6 \rangle$.

19. $$\ell_1 = \begin{cases} x = 1 + 2t \\ y = 3 - 2t \\ z = t \end{cases} \quad \text{and} \quad \ell_2 = \begin{cases} x = 3 - t \\ y = 3 + 5t \\ z = 2 + 7t \end{cases}$$

20. $$\ell_1 = \begin{cases} x = 1.1 + 0.6t \\ y = 3.77 + 0.9t \\ z = -2.3 + 1.5t \end{cases} \quad \text{and} \quad \ell_2 = \begin{cases} x = 3.11 + 3.4t \\ y = 2 + 5.1t \\ z = 2.5 + 8.5t \end{cases}$$

21. $$\ell_1 = \begin{cases} x = 0.2 + 0.6t \\ y = 1.33 - 0.45t \\ z = -4.2 + 1.05t \end{cases} \quad \text{and} \quad \ell_2 = \begin{cases} x = 0.86 + 9.2t \\ y = 0.835 - 6.9t \\ z = -3.045 + 16.1t \end{cases}$$

22. $$\ell_1 = \begin{cases} x = 0.1 + 1.1t \\ y = 2.9 - 1.5t \\ z = 3.2 + 1.6t \end{cases} \quad \text{and} \quad \ell_2 = \begin{cases} x = 4 - 2.1t \\ y = 1.8 + 7.2t \\ z = 3.1 + 1.1t \end{cases}$$

In Exercises 23 – 26, find the distance from the point to the line.

23. $P = (1, 1, 1)$, $\quad \vec{\ell}(t) = \langle 2, 1, 3 \rangle + t\langle 2, 1, -2 \rangle$

24. $P = (2, 5, 6)$, $\quad \vec{\ell}(t) = \langle -1, 1, 1 \rangle + t\langle 1, 0, 1 \rangle$

25. $P = (0, 3)$, $\quad \vec{\ell}(t) = \langle 2, 0 \rangle + t\langle 1, 1 \rangle$

26. $P = (1, 1)$, $\quad \vec{\ell}(t) = \langle 4, 5 \rangle + t\langle -4, 3 \rangle$

In Exercises 27 – 28, find the distance between the two lines, and find the points on each line that are closest together.

27. $\vec{\ell}_1(t) = \langle 1, 2, 1 \rangle + t\langle 2, -1, 1 \rangle$,
$\vec{\ell}_2(t) = \langle 3, 3, 3 \rangle + t\langle 4, 2, -2 \rangle$.

28. $\vec{\ell}_1(t) = \langle 0, 0, 1 \rangle + t\langle 1, 0, 0 \rangle$,
$\vec{\ell}_2(t) = \langle 0, 0, 3 \rangle + t\langle 0, 1, 0 \rangle$.

Exercises 29 – 31 explore special cases of the distance formulas (2.9) **and** (2.10)**.**

29. Let Q be a point on the line $\ell(t)$. Show why the distance formula correctly gives the distance from the point to the line as 0.

30. Let lines $\ell_1(t)$ and $\ell_2(t)$ be intersecting lines. Show why the distance formula correctly gives the distance between these lines as 0.

31. Let lines $\ell_1(t)$ and $\ell_2(t)$ be parallel.

    (a) Show why formula (2.10) for the distance between lines cannot be used as stated to find the distance between the lines.

    (b) Show why letting $\vec{c} = (\overrightarrow{P_1P_2} \times \vec{d}_2) \times \vec{d}_2$ allows one to the use the formula.

    (c) Show how one can use the formula for the distance between a point and a line to find the distance between parallel lines.

2.6 Planes

Any flat surface, such as a wall, table top or stiff piece of cardboard can be thought of as representing part of a plane. Consider a piece of cardboard with a point P marked on it. One can take a nail and stick it into the cardboard at P such that the nail is perpendicular to the cardboard; see Figure 2.46

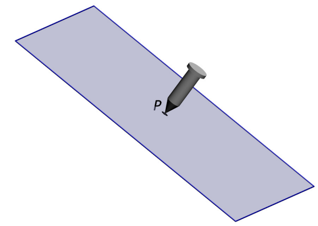


Figure 2.46: Illustrating defining a plane with a sheet of cardboard and a nail.

This nail provides a "handle" for the cardboard. Moving the cardboard around moves P to different locations in space. Tilting the nail (but keeping P fixed) tilts the cardboard. Both moving and tilting the cardboard defines a different plane in space. In fact, we can define a plane by: 1) the location of P in space, and 2) the direction of the nail.

The previous section showed that one can define a line given a point on the line and the direction of the line (usually given by a vector). One can make a similar statement about planes: we can define a plane in space given a point on the plane and the direction the plane "faces" (using the description above, the direction of the nail). Once again, the direction information will be supplied by a vector, called a **normal vector**, that is orthogonal to the plane.

What exactly does "orthogonal to the plane" mean? Choose any two points P and Q in the plane, and consider the vector $\overrightarrow{PQ}$. We say a vector $\vec{n}$ is orthogonal to the plane if $\vec{n}$ is perpendicular to $\overrightarrow{PQ}$ for all choices of P and Q; that is, if $\vec{n}\cdot\overrightarrow{PQ} = 0$ for all P and Q.

This gives us way of writing an equation describing the plane. Let $P = (x_0, y_0, z_0)$ be a point in the plane and let $\vec{n} = \langle a, b, c\rangle$ be a normal vector to the plane. A point $Q = (x, y, z)$ lies in the plane defined by P and $\vec{n}$ if, and only if, $\overrightarrow{PQ}$ is orthogonal to $\vec{n}$. Knowing $\overrightarrow{PQ} = \langle x-x_0, y-y_0, z-z_0\rangle$, consider:

$$\begin{aligned} \overrightarrow{PQ}\cdot\vec{n} &= 0\\ \langle x-x_0, y-y_0, z-z_0\rangle\cdot\langle a, b, c\rangle &= 0\\ a(x-x_0)+b(y-y_0)+c(z-z_0) &= 0 \qquad (2.13)\end{aligned}$$

More algebra produces:

$$ax+by+cz = d, \qquad (2.14)$$

where $d = ax_0+by_0+cz_0$ is a real number. Both of the equations (2.13) or (2.14) are referred to as **scalar equations** for the plane. Note that choosing the numbers a, b, c for the normal vector defines a whole *family* of parallel planes; the value of the constant d determines a particular member of that family.

As long as $c \neq 0$, we can solve for z:

$$z = \frac{1}{c}(d-ax-by). \qquad (2.15)$$

Equation (2.15) is especially useful as many computer programs can graph functions in this form. Equations (2.13) and (2.14) have specific names, given next.

Definition 22 Equations of a Plane in Standard and General Forms

The plane passing through the point $P = (x_0, y_0, z_0)$ with normal vector $\vec{n} = \langle a, b, c\rangle$ can be described by an equation with **standard form**

$$a(x-x_0)+b(y-y_0)+c(z-z_0) = 0;$$

the equation's **general form** is

$$ax+by+cz = d.$$

A key to remember throughout this section is this: to find the equation of a plane, we need a point and a normal vector. We will give several examples of finding the equation of a plane, and in each one different types of information are given. In each case, we need to use the given information to find a point on the plane and a normal vector.

Example 41 **Finding the equation of a plane.**

Write the equation of the plane that passes through the points $P = (1, 1, 0)$, $Q = (1, 2, -1)$ and $R = (0, 1, 2)$ in standard form.

Solution We need a vector $\vec{n}$ that is orthogonal to the plane. Since P, Q and R are in the plane, so are the vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$; $\overrightarrow{PQ} \times \overrightarrow{PR}$ is orthogonal to $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ and hence the plane itself.

It is straightforward to compute $\vec{n} = \overrightarrow{PQ} \times \overrightarrow{PR} = \langle 2, 1, 1\rangle$. We can use any point we wish in the plane (any of P, Q or R will do) and we arbitrarily choose P. Following Definition 22, the equation of the plane in standard form is

$$2(x-1)+(y-1)+z=0.$$

The plane is sketched in Figure 2.47.

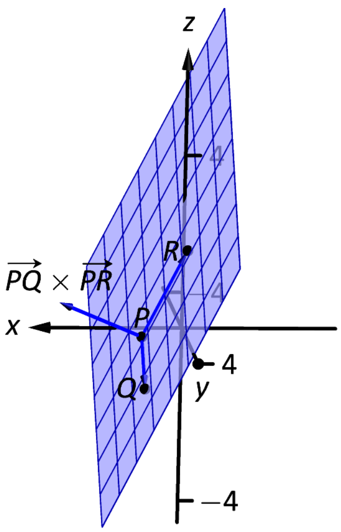


Figure 2.47: Sketching the plane in Example 41.

We have just demonstrated the fact that any three non-collinear points define a plane. (This is why a three-legged stool does not "rock;" it's three feet always lie in a plane. A four-legged stool will rock unless all four feet lie in the same plane.)

Example 42 **Finding the equation of a plane.**

Verify that lines ℓ_1 and ℓ_2, whose parametric equations are given below, intersect, then give the equation of the plane that contains these two lines in general form.

$$\ell_1: \begin{array}{rcl} x &=& -5+2s \\ y &=& 1+s \\ z &=& -4+2s \end{array} \qquad \ell_2: \begin{array}{rcl} x &=& 2+3t \\ y &=& 1-2t \\ z &=& 1+t \end{array}$$

Solution The lines clearly are not parallel. If they do not intersect, they are skew, meaning there is not a plane that contains them both. If they do intersect, there is such a plane.

To find their point of intersection, we set the x, y and z equations equal to each other and solve for s and t:

$$\begin{array}{rcl} -5+2s &=& 2+3t \\ 1+s &=& 1-2t \\ -4+2s &=& 1+t \end{array} \quad \Rightarrow \quad s=2, \quad t=-1.$$

When $s = 2$ and $t = -1$, the lines intersect at the point $P = (-1, 3, 0)$.

Let $\vec{d}_1 = \langle 2, 1, 2\rangle$ and $\vec{d}_2 = \langle 3, -2, 1\rangle$ be the directions of lines ℓ_1 and ℓ_2, respectively. A normal vector to the plane containing these the two lines will also be orthogonal to $\vec{d}_1$ and $\vec{d}_2$. Thus we find a normal vector $\vec{n}$ by computing $\vec{n} = \vec{d}_1 \times \vec{d}_2 = \langle 5, 4-7\rangle$.

We can pick any point in the plane with which to write our equation; each line gives us infinite choices of points. We choose P, the point of intersection.

We follow Definition 22 to write the plane's equation in general form:

$$\begin{aligned} 5(x+1)+4(y-3)-7z &= 0 \\ 5x+5+4y-12-7z &= 0 \\ 5x+4y-7z &= 7. \end{aligned}$$

The plane's equation in general form is $5x+4y-7z=7$; it is sketched in Figure 2.48.

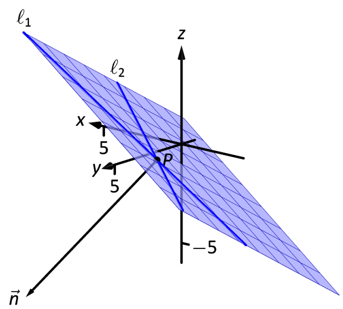


Figure 2.48: Sketching the plane in Example 42.

The two previous examples hint at an alternative method for describing a plane in $\mathbb{R}^3$: instead of providing a single direction orthogonal to the plane (given by the normal vector), we can give two directions that are *parallel* to the plane, such as the vectors $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ in Figure 2.47 of the direction vectors $\vec{d}_1$ and $\vec{d}_2$ to the lines in Figure 2.48. Suppose (x,y,z) is a point on the plane $5x+4y-7z=7$ from Example 42. We can treat the point $(-1,3,0)$ where the lines $\vec{\ell}_1$ and $\vec{\ell}_2$ intersect as our point of reference on the plane. From this point, we can reach the point (x,y,z) by first travelling some distance in the direction of $\vec{d}_1$ (parallel to $\vec{\ell}_1$), and then some distance in the direction of $\vec{d}_2$ (parallel to $\vec{\ell}_2$). We can express this mathematically as follows:

$$\begin{aligned} \langle x,y,z\rangle &= \langle -1,3,0\rangle + s\vec{d}_1 + t\vec{d}_2 \qquad (2.16) \\ &= \langle -1+2s+3t, 3+s-2t, 2s+t\rangle . \end{aligned}$$

We can think of the point $(-1,3,0)$ in Example 42 as defining a point of "origin" on the plane, and, even though they are not perpendicular, we can think of the lines $\vec{\ell}_1$ and $\vec{\ell}_2$ as defining a pair of coordinate axes on the plane. Any other point can be located with respect to these axes. (Any two non-parallel lines define a coordinate system in a plane; perpendicular lines are simply more convenient.)

Equation (2.16) can be viewed as a two-dimensional analogue of the vector equation of a line given in the previous section. It tells us that to get from the origin $(0,0,0)$ to the point (x,y,z) on the plane, we should first travel to the point $(-1,3,0)$ on the plane, and then move parallel to the lines $\vec{\ell}_1$ and $\vec{\ell}_2$ until we reach our point. This vector equation for a plane is not particularly useful in Science or Engineering applications, but it is useful mathematically. In particular, if we wanted to describe a two-dimensional plane in $\mathbb{R}^4$ (or any higher-dimensional space), we would have to resort to this method. (Keeping this method for describing a plane in mind will also help us to access some geometric intuition when we discuss span and linear independence later in the text.)

Example 43 **Finding the equation of a plane**

Give the equation, in standard form, of the plane that passes through the point $P=(-1,0,1)$ and is orthogonal to the line with vector equation $\vec{\ell}(t)=\langle -1,0,1\rangle + t\langle 1,2,2\rangle$.

Solution As the plane is to be orthogonal to the line, the plane must be orthogonal to the direction of the line given by $\vec{d}=\langle 1,2,2\rangle$. We use this as our normal vector. Thus the plane's equation, in standard form, is

$$(x+1)+2y+2(z-1)=0.$$

The line and plane are sketched in Figure 2.49.

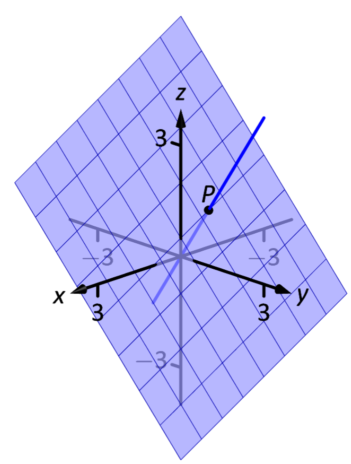


Figure 2.49: The line and plane in Example 43.

Example 44 **Finding the intersection of two planes**

Give the parametric equations of the line that is the intersection of the planes p_1 and p_2, where:

$$\begin{aligned} p_1 &: x-(y-2)+(z-1)=0 \\ p_2 &: -2(x-2)+(y+1)+(z-3)=0 \end{aligned}$$

Solution To find an equation of a line, we need a point on the line and the direction of the line.

We can find a point on the line by solving each equation of the planes for z:

$$\begin{aligned} p_1 &: z = -x + y - 1 \\ p_2 &: z = 2x - y - 2 \end{aligned}$$

We can now set these two equations equal to each other (i.e., we are finding values of x and y where the planes have the same z value):

$$\begin{aligned} -x + y - 1 &= 2x - y - 2 \\ 2y &= 3x - 1 \\ y &= \frac{1}{2}(3x - 1) \end{aligned}$$

We can choose any value for x; we choose $x = 1$. This determines that $y = 1$. We can now use the equations of either plane to find z: when $x = 1$ and $y = 1$, $z = -1$ on both planes. We have found a point P on the line: $P = (1, 1, -1)$.

We now need the direction of the line. Since the line lies in each plane, its direction is orthogonal to a normal vector for each plane. Considering the equations for p_1 and p_2, we can quickly determine their normal vectors. For p_1, $\vec{n}_1 = \langle 1, -1, 1 \rangle$ and for p_2, $\vec{n}_2 = \langle -2, 1, 1 \rangle$. A direction orthogonal to both of these directions is their cross product: $\vec{d} = \vec{n}_1 \times \vec{n}_2 = \langle -2, -3, -1 \rangle$.

The parametric equations of the line through $P = (1, 1, -1)$ in the direction of $d = \langle -2, -3, -1 \rangle$ is:

$$\ell: \quad x = -2t + 1 \quad y = -3t + 1 \quad z = -t - 1.$$

The planes and line are graphed in Figure 2.50.

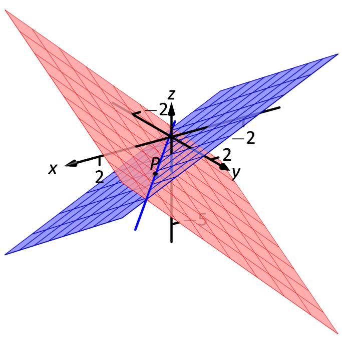


Figure 2.50: Graphing the planes and their line of intersection in Example 44.

In the previous example, note that any point (x, y, z) on the line ℓ is a point on *both* of the planes p_1 and p_2. We can view the pair of equations defining p_1 and p_2 as a *system*

$$\begin{array}{rcrcrcr} x & - & y & + & z & = & -1 \\ -2x & + & y & + & z & = & -2 \end{array}$$

of two linear equations in the three variables x, y, z. Any point (x, y, z) on ℓ belongs to both planes, and thus satisfies both equations; we say that (x, y, z) is a *solution* to the system above. We expect our line to be the set of all such solutions.

Another way to obtain our description of ℓ is as follows. Instead of solving both equations immediately for z, we first use the equation for p_1 to eliminate the x variable from the equation for p_2: multiplying both sides of the first equation by 2, we get $2x - 2y + 2z = -2$. Adding this equation to the equation for p_2, we obtain the equation

$$-y + 3z = -4.$$

Solving for y, we get $y = 3z + 4$; note that we have now solved for y in terms of z only. Note also that any point (x, y, z) that satisfied both of our original equations must also satisfy our new equation. (Make sure you understand why!) If we substitute $y = 3z + 4$ into the equation for p_1, we obtain

$$x - (3z + 4) + z = -1,$$

and solving for x gives $x = 2z - 1$. Summing up, we have

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

This looks a lot like the parametric equations for a line, except that we have the trivial equation $z = z$ at the bottom, and no parameter t. The key here is to realize that because $z = z$ is always satisfied, the variable z is free to take on any value that it wants, so there's no harm in declaring that z **is** the parameter. If we introduce a new variable t and declare that $z = t$, then we obtain the parametric equations

$$\begin{aligned} z &= -1 + 2t \\ Y &= 4 + 3t \\ z &= t, \end{aligned}$$

which are now definitely the parametric equations of a line. In vector form, we have

$$\vec{\ell}(t) = \langle\langle -1, 4, 0\rangle + t\,\langle 2, 3, 1\rangle\,.$$

Is this the same line as before? Our direction vector is note the same as before, but it is parallel: $\langle 2, 3, 1\rangle = -\vec{d}$. Our point on the line is also different. However, note that if we set $t = -1$ in the parametric equations we found in Example 44, we get

$$x = -2 + 1 = -1, y = -3(-1) + 1 = 4, \text{ and } z = -(-1) - 1 = 0,$$

so we have indeed found the same line as before.

Note: Once again, the system of equations we solved in the discussion following Example 44 is exactly the sort of thing we'll learn how to tackle in general in Chapter 4. In this case, we would say that the system has *infinitely many solutions*, or, if we wanted to be more specific, a *one-parameter family* of solutions.

Example 45 **Finding the intersection of a line and a plane**

Find the point of intersection, if any, of the line $\ell(t) = \langle 3, -3, -1\rangle + t\,\langle -1, 2, 1\rangle$ and the plane with equation in general form $2x + y + z = 4$.

Solution The equation of the plane shows that the vector $\vec{n} = \langle 2, 1, 1\rangle$ is a normal vector to the plane, and the equation of the line shows that the line moves parallel to $\vec{d} = \langle -1, 2, 1\rangle$. Since these are not orthogonal, we know there is a point of intersection. (If there were orthogonal, it would mean that the plane and line were parallel to each other, either never intersecting or the line was in the plane itself.)

To find the point of intersection, we need to find a t value such that $\ell(t)$ satisfies the equation of the plane. Rewriting the equation of the line with parametric equations will help:

$$\ell(t) = \begin{cases} x = 3 - t \\ y = -3 + 2t \\ z = -1 + t \end{cases}.$$

Replacing x, y and z in the equation of the plane with the expressions containing t found in the equation of the line allows us to determine a t value that indicates the point of intersection:

$$\begin{aligned} 2x + y + z &= 4 \\ 2(3 - t) + (-3 + 2t) + (-1 + t) &= 4 \\ t &= 2. \end{aligned}$$

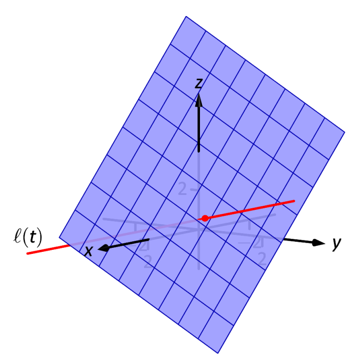


Figure 2.51: Illustrating the intersection of a line and a plane in Example 45.

When $t = 2$, the point on the line satisfies the equation of the plane; that point is $\ell(2) = \langle 1, 1, 1 \rangle$. Thus the point $(1, 1, 1)$ is the point of intersection between the plane and the line, illustrated in Figure 2.51.

Distances

Just as it was useful to find distances between points and lines in the previous section, it is also often necessary to find the distance from a point to a plane.

Consider Figure 2.52, where a plane with normal vector $\vec{n}$ is sketched containing a point P and a point Q, not on the plane, is given. We measure the distance from Q to the plane by measuring the length of the projection of $\overrightarrow{PQ}$ onto $\vec{n}$. That is, we want:

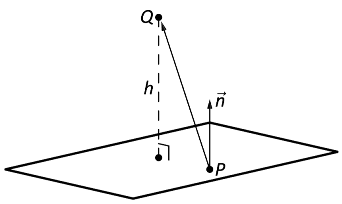


Figure 2.52: Illustrating finding the distance from a point to a plane.

$$\left\| \operatorname{proj}_{\vec{n}} \overrightarrow{PQ} \right\| = \left\| \frac{\vec{n} \cdot \overrightarrow{PQ}}{\| \vec{n} \|^2} \vec{n} \right\| = \frac{|\vec{n} \cdot \overrightarrow{PQ}|}{\| \vec{n} \|} \tag{2.17}$$

Note: Equation (2.17) is useful as it does more than just give the distance between a point and a plane. We will see how it allows us to find several other distances as well: the distance between parallel planes and the distance from a line and a plane.
However, as with the distance problems in the previous section, learning to follow the steps in Key Idea 9 will pay off more in the long run than memorizing a formula. Here, our key steps are to draw a diagram such as Figure 2.52, which doesn't need to be accurate, but does need to contain all the information needed to construct the projection whose length gives us the desired distance.

Example 46 Distance between a point and a plane

Find the distance bewteen the point $Q = (2, 1, 4)$ and the plane with equation $2x - 5y + 6z = 9$.

Solution Referring to Figure 2.52, we need to determine the normal vector $\vec{n}$ and a point P on the plane. Using the equation of the plane, we find the normal vector $\vec{n} = \langle 2, -5, 6 \rangle$. To find a point on the plane, we can let x and y be anything we choose, then let z be whatever satisfies the equation. Letting x and y be 0 seems simple; this makes $z = \frac{3}{2}$. Thus we let $P = \left\langle 0, 0, \frac{3}{2} \right\rangle$, and $\overrightarrow{PQ} = \left\langle 2, 1, \frac{5}{2} \right\rangle$.

We can now compute the projection of $\overrightarrow{PQ}$ onto $\vec{n}$. We have:

$$\begin{aligned}
\operatorname{proj}_{\vec{n}} \overrightarrow{PQ} &= \left(\frac{\overrightarrow{PQ} \cdot \vec{n}}{\|\vec{n}\|^2} \right) \vec{n} \\
&= \left(\frac{\left\langle 2, 1, \frac{5}{2} \right\rangle \cdot \langle 2, -5, 6 \rangle}{(2^2 + 5^2 + 6^2)} \right) \langle 2, -5, 6 \rangle \\
&= \frac{14}{65} \langle 2, -5, 6 \rangle .
\end{aligned}$$

The desired distance is then given by

$$\left\| \operatorname{proj}_{\vec{n}} \overrightarrow{PQ} \right\| = \frac{14}{65} \left\| \langle 2, -5, 6 \rangle \right\| = \frac{14}{\sqrt{65}}.$$

Although it was not requested in Example 46, note that we can also find the point R on the plane that is closest to Q. The desired point must be such that $\overrightarrow{RQ} = \operatorname{proj}_{\vec{n}} \overrightarrow{PQ}$. Since we know the point Q and the vector $\overrightarrow{RQ}$, we can find the point R: since $\overrightarrow{RQ} = \overrightarrow{OQ} - \overrightarrow{OR}$, we find that

$$\begin{aligned}
\overrightarrow{OR} &= \overrightarrow{OQ} - \overrightarrow{RQ} \\
&= \langle 2, 1, 4 \rangle - \frac{14}{65} \langle 2, -5, 6 \rangle \\
&= \frac{1}{65} \langle 102, 135, 176 \rangle .
\end{aligned}$$

The desired point R thus has coordinates $\left(\frac{102}{65}, \frac{135}{65}, \frac{176}{65} \right)$. To make sure that we haven't made any mistakes, let's make sure that this point is indeed on the plane. We have

$$2\left(\frac{102}{65}\right) - 5\left(\frac{135}{65}\right) + 6\left(\frac{176}{65}\right) = \frac{1}{65}(204 - 675 + 1056) = \frac{585}{65} = 9,$$

as expected.

Example 47 **Distance between a line and a plane**
Let ℓ be the line with vector equation

$$\vec{\ell}(t) = \langle 3, 2, -4\rangle + t\langle 3, 1, -1\rangle,$$

and let p be the plane with equation $x - 2y + z = 4$. Verify that the ℓ is parallel to the plane p, and find the distance between them.

Solution From the vector equation for ℓ we have the direction vector $\vec{d} = \langle 3, 1, -1\rangle$, and from the equation for p we can read off the normal vector $\vec{n} = \langle 1, -2, 1\rangle$. Since

$$\vec{d} \cdot \vec{n} = 3(1) + 1(-2) - 1(1) = 0,$$

we know that $\vec{d}$ is orthogonal to $\vec{n}$, and thus ℓ is parallel to p. To find the distance from ℓ to p, we first choose a point on each object. From the vector equation for ℓ we have the point $P = (2, 0, -4)$, and setting $y = z = 0$ in the equation for p, we get $x = 4$ and the point $Q = (4, 0, 0)$.

From these two points we can construct the vector

$$\vec{v} = \overrightarrow{PQ} = \langle 1, -2, 4\rangle$$

which begins on ℓ and ends on p. The distance from ℓ to p is then given by the normal component of $\vec{v}$: we have

$$h = \|\text{proj}_{\vec{n}}\,\vec{v}\| = \frac{|\vec{n} \cdot \vec{v}|}{\|\vec{n}\|} = \frac{9}{6} = \frac{3}{2}.$$

In the previous section we used Equation (2.10) to find the shortest distance between a pair of skew lines. Although we provided some discussion of how this formula was obtained, it's once again the case that memorizing such a formula is not as effective as understanding the process that leads to it. In the next example, we repeat Example 39, but this time we try to understand the problem using planes.

Example 48 **Distance between skew lines**
Find the distance between the skew lines

$$\begin{aligned}\ell_1 &: \langle x, y, z\rangle = \langle 1, 2, 0\rangle + t\langle 3, -1, 1\rangle \\ \ell_2 &: \langle x, y, z\rangle = \langle -2, 3, 5\rangle + t\langle 4, 1, 2\rangle.\end{aligned}$$

Solution We already found the distance between these two lines in Example 39 using Equation (2.10). Supposing that we forgot this formula, how would we proceed? The key is to realize that whenever we have a pair of skew lines, we also have a pair of parallel planes, each of which contains one of the lines. To see this, we first compute the cross product of the direction vectors $\vec{d}_1 = \langle 3, -1, 1\rangle$ and $\vec{d}_2 = \langle 4, 1, 2\rangle$ for the two lines. We find

$$\vec{n} = \vec{d}_1 \times \vec{d}_2 = \langle = \langle -3, -2, 7\rangle.$$

Since $\vec{n}$ is orthogonal to $\vec{d}_1$, the plane through the point $(1,2,0)$ with normal vector $\vec{n}$ contains the line ℓ_1. Similarly, the plane through $(-2,3,5)$ with normal vector $\vec{n}$ contains ℓ_2. We now have our parallel planes.

The next step is to realize that at this point, the problem is no different from the ones we solved in Examples 46: the distance from ℓ_1 to ℓ_2 is the same as the distance between the parallel planes, and the distance between parallel planes is equal to the distance between the first plane, and any point on the second plane.

By definition, the point $P_1 = (1,2,0)$ on ℓ_1 lies on the first plane, and the point $P_2 = (-2,3,5)$ on ℓ_2 lies on the second plane. We compute the vector $\overrightarrow{P_1P_2} = \langle -3,1,5\rangle$, and then find the projection of this vector onto $\vec{n}$, as in Example 46. We have

$$\begin{aligned}
\operatorname{proj}_{\vec{n}} \overrightarrow{P_1P_2} &= \left(\frac{\overrightarrow{P_1P_2}\cdot\vec{n}}{\|\vec{n}\|^2}\right)\vec{n} \\
&= \left(\frac{\langle -3,1,5\rangle\cdot\langle -3,-2,7\rangle}{(3^2+2^2+7^2)}\right)\langle -3,-2,7\rangle \\
&= \frac{42}{62}\langle -3,-2,7\rangle\,.
\end{aligned}$$

The distance is then given by

$$\left\|\operatorname{proj}_{\vec{n}} \overrightarrow{P_1P_2}\right\| = \frac{42}{\sqrt{62}},$$

as before.

These past two sections have not explored lines and planes in space as merely an exercise of mathematical curiosity. There are many, many applications of these fundamental concepts. Complex shapes can be modelled (or, *approximated*) using planes. For instance, part of the exterior of an aircraft may have a complex, yet smooth, shape, and engineers will want to know how air flows across this piece as well as how heat might build up due to air friction. Many equations that help determine air flow and heat dissipation are difficult to apply to arbitrary surfaces, but simple to apply to planes. By approximating a surface with millions of small planes one can more readily model the needed behaviour.

Exercises 2.6

Terms and Concepts

1. In order to find the equation of a plane, what two pieces of information must one have?

2. What is the relationship between a plane and one of its normal vectors?

Problems

In Exercises 3 – 6, give any two points in the given plane.

3. $2x - 4y + 7z = 2$

4. $3(x+2) + 5(y-9) - 4z = 0$

5. $x = 2$

6. $4(y+2) - (z-6) = 0$

In Exercises 7 – 20, give the equation of the described plane in standard and general forms.

7. Passes through $(2, 3, 4)$ and has normal vector $\vec{n} = \langle 3, -1, 7 \rangle$.

8. Passes through $(1, 3, 5)$ and has normal vector $\vec{n} = \langle 0, 2, 4 \rangle$.

9. Passes through the points $(1, 2, 3)$, $(3, -1, 4)$ and $(1, 0, 1)$.

10. Passes through the points $(5, 3, 8)$, $(6, 4, 9)$ and $(3, 3, 3)$.

11. Contains the intersecting lines $\ell_1(t) = \langle 2, 1, 2 \rangle + t\,\langle 1, 2, 3 \rangle$ and $\ell_2(t) = \langle 2, 1, 2 \rangle + t\,\langle 2, 5, 4 \rangle$.

12. Contains the intersecting lines $\ell_1(t) = \langle 5, 0, 3 \rangle + t\,\langle -1, 1, 1 \rangle$ and $\ell_2(t) = \langle 1, 4, 7 \rangle + t\,\langle 3, 0, -3 \rangle$.

13. Contains the parallel lines $\ell_1(t) = \langle 1, 1, 1 \rangle + t\,\langle 1, 2, 3 \rangle$ and $\ell_2(t) = \langle 1, 1, 2 \rangle + t\,\langle 1, 2, 3 \rangle$.

14. Contains the parallel lines $\ell_1(t) = \langle 1, 1, 1 \rangle + t\,\langle 4, 1, 3 \rangle$ and $\ell_2(t) = \langle 2, 2, 2 \rangle + t\,\langle 4, 1, 3 \rangle$.

15. Contains the point $(2, -6, 1)$ and the line
$$\ell(t) = \begin{cases} x = 2 + 5t \\ y = 2 + 2t \\ z = -1 + 2t \end{cases}$$

16. Contains the point $(5, 7, 3)$ and the line
$$\ell(t) = \begin{cases} x = t \\ y = t \\ z = t \end{cases}$$

17. Contains the point $(5, 7, 3)$ and is orthogonal to the line $\ell(t) = \langle 4, 5, 6 \rangle + t\,\langle 1, 1, 1 \rangle$.

18. Contains the point $(4, 1, 1)$ and is orthogonal to the line
$$\ell(t) = \begin{cases} x = 4 + 4t \\ y = 1 + 1t \\ z = 1 + 1t \end{cases}$$

19. Contains the point $(-4, 7, 2)$ and is parallel to the plane $3(x-2) + 8(y+1) - 10z = 0$.

20. Contains the point $(1, 2, 3)$ and is parallel to the plane $x = 5$.

In Exercises 21 – 22, give the equation of the line that is the intersection of the given planes.

21. $p1: \; 3(x-2) + (y-1) + 4z = 0$, and $p2: \; 2(x-1) - 2(y+3) + 6(z-1) = 0$.

22. $p1: \; 5(x-5) + 2(y+2) + 4(z-1) = 0$, and $p2: \; 3x - 4(y-1) + 2(z-1) = 0$.

In Exercises 23 – 26, find the point of intersection between the line and the plane.

23. line: $\langle 5, 1, -1 \rangle + t\,\langle 2, 2, 1 \rangle$, plane: $5x - y - z = -3$

24. line: $\langle 4, 1, 0 \rangle + t\,\langle 1, 0, -1 \rangle$, plane: $3x + y - 2z = 8$

25. line: $\langle 1, 2, 3 \rangle + t\,\langle 3, 5, -1 \rangle$, plane: $3x - 2y - z = 4$

26. line: $\langle 1, 2, 3 \rangle + t\,\langle 3, 5, -1 \rangle$, plane: $3x - 2y - z = -4$

In Exercises 27 – 30, find the given distances.

27. The distance from the point $(1, 2, 3)$ to the plane $3(x-1) + (y-2) + 5(z-2) = 0$.

28. The distance from the point $(2, 6, 2)$ to the plane $2(x-1) - y + 4(z+1) = 0$.

29. The distance between the parallel planes $x + y + z = 0$ and $(x-2) + (y-3) + (z+4) = 0$

30. The distance between the parallel planes $2(x-1)+2(y+1)+(z-2)=0$ and $2(x-3)+2(y-1)+(z-3)=0$

31. Show why if the point Q lies in a plane, then the distance formula correctly gives the distance from the point to the plane as 0.

32. How is Exercise 30 in Section 2.5 easier to answer once we have an understanding of planes?

3: Matrix Arithmetic

A fundamental topic of mathematics is arithmetic; adding, subtracting, multiplying and dividing numbers. After learning how to do this, most of us went on to learn how to add, subtract, multiply and divide "x". We are comfortable with expressions such as

$$x + 3x - x \cdot x^2 + x^5 \cdot x^{-1}$$

and know that we can "simplify" this to

$$4x - x^3 + x^4.$$

This chapter deals with the idea of doing similar operations, but instead of an unknown number x, we will be using a matrix A. So what exactly does the expression

$$A + 3A - A \cdot A^2 + A^5 \cdot A^{-1}$$

mean? We are going to need to learn to define what matrix addition, scalar multiplication, matrix multiplication and matrix inversion are. We will learn just that, plus some more good stuff, in this chapter.

3.1 Matrix Addition and Scalar Multiplication

AS YOU READ . . .

1. When are two matrices equal?
2. Write an explanation of how to add matrices as though writing to someone who knows what a matrix is but not much more.
3. T/F: There is only 1 zero matrix.
4. T/F: To multiply a matrix by 2 means to multiply each entry in the matrix by 2.

In the past, when we dealt with expressions that used "x," we didn't just add and multiply x's together for the fun of it, but rather because we were usually given some sort of *equation* that had x in it and we had to "solve for x."

This begs the question, "What does it mean to be equal?" Two numbers are equal, when, . . ., uh, . . ., never mind. What does it mean for two matrices to be equal? We say that matrices A and B are equal when their corresponding entries are equal. This seems like a very simple definition, but it is rather important, so we give it a box.

Definition 23 Matrix Equality

Two $m \times n$ matrices A and B are *equal* if their corresponding entries are equal.

Notice that our more formal definition specifies that if matrices are equal, they have the same dimensions. This should make sense.

Now we move on to describing how to add two matrices together. To start off, take a wild stab: what do you think the following sum is equal to?

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 2 & -1 \\ 5 & 7 \end{bmatrix} = ?$$

If you guessed

$$\begin{bmatrix} 3 & 1 \\ 8 & 11 \end{bmatrix},$$

you guessed correctly. That wasn't so hard, was it?

Let's keep going, hoping that we are starting to get on a roll. Make another wild guess: what do you think the following expression is equal to?

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

If you guessed

$$\begin{bmatrix} 3 & 6 \\ 9 & 12 \end{bmatrix},$$

you guessed correctly!

Even if you guessed wrong both times, you probably have seen enough in these two examples to have a fair idea now what matrix addition and scalar multiplication are all about.

Before we formally define how to perform the above operations, let us first recall that if A is an $m \times n$ matrix, then we can write A as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Secondly, we should define what we mean by the word *scalar*. A scalar is any number that we multiply a matrix by. (In some sense, we use that number to *scale* the matrix.) We are now ready to define our first arithmetic operations.

Definition 24 Matrix Addition

Let A and B be $m \times n$ matrices. The *sum* of A and B, denoted $A + B$, is

$$\begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}.$$

Definition 25 Scalar Multiplication

Let A be an $m \times n$ matrix and let k be a scalar. The *scalar multiplication* of k and A, denoted kA, is

$$\begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{bmatrix}.$$

We are now ready for an example.

Example 49 Matrix addition and scalar multiplication

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 2 & 1 \\ 5 & 5 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 2 \\ -1 & 0 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 & 3 \\ 9 & 8 & 7 \end{bmatrix}.$$

Simplify the following matrix expressions.

1. $A + B$
2. $B + A$
3. $A - B$
4. $A + C$
5. $-3A + 2B$
6. $A - A$
7. $5A + 5B$
8. $5(A + B)$

Solution

1. $A + B = \begin{bmatrix} 3 & 6 & 9 \\ 0 & 4 & 3 \\ 4 & 5 & 9 \end{bmatrix}.$

2. $B + A = \begin{bmatrix} 3 & 6 & 9 \\ 0 & 4 & 3 \\ 4 & 5 & 9 \end{bmatrix}.$

3. $A - B = \begin{bmatrix} -1 & -2 & -3 \\ -2 & 0 & -1 \\ 6 & 5 & 1 \end{bmatrix}.$

4. $A + C$ is not defined. If we look at our definition of matrix addition, we see that the two matrices need to be the same size. Since A and C have different dimensions, we don't even try to create something as an addition; we simply say that the sum is not defined.

5. $-3A + 2B = \begin{bmatrix} 1 & 2 & 3 \\ 5 & -2 & 1 \\ -17 & -15 & -7 \end{bmatrix}.$

6. $A - A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$

7. Strictly speaking, this is $\begin{bmatrix} 5 & 10 & 15 \\ -5 & 10 & 5 \\ 25 & 25 & 25 \end{bmatrix} + \begin{bmatrix} 10 & 20 & 30 \\ 5 & 10 & 10 \\ -5 & 0 & 20 \end{bmatrix} = \begin{bmatrix} 15 & 30 & 45 \\ 0 & 20 & 15 \\ 20 & 25 & 45 \end{bmatrix}.$

8. Strictly speaking, this is

$$5\left(\begin{bmatrix}1 & 2 & 3\\ -1 & 2 & 1\\ 5 & 5 & 5\end{bmatrix}+\begin{bmatrix}2 & 4 & 6\\ 1 & 2 & 2\\ -1 & 0 & 4\end{bmatrix}\right)=5\cdot\begin{bmatrix}3 & 6 & 9\\ 0 & 4 & 3\\ 4 & 5 & 9\end{bmatrix}$$
$$=\begin{bmatrix}15 & 30 & 45\\ 0 & 20 & 15\\ 20 & 25 & 45\end{bmatrix}.$$

Our example raised a few interesting points. Notice how $A+B=B+A$. We probably aren't surprised by this, since we know that when dealing with numbers, $a+b=b+a$. Also, notice that $5A+5B=5(A+B)$. In our example, we were careful to compute each of these expressions following the proper order of operations; knowing these are equal allows us to compute similar expressions in the most convenient way.

We use the bold face to distinguish the zero matrix, **0**, from the number zero, 0.

Another interesting thing that came from our previous example is that

$$A-A=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}.$$

It seems like this should be a special matrix; after all, every entry is 0 and 0 is a special number.

In fact, this is a special matrix. We define **0**, which we read as "the zero matrix," to be the matrix of all zeros. We should be careful; this previous "definition" is a bit ambiguous, for we have not stated what size the zero matrix should be. Is $\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$ the zero matrix? How about $\begin{bmatrix}0 & 0\end{bmatrix}$?

Let's not get bogged down in semantics. If we ever see **0** in an expression, we will usually know right away what size **0** should be; it will be the size that allows the expression to make sense. If A is a 3×5 matrix, and we write $A+\mathbf{0}$, we'll simply assume that **0** is also a 3×5 matrix. If we are ever in doubt, we can add a subscript; for instance, $\mathbf{0}_{2\times 7}$ is the 2×7 matrix of all zeros.

Since the zero matrix is an important concept, we give it its own definition box.

Definition 26 The Zero Matrix

The $m\times n$ matrix of all zeros, denoted $\mathbf{0}_{m\times n}$, is the *zero matrix*.

When the dimensions of the zero matrix are clear from the context, the subscript is generally omitted.

The following presents some of the properties of matrix addition and scalar multiplication that we discovered above, plus a few more.

Theorem 8 Properties of Matrix Addition and Scalar Multiplication

The following equalities hold for all $m \times n$ matrices A, B and C and scalars k.

1. $A + B = B + A$ (Commutative Property)
2. $(A + B) + C = A + (B + C)$ (Associative Property)
3. $k(A + B) = kA + kB$ (Scalar Multiplication Distributive Property)
4. $kA = Ak$
5. $A + \mathbf{0} = \mathbf{0} + A = A$ (Additive Identity)
6. $0A = \mathbf{0}$

Be sure that this last property makes sense; it says that if we multiply any matrix by the *number* 0, the result is the *zero matrix*, or **0**.

We began this section with the concept of matrix equality. Let's put our matrix addition properties to use and solve a matrix equation.

Example 50 **Solving a matrix equation**
Let

$$A = \begin{bmatrix} 2 & -1 \\ 3 & 6 \end{bmatrix}.$$

Find the matrix X such that

$$2A + 3X = -4A.$$

Solution We can use basic algebra techniques to manipulate this equation for X; first, let's subtract $2A$ from both sides. This gives us

$$3X = -6A.$$

Now divide both sides by 3 to get

$$X = -2A.$$

Now we just need to compute $-2A$; we find that

$$X = \begin{bmatrix} -4 & 2 \\ -6 & -12 \end{bmatrix}.$$

Our matrix properties identified **0** as the Additive Identity; i.e., if you add **0** to any matrix A, you simply get A. This is similar in notion to the fact that for all numbers a, $a + 0 = a$. A *Multiplicative Identity* would be a matrix I where $I \times A = A$ for all matrices A. (What would such a matrix look like? A matrix of all 1s, perhaps?) However, in order for this to make sense, we'll need to learn to multiply matrices together, which we'll do in the next section.

Exercises 3.1

Problems

Matrices A and B are given below. In Exercises 1 – 6, simplify the given expression.

$$A = \begin{bmatrix} 1 & -1 \\ 7 & 4 \end{bmatrix} \quad B = \begin{bmatrix} -3 & 2 \\ 5 & 9 \end{bmatrix}$$

1. $A + B$
2. $2A - 3B$
3. $3A - A$
4. $4B - 2A$
5. $3(A - B) + B$
6. $2(A - B) - (A - 3B)$

Matrices A and B are given below. In Exercises 7 – 10, simplify the given expression.

$$A = \begin{bmatrix} 3 \\ 5 \end{bmatrix} \quad B = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$$

7. $4B - 2A$
8. $-2A + 3A$
9. $-2A - 3A$
10. $-B + 3B - 2B$

Matrices A and B are given below. In Exercises 11 – 14, find X that satisfies the equation.

$$A = \begin{bmatrix} 3 & -1 \\ 2 & 5 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 7 \\ 3 & -4 \end{bmatrix}$$

11. $2A + X = B$
12. $A - X = 3B$
13. $3A + 2X = -1B$
14. $A - \frac{1}{2}X = -B$

In Exercises 15 – 21, find values for the scalars a and b that satisfy the given equation.

15. $a\begin{bmatrix} 1 \\ 2 \end{bmatrix} + b\begin{bmatrix} -1 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \end{bmatrix}$
16. $a\begin{bmatrix} -3 \\ 1 \end{bmatrix} + b\begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$
17. $a\begin{bmatrix} 4 \\ -2 \end{bmatrix} + b\begin{bmatrix} -6 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ -5 \end{bmatrix}$
18. $a\begin{bmatrix} 1 \\ 1 \end{bmatrix} + b\begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$
19. $a\begin{bmatrix} 1 \\ 3 \end{bmatrix} + b\begin{bmatrix} -3 \\ -9 \end{bmatrix} = \begin{bmatrix} 4 \\ -12 \end{bmatrix}$
20. $a\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + b\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}$
21. $a\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + b\begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix}$

3.2 Matrix Multiplication

AS YOU READ . . .

1. T/F: Column vectors are used more in this text than row vectors, although some other texts do the opposite.
2. T/F: To multiply $A \times B$, the number of rows of A and B need to be the same.
3. T/F: The entry in the 2nd row and 3rd column of the product AB comes from multipling the 2nd row of A with the 3rd column of B.
4. Name two properties of matrix multiplication that also hold for "regular multiplication" of numbers.
5. Name a property of "regular multiplication" of numbers that does not hold for matrix multiplication.
6. T/F: $A^3 = A \cdot A \cdot A$

In the previous section we found that the definition of matrix addition was very intuitive, and we ended that section discussing the fact that eventually we'd like to know what it means to multiply matrices together.

In the spirit of the last section, take another wild stab: what do you think

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

means?

You are likely to have guessed

$$\begin{bmatrix} 1 & -2 \\ 6 & 8 \end{bmatrix}$$

but this is, in fact, *not* right. (I guess you *could* define multiplication this way. If you'd prefer this type of multiplication, you'll have to write your own book.) The actual answer is

$$\begin{bmatrix} 5 & 3 \\ 11 & 5 \end{bmatrix}.$$

If you can look at this one example and suddenly understand exactly how matrix multiplication works, then you are probably smarter than the author. While matrix multiplication isn't hard, it isn't nearly as intuitive as matrix addition is.

To further muddy the waters (before we clear them), consider

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

Our experience from the last section would lend us to believe that this is not defined, but our confidence is probably a bit shaken by now. In fact, this multiplication *is* defined, and it is

$$\begin{bmatrix} 5 & 3 & -2 \\ 11 & 5 & -4 \end{bmatrix}.$$

You may see some similarity in this answer to what we got before, but again, probably not enough to really figure things out.

So let's take a step back and progress slowly. The first thing we'd like to do is define a special type of matrix called a vector.

Definition 27 Column and Row Vectors

A $m \times 1$ matrix is called a *column vector*.
A $1 \times n$ matrix is called a *row vector*.

While it isn't obvious right now, column vectors are going to become far more useful to us than row vectors. Therefore, we often omit the word "column" when referring to column vectors, and we just call them "vectors."

In this text, row vectors are only used in this section when we discuss matrix multiplication, whereas we'll make extensive use of column vectors. Other texts make great use of row vectors, but little use of column vectors. It is a matter of preference and tradition: "most" texts use column vectors more. In some more advanced textbooks, row vectors are considered to be "dual" to column vectors. Abstractly, a *dual vector* is an object that eats a vector and spits out a number. Here, we see that the way a row vector eats a column vector and produces a number is via multiplication.

We have been using upper case letters to denote matrices; we use lower case letters with an arrow overtop to denote row and column vectors. An example of a row vector is

$$\vec{u} = \begin{bmatrix} 1 & 2 & -1 & 0 \end{bmatrix}$$

and an example of a column vector is

$$\vec{v} = \begin{bmatrix} 1 \\ 7 \\ 8 \end{bmatrix}.$$

Before we learn how to multiply matrices in general, we will learn what it means to multiply a row vector by a column vector.

Definition 28 Multiplying a row vector by a column vector

Let $\vec{u}$ be an $1 \times n$ row vector with entries $u_1, u_2, \cdots, u_n$ and let $\vec{v}$ be an $n \times 1$ column vector with entries $v_1, v_2, \cdots, v_n$. The *product of $\vec{u}$ and $\vec{v}$*, denoted $\vec{u} \cdot \vec{v}$ or $\vec{u}\vec{v}$, is

$$\sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Don't worry if this definition doesn't make immediate sense. It is really an easy concept; an example will make things more clear.

Example 51 Multiplying row and column vectors

Let

$$\vec{u} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, \vec{v} = \begin{bmatrix} 2 & 0 & 1 & -1 \end{bmatrix}, \vec{x} = \begin{bmatrix} -2 \\ 4 \\ 3 \end{bmatrix}, \vec{y} = \begin{bmatrix} 1 \\ 2 \\ 5 \\ 0 \end{bmatrix}.$$

Find the following products.

1. $\vec{u}\vec{x}$
2. $\vec{v}\vec{y}$
3. $\vec{u}\vec{y}$
4. $\vec{u}\vec{v}$
5. $\vec{x}\vec{u}$

Solution

1. $\vec{u}\vec{x} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} -2 \\ 4 \\ 3 \end{bmatrix} = 1(-2) + 2(4) + 3(3) = 15$

2. $\vec{v}\vec{y} = \begin{bmatrix} 2 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 5 \\ 0 \end{bmatrix} = 2(1) + 0(2) + 1(5) - 1(0) = 7$

3. $\vec{u}\vec{y}$ is not defined; Definition 28 specifies that in order to multiply a row vector and column vector, they must have the same number of entries.

4. $\vec{u}\vec{v}$ is not defined; we only know how to multipy row vectors by column vectors. We haven't defined how to multiply two row vectors (in general, it can't be done).

5. The product $\vec{x}\vec{u}$ *is* defined, but we don't know how to do it yet. Right now, we only know how to multiply a row vector times a column vector; we don't know how to multiply a column vector times a row vector. (That's right: $\vec{u}\vec{x} \neq \vec{x}\vec{u}$!)

Now that we understand how to multiply a row vector by a column vector, we are ready to define matrix multiplication.

Definition 29 Matrix Multiplication

Let A be an $m \times r$ matrix, and let B be an $r \times n$ matrix. The *matrix product of A and B*, denoted $A \cdot B$, or simply AB, is the $m \times n$ matrix M whose entry in the i^{th} row and j^{th} column is the product of the i^{th} row of A and the j^{th} column of B.

It may help to illustrate it in this way. Let matrix A have rows $\vec{a_1}, \vec{a_2}, \cdots, \vec{a_m}$ and let B have columns $\vec{b_1}, \vec{b_2}, \cdots, \vec{b_n}$. Thus A looks like

$$\begin{bmatrix} - & \vec{a_1} & - \\ - & \vec{a_2} & - \\ & \vdots & \\ - & \vec{a_m} & - \end{bmatrix}$$

where the "$-$" symbols just serve as reminders that the $\vec{a_i}$ represent rows, and B looks like

$$\begin{bmatrix} | & | & & | \\ \vec{b_1} & \vec{b_2} & \cdots & \vec{b_n} \\ | & | & & | \end{bmatrix}$$

where again, the "$|$" symbols just remind us that the $\vec{b_i}$ represent column vectors. Then

$$AB = \begin{bmatrix} \vec{a_1}\vec{b_1} & \vec{a_1}\vec{b_2} & \cdots & \vec{a_1}\vec{b_n} \\ \vec{a_2}\vec{b_1} & \vec{a_2}\vec{b_2} & \cdots & \vec{a_2}\vec{b_n} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a_m}\vec{b_1} & \vec{a_m}\vec{b_2} & \cdots & \vec{a_m}\vec{b_n} \end{bmatrix}.$$

Two quick notes about this definition. First, notice that in order to multiply A and B, the number of *columns* of A must be the same as the number of *rows* of B (we refer to these as the "inner dimensions"). Secondly, the resulting matrix has the same number of *rows* as A and the same number of *columns* as B (we refer to these as the "outer dimensions").

$$\overbrace{(m \times \underbrace{r) \times (r}_{\text{these inner dimensions must match}} \times n)}^{\text{final dimensions are the outer dimensions}}$$

Of course, this will make much more sense when we see an example.

Example 52 **A more general matrix product**

Revisit the matrix product we saw at the beginning of this section; multiply

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

Solution Let's call our first matrix A and the second B. We should first check to see that we can actually perform this multiplication. Matrix A is 2×2 and B is 2×3. The "inner" dimensions match up, so we can compute the product; the "outer" dimensions tell us that the product will be 2×3. Let

$$AB = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix}.$$

Let's find the value of each of the entries.

The entry m_{11} is in the first row and first column; therefore to find its value, we need to multiply the first row of A by the first column of B. Thus

$$m_{11} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1(1) + 2(2) = 5.$$

So now we know that

$$AB = \begin{bmatrix} 5 & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix}.$$

Finishing out the first row, we have

$$m_{12} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = 1(-1) + 2(2) = 3$$

using the first row of A and the second column of B, and

$$m_{13} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = 1(0) + 2(-1) = -2$$

using the first row of A and the third column of B. Thus we have

$$AB = \begin{bmatrix} 5 & 3 & -2 \\ m_{21} & m_{22} & m_{23} \end{bmatrix}.$$

To compute the second row of AB, we multiply with the second row of A. We find

$$m_{21} = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 11,$$

$$m_{22} = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = 5, \text{ and}$$

$$m_{23} = \begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = -4.$$

Thus

$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 2 & 2 & -1 \end{bmatrix} = \begin{bmatrix} 5 & 3 & -2 \\ 11 & 5 & -4 \end{bmatrix}.$$

Example 53 **Multiplying matrices**

Multiply

$$\begin{bmatrix} 1 & -1 \\ 5 & 2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 6 & 7 & 9 \end{bmatrix}.$$

Solution Let's first check to make sure this product is defined. Again calling the first matrix A and the second B, we see that A is a 3×2 matrix and B is a 2×4 matrix; the inner dimensions match so the product is defined, and the product will be a 3×4 matrix,

$$AB = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix}.$$

We will demonstrate how to compute some of the entries, then give the final answer. The reader can fill in the details of how each entry was computed.

$$m_{11} = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = -1.$$

$$m_{13} = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 7 \end{bmatrix} = -6.$$

$$m_{23} = \begin{bmatrix} 5 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 7 \end{bmatrix} = 19.$$

$$m_{24} = \begin{bmatrix} 5 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 9 \end{bmatrix} = 23.$$

$$m_{32} = \begin{bmatrix} -2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix} = 16.$$

$$m_{34} = \begin{bmatrix} -2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 9 \end{bmatrix} = 25.$$

So far, we've computed this much of AB:

$$AB = \begin{bmatrix} -1 & m_{12} & -6 & m_{14} \\ m_{21} & m_{22} & 19 & 23 \\ m_{31} & 16 & m_{33} & 25 \end{bmatrix}.$$

The final product is

$$AB = \begin{bmatrix} -1 & -5 & -6 & -8 \\ 9 & 17 & 19 & 23 \\ 4 & 16 & 19 & 25 \end{bmatrix}.$$

Example 54 **An undefined product**

Multiply, if possible,

$$\begin{bmatrix} 2 & 3 & 4 \\ 9 & 8 & 7 \end{bmatrix} \begin{bmatrix} 3 & 6 \\ 5 & -1 \end{bmatrix}.$$

SOLUTION Again, we'll call the first matrix A and the second B. Checking the dimensions of each matrix, we see that A is a 2×3 matrix, whereas B is a 2×2 matrix. The inner dimensions do not match, therefore this multiplication is not defined.

Example 55 **A vector product revisited**
In Example 50, we were told that the product $\vec{x}\vec{u}$ was defined, where

$$\vec{x} = \begin{bmatrix} -2 \\ 4 \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{u} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix},$$

although we were not shown what that product was. Find $\vec{x}\vec{u}$.

SOLUTION Again, we need to check to make sure the dimensions work correctly (remember that even though we are referring to $\vec{u}$ and $\vec{x}$ as vectors, they are, in fact, just matrices).

The column vector $\vec{x}$ has dimensions 3×1, whereas the row vector $\vec{u}$ has dimensions 1×3. Since the inner dimensions do match, the matrix product is defined; the outer dimensions tell us that the product will be a 3×3 matrix, as shown below:

$$\vec{x}\vec{u} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}.$$

To compute the entry m_{11}, we multiply the first row of $\vec{x}$ by the first column of $\vec{u}$. What is the first row of $\vec{x}$? Simply the number -2. What is the first column of $\vec{u}$? Just the number 1. Thus $m_{11} = -2$. (This does seem odd, but through checking, you can see that we are indeed following the rules.)

What about the entry m_{12}? Again, we multiply the first row of $\vec{x}$ by the first column of $\vec{u}$; that is, we multiply $-2(2)$. So $m_{12} = -4$.

What about m_{23}? Multiply the second row of $\vec{x}$ by the third column of $\vec{u}$; multiply $4(3)$, so $m_{23} = 12$.

One final example: m_{31} comes from multiplying the third row of $\vec{x}$, which is 3, by the first column of $\vec{u}$, which is 1. Therefore $m_{31} = 3$.

So far we have computed

$$\vec{x}\vec{u} = \begin{bmatrix} -2 & -4 & m_{13} \\ m_{21} & m_{22} & 12 \\ 3 & m_{32} & m_{33} \end{bmatrix}.$$

After performing all 9 multiplications, we find

$$\vec{x}\vec{u} = \begin{bmatrix} -2 & -4 & -6 \\ 4 & 8 & 12 \\ 3 & 6 & 9 \end{bmatrix}.$$

In this last example, we saw a "nonstandard" multiplication (at least, it felt nonstandard). Studying the entries of this matrix, it seems that there are several different patterns that can be seen amongst the entries. (Remember that mathematicians like to look for patterns. Also remember that we often guess wrong at first; don't be scared and try to identify some patterns.)

In Section 3.1, we identified the zero matrix $\mathbf{0}$ that had a nice property in relation to matrix addition (i.e., $A + \mathbf{0} = A$ for any matrix A). In the following example we'll identify a matrix that works well with multiplication as well as some multiplicative properties. For instance, we've learned how $1 \cdot A = A$; is there a *matrix* that acts like the number 1? That is, can we find a matrix X where $X \cdot A = A$? (We made a guess in Section 3.1 that maybe a matrix of all 1s would work, but you can probably already see that this guess is doomed to failure.)

Example 56 **Computing matrix products**

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}, \qquad I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Find the following products.

1. AB
2. BA
3. $A\mathbf{0}_{3\times 4}$
4. AI
5. IA
6. I^2
7. BC
8. B^2

Solution We will find each product, but we leave the details of each computation to the reader.

1. $AB = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 6 & 6 \\ 0 & 0 & 0 \\ -7 & -7 & -7 \end{bmatrix}$

2. $BA = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix} = \begin{bmatrix} 1 & -13 & 11 \\ 1 & -13 & 11 \\ 1 & -13 & 11 \end{bmatrix}$

3. $A\mathbf{0}_{3\times 4} = \mathbf{0}_{3\times 4}$.

4. $AI = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix}$

5. $IA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -7 & 5 \\ -2 & -8 & 3 \end{bmatrix}$

6. We haven't formally defined what I^2 means, but we could probably make the reasonable guess that $I^2 = I \cdot I$. Thus

$$I^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

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

8. $B^2 = BB = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{bmatrix}$

This example is simply chock full of interesting ideas; it is almost hard to think about where to start.

Interesting Idea #1: Notice that in our example, $AB \neq BA$! When dealing with numbers, we were used to the idea that $ab = ba$. With matrices, multiplication is *not* commutative. (Of course, we can find special situations where it does work. In general, though, it doesn't.)

Interesting Idea #2: Right before this example we wondered if there was a matrix that "acted like the number 1," and guessed it may be a matrix of all 1s. However, we found out that such a matrix does not work in that way; in our example, $AB \neq A$. We did find that $AI = IA = A$. There is a Multiplicative Identity; it just isn't what we thought it would be. And just as $1^2 = 1$, $I^2 = I$.

Interesting Idea #3: When dealing with numbers, we are very familiar with the notion that "If $ax = bx$, then $a = b$." (As long as $x \neq 0$.) Notice that, in our example, $BB = BC$, yet $B \neq C$. In general, just because $AX = BX$, we *cannot* conclude that $A = B$.

Matrix multiplication is turning out to be a very strange operation. We are very used to multiplying numbers, and we know a bunch of properties that hold when using this type of multiplication. When multiplying matrices, though, we probably find ourselves asking two questions, "What *does* work?" and "What *doesn't* work?" We'll answer these questions; first we'll do an example that demonstrates some of the things that do work.

Example 57 **Exploring properties of matrix multiplication**

Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Find the following:

1. $A(B + C)$
2. $AB + AC$
3. $A(BC)$
4. $(AB)C$

SOLUTION We'll compute each of these without showing all the intermediate steps. Keep in mind order of operations: things that appear inside of parentheses are computed first.

1.

$$\begin{aligned} A(B+C) &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 7 & 4 \\ 17 & 10 \end{bmatrix} \end{aligned}$$

2.

$$\begin{aligned} AB + AC &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 3 & -1 \\ 7 & -1 \end{bmatrix} + \begin{bmatrix} 4 & 5 \\ 10 & 11 \end{bmatrix} \\ &= \begin{bmatrix} 7 & 4 \\ 17 & 10 \end{bmatrix} \end{aligned}$$

3.
$$\begin{aligned} A(BC) &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 3 & 3 \\ 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 1 \\ 13 & 5 \end{bmatrix} \end{aligned}$$

4.
$$\begin{aligned} (AB)\,C &= \left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 3 & -1 \\ 7 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 1 \\ 13 & 5 \end{bmatrix} \end{aligned}$$

In looking at our example, we should notice two things. First, it looks like the "distributive property" holds; that is, $A(B+C) = AB + AC$. This is nice as many algebraic techniques we have learned about in the past (when doing "ordinary algebra") will still work. Secondly, it looks like the "associative property" holds; that is, $A(BC) = (AB)C$. This is nice, for it tells us that when we are multiplying several matrices together, we don't have to be particularly careful in what order we multiply certain pairs of matrices together.

Be careful: in computing ABC together, we can first multiply AB or BC, but we cannot change the *order* in which these matrices appear. We cannot multiply BA or AC, for instance.

In leading to an important theorem, let's define a matrix we saw in an earlier example.

Definition 30 Identity Matrix

The $n \times n$ matrix with 1's on the diagonal and zeros elsewhere is the $n \times n$ *identity matrix*, denoted I_n. When the context makes the dimension of the identity clear, the subscript is generally omitted.

Definition 30 uses a term we won't define until Definition 45 on page 218: *diagonal*. In short, a "diagonal matrix" is one in which the only nonzero entries are the "diagonal entries." The examples given here and in the exercises should suffice until we meet the full definition later.

Note that while the zero matrix can come in all different shapes and sizes, the identity matrix is always a square matrix. We show a few identity matrices below.

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

In our examples above, we have seen examples of things that do and do not work. We should be careful about what examples *prove*, though. If someone were to claim that $AB = BA$ is always true, one would only need to show them one example where they were false, and we would know the person was wrong. However, if someone claims that $A(B+C) = AB + AC$ is always true, we can't prove this with just one example. We need something more powerful; we need a true proof.

In this text, we forgo most proofs. The reader should know, though, that when we state something in a theorem, there is a proof that backs up what we state. Our justification comes from something stronger than just examples.

Now we give the good news of what does work when dealing with matrix multiplication.

Theorem 9 Properties of Matrix Multiplication

Let A, B and C be matrices with dimensions so that the following operations make sense, and let k be a scalar. The following equalities hold:

1. $A(BC) = (AB)C$ (Associative Property)
2. $A(B + C) = AB + AB$ and
   $(B + C)A = BA + CA$ (Distributive Property)
3. $k(AB) = (kA)B = A(kB)$
4. $AI = IA = A$

The above box contains some very good news, and probably some very surprising news. Matrix multiplication probably seems to us like a very odd operation, so we probably wouldn't have been surprised if we were told that $A(BC) \neq (AB)C$. It is a very nice thing that the Associative Property does hold.

As we near the end of this section, we raise one more issue of notation. We define $A^0 = I$. If n is a positive integer, we define

$$A^n = \underbrace{A \cdot A \cdot \cdots \cdot A}_{n \text{ times}}.$$

With numbers, we are used to $a^{-n} = \frac{1}{a^n}$. Do negative exponents work with matrices, too? The answer is yes, sort of. We'll have to be careful, and we'll cover the topic in detail once we define the inverse of a matrix. For now, though, we recognize the fact that $A^{-1} \neq \frac{1}{A}$, for $\frac{1}{A}$ makes no sense; we don't know how to "divide" by a matrix.

We end this section with a reminder of some of the things that do not work with matrix multiplication. The good news is that there are really only two things on this list.

1. Matrix multiplication is not commutative; that is, $AB \neq BA$.
2. In general, just because $AX = BX$, we cannot conclude that $A = B$.

The bad news is that these ideas pop up in many places where we don't expect them. For instance, we are used to

$$(a + b)^2 = a^2 + 2ab + b^2.$$

What about $(A + B)^2$? All we'll say here is that

$$(A + B)^2 \neq A^2 + 2AB + B^2;$$

we leave it to the reader to figure out why.

The next section is devoted to visualizing column vectors and "seeing" how some of these arithmetic properties work together.

Exercises 3.2

Problems

In Exercises 1 – 12, row and column vectors $\vec{u}$ and $\vec{v}$ are defined. Find the product $\vec{u}\vec{v}$, where possible.

1. $\vec{u} = \begin{bmatrix} 1 & -4 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$

2. $\vec{u} = \begin{bmatrix} 2 & 3 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$

3. $\vec{u} = \begin{bmatrix} 1 & -1 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

4. $\vec{u} = \begin{bmatrix} 0.6 & 0.8 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}$

5. $\vec{u} = \begin{bmatrix} 1 & 2 & -1 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$

6. $\vec{u} = \begin{bmatrix} 3 & 2 & -2 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} -1 \\ 0 \\ 9 \end{bmatrix}$

7. $\vec{u} = \begin{bmatrix} 8 & -4 & 3 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$

8. $\vec{u} = \begin{bmatrix} -3 & 6 & 1 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

9. $\vec{u} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$
   $\vec{v} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$

10. $\vec{u} = \begin{bmatrix} 6 & 2 & -1 & 2 \end{bmatrix}$
    $\vec{v} = \begin{bmatrix} 3 \\ 2 \\ 9 \\ 5 \end{bmatrix}$

11. $\vec{u} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

12. $\vec{u} = \begin{bmatrix} 2 & -5 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

In Exercises 13 – 27, matrices A and B are defined.

(a) Give the dimensions of A and B. If the dimensions properly match, give the dimensions of AB and BA.

(b) Find the products AB and BA, if possible.

13. $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \end{bmatrix} \; B = \begin{bmatrix} 2 & 5 \\ 3 & -1 \end{bmatrix}$

14. $A = \begin{bmatrix} 3 & 7 \\ 2 & 5 \end{bmatrix} \; B = \begin{bmatrix} 1 & -1 \\ 3 & -3 \end{bmatrix}$

15. $A = \begin{bmatrix} 3 & -1 \\ 2 & 2 \end{bmatrix}$
    $B = \begin{bmatrix} 1 & 0 & 7 \\ 4 & 2 & 9 \end{bmatrix}$

16. $A = \begin{bmatrix} 0 & 1 \\ 1 & -1 \\ -2 & -4 \end{bmatrix}$
    $B = \begin{bmatrix} -2 & 0 \\ 3 & 8 \end{bmatrix}$

17. $A = \begin{bmatrix} 9 & 4 & 3 \\ 9 & -5 & 9 \end{bmatrix}$
    $B = \begin{bmatrix} -2 & 5 \\ -2 & -1 \end{bmatrix}$

18. $A = \begin{bmatrix} -2 & -1 \\ 9 & -5 \\ 3 & -1 \end{bmatrix}$
    $B = \begin{bmatrix} -5 & 6 & -4 \\ 0 & 6 & -3 \end{bmatrix}$

19. $A = \begin{bmatrix} 2 & 6 \\ 6 & 2 \\ 5 & -1 \end{bmatrix}$
    $B = \begin{bmatrix} -4 & 5 & 0 \\ -4 & 4 & -4 \end{bmatrix}$

20. $A = \begin{bmatrix} -5 & 2 \\ -5 & -2 \\ -5 & -4 \end{bmatrix}$
    $B = \begin{bmatrix} 0 & -5 & 6 \\ -5 & -3 & -1 \end{bmatrix}$

21. $A = \begin{bmatrix} 8 & -2 \\ 4 & 5 \\ 2 & -5 \end{bmatrix}$
    $B = \begin{bmatrix} -5 & 1 & -5 \\ 8 & 3 & -2 \end{bmatrix}$

22. $A = \begin{bmatrix} 1 & 4 \\ 7 & 6 \end{bmatrix}$
    $B = \begin{bmatrix} 1 & -1 & -5 & 5 \\ -2 & 1 & 3 & -5 \end{bmatrix}$

23. $A = \begin{bmatrix} -1 & 5 \\ 6 & 7 \end{bmatrix}$
    $B = \begin{bmatrix} 5 & -3 & -4 & -4 \\ -2 & -5 & -5 & -1 \end{bmatrix}$

24. $A = \begin{bmatrix} -1 & 2 & 1 \\ -1 & 2 & -1 \\ 0 & 0 & -2 \end{bmatrix}$

$B = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & -1 \\ 1 & 0 & 0 \end{bmatrix}$

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

$B = \begin{bmatrix} -2 & -2 & -2 \\ 0 & -2 & 0 \\ -2 & 0 & 2 \end{bmatrix}$

26. $A = \begin{bmatrix} -4 & 3 & 3 \\ -5 & -1 & -5 \\ -5 & 0 & -1 \end{bmatrix}$

$B = \begin{bmatrix} 0 & 5 & 0 \\ -5 & -4 & 3 \\ 5 & -4 & 3 \end{bmatrix}$

27. $A = \begin{bmatrix} -4 & -1 & 3 \\ 2 & -3 & 5 \\ 1 & 5 & 3 \end{bmatrix}$

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

In Exercises 28 – 33, a *diagonal* matrix D and a matrix A are given. Find the products DA and AD, where possible.

28. $D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$

$A = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}$

29. $D = \begin{bmatrix} 4 & 0 \\ 0 & -3 \end{bmatrix}$

$A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$

30. $D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

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

$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

32. $D = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}$

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

33. $D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$

$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$

In Exercises 34 – 39, a matrix A and a vector $\vec{x}$ are given. Find the product $A\vec{x}$.

34. $A = \begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}$

35. $A = \begin{bmatrix} -1 & 4 \\ 7 & 3 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

36. $A = \begin{bmatrix} 2 & 0 & 3 \\ 1 & 1 & 1 \\ 3 & -1 & 2 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}$

37. $A = \begin{bmatrix} -2 & 0 & 3 \\ 1 & 1 & -2 \\ 4 & 2 & -1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix}$

38. $A = \begin{bmatrix} 2 & -1 \\ 4 & 3 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

39. $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 2 \\ 2 & 3 & 1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

40. Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Find A^2 and A^3.

41. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$. Find A^2 and A^3.

42. Let $A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$. Find A^2 and A^3.

43. Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$. Find A^2 and A^3.

44. Let $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Find A^2 and A^3.

45. In the text we state that $(A+B)^2 \neq A^2 + 2AB + B^2$. We investigate that claim here.
    (a) Let $A = \begin{bmatrix} 5 & 3 \\ -3 & -2 \end{bmatrix}$ and let $B = \begin{bmatrix} -5 & -5 \\ -2 & 1 \end{bmatrix}$. Compute $A + B$.
    (b) Find $(A+B)^2$ by using your answer from (a).
    (c) Compute $A^2 + 2AB + B^2$.

(d) Are the results from (a) and (b) the same?

(e) Carefully expand the expression $(A + B)^2 = (A + B)(A + B)$ and show why this is not equal to $A^2 + 2AB + B^2$.

3.3 Visualizing Matrix Arithmetic in 2D

AS YOU READ . . .

1. T/F: Two vectors with the same length and direction are equal even if they start from different places.
2. One can visualize vector addition using what law?
3. T/F: Multiplying a vector by 2 doubles its length.
4. What do mathematicians do?
5. T/F: Multiplying a vector by a matrix always changes its length and direction.

When we first learned about adding numbers together, it was useful to picture a number line: $2 + 3 = 5$ could be pictured by starting at 0, going out 2 tick marks, then another 3, and then realizing that we moved 5 tick marks from 0. Similar visualizations helped us understand what $2 - 3$ meant and what 2×3 meant.

We now investigate a way to picture matrix arithmetic – in particular, operations involving column vectors. This not only will help us better understand the arithmetic operations, it will open the door to a great wealth of interesting study. Visualizing matrix arithmetic has a wide variety of applications, the most common being computer graphics. While we often think of these graphics in terms of video games, there are numerous other important applications. For example, chemists and biologists often use computer models to "visualize" complex molecules to "see" how they interact with other molecules.

We will start with vectors in two dimensions (2D) – that is, vectors with only two entries. We assume the reader is familiar with the Cartesian plane, that is, plotting points and graphing functions on "the x–y plane." We graph vectors in a manner very similar to plotting points. Given the vector

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

we draw $\vec{x}$ by drawing an arrow whose tip is 1 unit to the right and 2 units up from its origin. (To help reduce clutter, in all figures each tick mark represents one unit.)

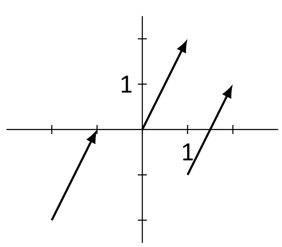


Figure 3.1: Various drawings of $\vec{x}$

When drawing vectors, we do not specify where you start drawing; all we specify is where the tip lies based on where we started. Figure 3.1 shows vector $\vec{x}$ drawn 3 ways. In some ways, the "most common" way to draw a vector has the arrow start at the origin,but this is by no means the only way of drawing the vector.

Let's practice this concept by drawing various vectors from given starting points.

Example 58 **Sketching vectors**

Let

$$\vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \vec{y} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \text{and} \quad \vec{z} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}.$$

Draw $\vec{x}$ starting from the point $(0, -1)$; draw $\vec{y}$ starting from the point $(-1, -1)$, and draw $\vec{z}$ starting from the point $(2, -1)$.

Solution To draw $\vec{x}$, start at the point $(0, -1)$ as directed, then move to the right one unit and down one unit and draw the tip. Thus the arrow "points" from $(0, -1)$ to $(1, -2)$.

To draw $\vec{y}$, we are told to start and the point $(-1, -1)$. We draw the tip by moving to the right 2 units and up 3 units; hence $\vec{y}$ points from $(-1, -1)$ to (1,2).

To draw $\vec{z}$, we start at $(2, -1)$ and draw the tip 3 units to the left and 2 units up; $\vec{z}$ points from $(2, -1)$ to $(-1, 1)$.

Each vector is drawn as shown in Figure 3.2.

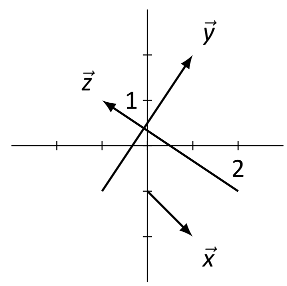


Figure 3.2: Drawing vectors $\vec{x}$, $\vec{y}$ and $\vec{z}$ in Example 57

Vectors are just special types of matrices. The zero vector, $\vec{0}$, is a special type of zero matrix, **0**. It helps to distinguish the two by using different notation.

How does one draw the zero vector, $\vec{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$? Following our basic procedure, we start by going 0 units in the *x* direction, followed by 0 units in the *y* direction. In other words, we don't go anywhere. In general, we don't actually draw $\vec{0}$. At best, one can draw a dark circle at the origin to convey the idea that $\vec{0}$, when starting at the origin, points to the origin.

In section 3.1 we learned about matrix arithmetic operations: matrix addition and scalar multiplication. Let's investigate how we can "draw" these operations.

Vector Addition

Given two vectors $\vec{x}$ and $\vec{y}$, how do we draw the vector $\vec{x}+\vec{y}$? Let's look at this in the context of an example, then study the result.

Example 59 **Sketching vectors and their sum**

Let

$$\vec{x}=\begin{bmatrix}1\\1\end{bmatrix} \quad \text{and} \quad \vec{y}=\begin{bmatrix}3\\1\end{bmatrix}.$$

Sketch $\vec{x}$, $\vec{y}$ and $\vec{x}+\vec{y}$.

Solution A starting point for drawing each vector was not given; by default, we'll start at the origin. (This is in many ways nice; this means that the *vector* $\begin{bmatrix}3\\1\end{bmatrix}$ "points" to the *point* (3,1).) We first compute $\vec{x}+\vec{y}$:

$$\vec{x}+\vec{y}=\begin{bmatrix}1\\1\end{bmatrix}+\begin{bmatrix}3\\1\end{bmatrix}=\begin{bmatrix}4\\2\end{bmatrix}$$

Sketching each gives the picture in Figure 3.3.

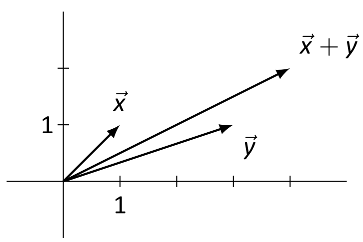


Figure 3.3: Adding vectors $\vec{x}$ and $\vec{y}$ in Example 58

This example is pretty basic; we were given two vectors, told to add them together, then sketch all three vectors. Our job now is to go back and try to see a relationship between the drawings of $\vec{x}$, $\vec{y}$ and $\vec{x}+\vec{y}$. Do you see any?

Here is one way of interpreting the adding of $\vec{x}$ to $\vec{y}$. Regardless of where we start, we draw $\vec{x}$. Now, from the tip of $\vec{x}$, draw $\vec{y}$. The vector $\vec{x}+\vec{y}$ is the vector found by drawing an arrow from the *origin* of $\vec{x}$ to the *tip* of $\vec{y}$. Likewise, we could start by drawing $\vec{y}$. Then, starting from the tip of $\vec{y}$, we can draw $\vec{x}$. Finally, draw $\vec{x}+\vec{y}$ by drawing the vector that starts at the origin of $\vec{y}$ and ends at the tip of $\vec{x}$.

The picture in Figure 3.4 illustrates this. The gray vectors demonstrate drawing the second vector from the tip of the first; we draw the vector $\vec{x}+\vec{y}$ dashed to set it apart from the rest. We also lightly filled the *parallelogram* whose opposing sides are the vectors $\vec{x}$ and $\vec{y}$. This highlights what is known as the *Parallelogram Law*.

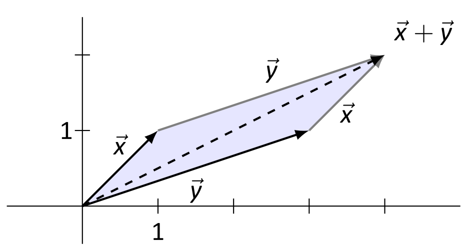


Figure 3.4: Adding vectors graphically using the Parallelogram Law

> **Key Idea 10 Parallelogram Law**
>
> To draw the vector $\vec{x}+\vec{y}$, one can draw the parallelogram with $\vec{x}$ and $\vec{y}$ as its sides. The vector that points from the vertex where $\vec{x}$ and $\vec{y}$ originate to the vertex where $\vec{x}$ and $\vec{y}$ meet is the vector $\vec{x}+\vec{y}$.

Knowing all of this allows us to draw the sum of two vectors without knowing specifically what the vectors are, as we demonstrate in the following example.

Example 60 Addition of vectors
Consider the vectors $\vec{x}$ and $\vec{y}$ as drawn in Figure 3.5. Sketch the vector $\vec{x}+\vec{y}$.

Solution

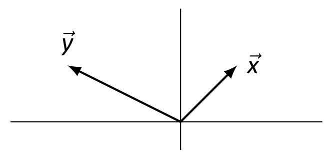


Figure 3.5: Vectors $\vec{x}$ and $\vec{y}$ in Example 59

We'll apply the Parallelogram Law, as given in Key Idea 10. As before, we draw $\vec{x}+\vec{y}$ dashed to set it apart. The result is given in Figure 3.6.

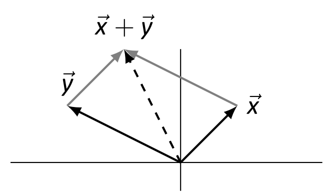


Figure 3.6: Vectors $\vec{x}$, $\vec{y}$ and $\vec{x}+\vec{y}$ in Example 59

Scalar Multiplication

After learning about matrix addition, we learned about scalar multiplication. We apply that concept now to vectors and see how this is represented graphically.

Example 61 Visualizing scalar multiplication
Let

$$\vec{x}=\begin{bmatrix}1\\1\end{bmatrix} \quad \text{and} \quad \vec{y}=\begin{bmatrix}-2\\1\end{bmatrix}.$$

Sketch $\vec{x}$, $\vec{y}$, $3\vec{x}$ and $-1\vec{y}$.

Solution We begin by computing $3\vec{x}$ and $-\vec{y}$:

$$3\vec{x}=\begin{bmatrix}3\\3\end{bmatrix} \quad \text{and} \quad -\vec{y}=\begin{bmatrix}2\\-1\end{bmatrix}.$$

All four vectors are sketched in Figure 3.7.

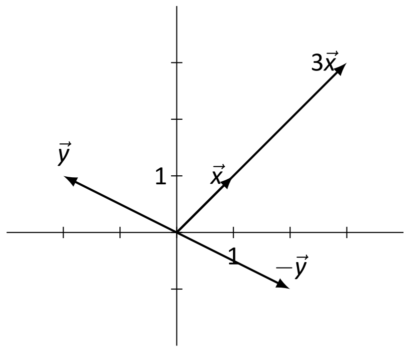


Figure 3.7: Vectors $\vec{x}$, $\vec{y}$, $3\vec{x}$ and $-\vec{y}$ in Example 60

As we often do, let us look at the previous example and see what we can learn from it. We can see that $\vec{x}$ and $3\vec{x}$ point in the same direction (they lie on the same line), but $3\vec{x}$ is just longer than $\vec{x}$. (In fact, it looks like $3\vec{x}$ is *3* times longer than $\vec{x}$. Is it? How do we measure length?)

We also see that $\vec{y}$ and $-\vec{y}$ seem to have the same length and lie on the same line, but point in the opposite direction.

A vector inherently conveys two pieces of information: length and direction. Multiplying a vector by a positive scalar c stretches the vectors by a factor of c; multiplying by a negative scalar c both stretches the vector and makes it point in the opposite direction.

Knowing this, we can sketch scalar multiples of vectors without knowing specifically what they are, as we do in the following example.

Example 62 **Sketching scalar multiples**

Let vectors $\vec{x}$ and $\vec{y}$ be as in Figure 3.8. Draw $3\vec{x}$, $-2\vec{x}$, and $\frac{1}{2}\vec{y}$.

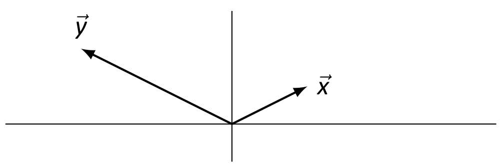


Figure 3.8: Vectors $\vec{x}$ and $\vec{y}$ in Example 61

Solution To draw $3\vec{x}$, we draw a vector in the same direction as $\vec{x}$, but 3 times as long. To draw $-2\vec{x}$, we draw a vector twice as long as $\vec{x}$ in the opposite direction; to draw $\frac{1}{2}\vec{y}$, we draw a vector half the length of $\vec{y}$ in the same direction as $\vec{y}$. We again use the default of drawing all the vectors starting at the origin. All of this is shown in Figure 3.9.

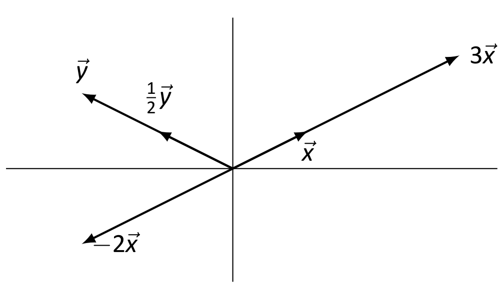


Figure 3.9: Vectors $\vec{x}$, $\vec{y}$, $3\vec{x}$, $-2x$ and $\frac{1}{2}\vec{x}$ in Example 61

Vector Subtraction

The final basic operation to consider between two vectors is that of vector subtraction: given vectors $\vec{x}$ and $\vec{y}$, how do we draw $\vec{x}-\vec{y}$?

If we know explicitly what $\vec{x}$ and $\vec{y}$ are, we can simply compute what $\vec{x}-\vec{y}$ is and then draw it. We can also think in terms of vector addition and scalar multiplication: we can *add* the vectors $\vec{x}+(-1)\vec{y}$. That is, we can draw $\vec{x}$ and draw $-\vec{y}$, then add them as we did in Example 59. This is especially useful we don't know explicitly what $\vec{x}$ and $\vec{y}$ are.

Example 63 **Sketching the difference of two vectors**
Let vectors $\vec{x}$ and $\vec{y}$ be as in Figure 3.10. Draw $\vec{x}-\vec{y}$.

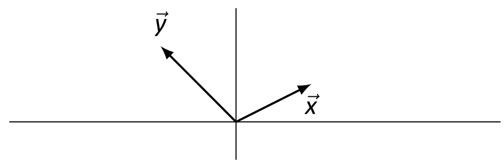


Figure 3.10: Vectors $\vec{x}$ and $\vec{y}$ in Example 62

Solution To draw $\vec{x}-\vec{y}$, we will first draw $-\vec{y}$ and then apply the Parallelogram Law to add $\vec{x}$ to $-\vec{y}$. See Figure 3.11.

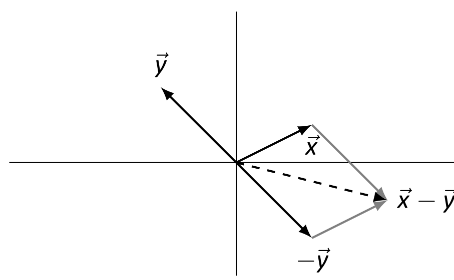


Figure 3.11: Vectors $\vec{x}$, $\vec{y}$ and $\vec{x}-\vec{y}$ in Example 62

In Figure 3.12, we redraw Figure 3.11 from Example 62 but remove the gray vectors that tend to add clutter, and we redraw the vector $\vec{x}-\vec{y}$ dotted so that it starts from the tip of $\vec{y}$. (Remember that we can draw vectors starting from anywhere.) Note that the dotted version of $\vec{x}-\vec{y}$ points from $\vec{y}$ to $\vec{x}$. This is a "shortcut" to drawing $\vec{x}-\vec{y}$; simply draw the vector that starts at the tip of $\vec{y}$ and ends at the tip of $\vec{x}$. This is important so we make it a Key Idea.

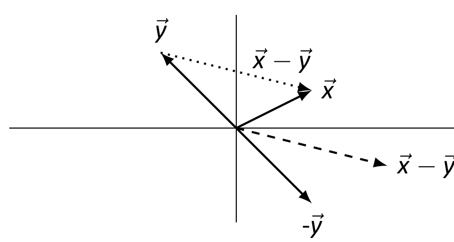


Figure 3.12: Redrawing vector $\vec{x}-\vec{y}$

Key Idea 11 Vector Subtraction

To draw the vector $\vec{x} - \vec{y}$, draw $\vec{x}$ and $\vec{y}$ so that they have the same origin. The vector $\vec{x} - \vec{y}$ is the vector that starts from the tip of $\vec{y}$ and points to the tip of $\vec{x}$.

Let's practice this once more with a quick example.

Example 64 Sketching a vector difference
Let $\vec{x}$ and $\vec{y}$ be as in Figure 3.13 (a). Draw $\vec{x} - \vec{y}$.

Solution We simply apply Key Idea 11: we draw an arrow from $\vec{y}$ to $\vec{x}$. We do so in Figure 3.13 (b); $\vec{x} - \vec{y}$ is dashed.

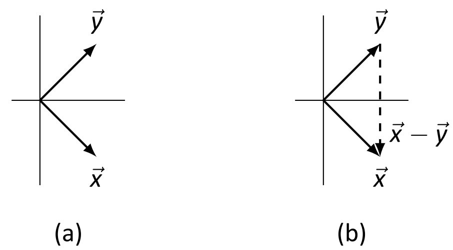


Figure 3.13: Vectors $\vec{x}$, $\vec{y}$ and $\vec{x} - \vec{y}$ in Example 63

Vector Length

When we discussed scalar multiplication, we made reference to a fundamental question: How do we measure the length of a vector? Basic geometry gives us an answer in the two dimensional case that we are dealing with right now, and later we can extend these ideas to higher dimensions.

Consider Figure 3.14. A vector $\vec{x}$ is drawn in black, and dashed and dotted lines have been drawn to make it the hypotenuse of a right triangle.

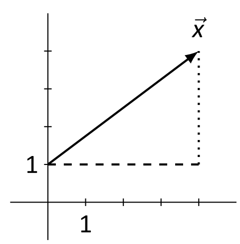


Figure 3.14: Measuring the length of a vector

It is easy to see that the dashed line has length 4 and the dotted line has length 3. We'll let c denote the length of $\vec{x}$; according to the Pythagorean Theorem, $4^2 + 3^2 = c^2$. Thus $c^2 = 25$ and we quickly deduce that $c = 5$.

Notice that in our figure, $\vec{x}$ goes to the right 4 units and then up 3 units. In other words, we can write

$$\vec{x} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

We learned above that the length of $\vec{x}$ is $\sqrt{4^2 + 3^2}$. (Remember that $\sqrt{4^2 + 3^2} \neq 4 + 3$!) This hints at a basic calculation that works for all vectors $\vec{x}$, and we define the length of a vector according to this rule.

Definition 31 Vector Length

Let

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

The *length* of $\vec{x}$, denoted $||\vec{x}||$, is

$$||\vec{x}|| = \sqrt{x_1^2 + x_2^2}.$$

Example 65 Calculating the length of vectors
Find the length of each of the vectors given below.

$$\vec{x_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \vec{x_2} = \begin{bmatrix} 2 \\ -3 \end{bmatrix} \quad \vec{x_3} = \begin{bmatrix} .6 \\ .8 \end{bmatrix} \quad \vec{x_4} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

SOLUTION We apply Definition 31 to each vector.

$$\begin{aligned} ||\vec{x_1}|| &= \sqrt{1^2 + 1^2} = \sqrt{2}. \\ ||\vec{x_2}|| &= \sqrt{2^2 + (-3)^2} = \sqrt{13}. \\ ||\vec{x_3}|| &= \sqrt{.6^2 + .8^2} = \sqrt{.36 + .64} = 1. \\ ||\vec{x_4}|| &= \sqrt{3^2 + 0} = 3. \end{aligned}$$

Now that we know how to compute the length of a vector, let's revisit a statement we made as we explored Examples 60 and 61: "Multiplying a vector by a positive scalar c stretches the vectors by a factor of c ..." At that time, we did not know how to measure the length of a vector, so our statement was unfounded. In the following example, we will confirm the truth of our previous statement.

Example 66 Measuring the effect of scalar multiplication on vector length
Let $\vec{x} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Compute $||\vec{x}||$, $||3\vec{x}||$, $||-2\vec{x}||$, and $||c\vec{x}||$, where c is a scalar.

SOLUTION We apply Definition 31 to each of the vectors.

$||\vec{x}|| = \sqrt{4+1} = \sqrt{5}.$

Before computing the length of $||3\vec{x}||$, we note that $3\vec{x} = \begin{bmatrix} 6 \\ -3 \end{bmatrix}$.

$||3\vec{x}|| = \sqrt{36+9} = \sqrt{45} = 3\sqrt{5} = 3\,||\vec{x}||\,.$

Before computing the length of $||-2\vec{x}||$, we note that $-2\vec{x} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$.

$||-2\vec{x}|| = \sqrt{16+4} = \sqrt{20} = 2\sqrt{5} = 2\,||\vec{x}||\,.$

Finally, to compute $||c\vec{x}||$, we note that $c\vec{x} = \begin{bmatrix} 2c \\ -c \end{bmatrix}$. Thus:

$||c\vec{x}|| = \sqrt{(2c)^2 + (-c)^2} = \sqrt{4c^2 + c^2} = \sqrt{5c^2} = |c|\sqrt{5}.$

This last line is true because the square root of any number squared is the *absolute value* of that number (for example, $\sqrt{(-3)^2} = 3$).

The last computation of our example is the most important one. It shows that, in general, multiplying a vector $\vec{x}$ by a scalar c stretches $\vec{x}$ by a factor of $|c|$ (and the direction will change if c is negative). This is important so we'll make it a Theorem.

We can multiply a 3×2 matrix by a 2D vector and get a 3D vector back, and this gives very interesting results. See section 3.5.

Theorem 10 Vector Length and Scalar Multiplication

Let $\vec{x}$ be a vector and let c be a scalar. Then the length of $c\vec{x}$ is

$$||c\vec{x}|| = |c| \cdot ||\vec{x}|| \, .$$

Matrix $-$ Vector Multiplication

The last arithmetic operation to consider visualizing is matrix multiplication. Specifically, we want to visualize the result of multiplying a vector by a matrix. In order to multiply a 2D vector by a matrix and get a 2D vector back, our matrix must be a square, 2×2 matrix.

We'll start with an example. Given a matrix A and several vectors, we'll graph the vectors before and after they've been multiplied by A and see what we learn.

Example 67 **Multiplying a vector by a matrix**
Let A be a matrix, and $\vec{x}$, $\vec{y}$, and $\vec{z}$ be vectors as given below.

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \vec{z} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

Graph $\vec{x}$, $\vec{y}$ and $\vec{z}$, as well as $A\vec{x}$, $A\vec{y}$ and $A\vec{z}$.

SOLUTION

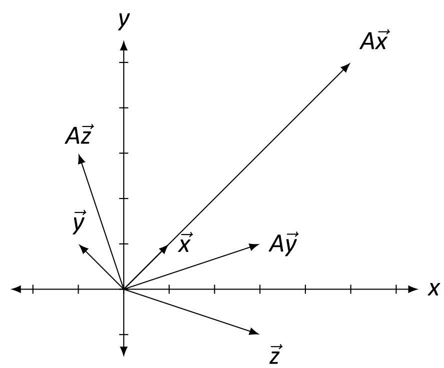


Figure 3.15: Multiplying vectors by a matrix in Example 66.

It is straightforward to compute:

$$A\vec{x} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}, \quad A\vec{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \text{and} \quad A\vec{z} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}.$$

The vectors are sketched in Figure 3.15

There are several things to notice. When each vector is multiplied by A, the result is a vector with a different length (in this example, always longer), and in two of the cases (for $\vec{y}$ and $\vec{z}$), the resulting vector points in a different direction.

This isn't surprising. In the previous section we learned about matrix multiplication, which is a strange and seemingly unpredictable operation. Would you expect to see some sort of immediately recognizable pattern appear from multiplying a matrix and a vector? (This is a rhetorical question; the expected answer is "No.") In fact, the surprising thing from the example is that $\vec{x}$ and $A\vec{x}$ point in the same direction! Why does the direction of $\vec{x}$ not change after multiplication by A? (We'll answer this in Section 8.1 when we learn about something called "eigenvectors.")

Different matrices act on vectors in different ways. (That's one reason we call them "different.") Some always increase the length of a vector through multiplication, others always decrease the length, others increase the length of some vectors and decrease the length of others, and others still don't change the length at all. A similar statement can be made about how matrices affect the direction of vectors through multiplication: some change every vector's direction, some change "most" vector's direction but leave some the same, and others still don't change the direction of any vector.

How do we set about studying how matrix multiplication affects vectors? We could just create lots of different matrices and lots of different vectors, multiply, then graph, but this would be a lot of work with very little useful result. It would be too hard to find a pattern of behaviour in this. (Remember, that's what mathematicians do. We look for patterns.)

Instead, we'll begin by using a technique we've employed often in the past. We have a "new" operation; let's explore how it behaves with "old" operations. Specifically, we know how to sketch vector addition. What happens when we throw matrix multiplication into the mix? Let's try an example.

Example 68 **Combining addition and matrix multiplication**
Let A be a matrix and $\vec{x}$ and $\vec{y}$ be vectors as given below.

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Sketch $\vec{x}+\vec{y}$, $A\vec{x}$, $A\vec{y}$, and $A(\vec{x}+\vec{y})$.

Solution It is pretty straightforward to compute:

$$\vec{x}+\vec{y} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}; \quad A\vec{x} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}; \quad A\vec{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad A(\vec{x}+\vec{y}) = \begin{bmatrix} 3 \\ 5 \end{bmatrix}.$$

In Figure 3.16, we have graphed the above vectors and have included dashed gray vectors to highlight the additive nature of $\vec{x}+\vec{y}$ and $A(\vec{x}+\vec{y})$. Does anything strike you as interesting?

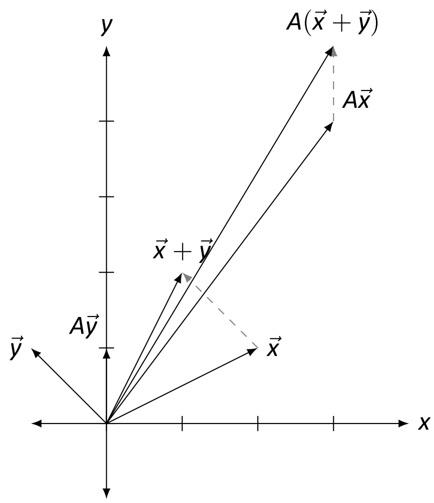


Figure 3.16: Vector addition and matrix multiplication in Example 67.

Let's not focus on things which don't matter right now: let's not focus on how long certain vectors became, nor necessarily how their direction changed. Rather, think about how matrix multiplication interacted with the vector addition.

In some sense, we started with three vectors, $\vec{x}$, $\vec{y}$, and $\vec{x}+\vec{y}$. This last vector is special; it is the sum of the previous two. Now, multiply all three by A. What happens? We get three new vectors, but the significant thing is this: the last vector is still the sum of the previous two! (We emphasize this by drawing dotted vectors to represent part of the Parallelogram Law.)

Of course, we knew this already: we already knew that $A\vec{x}+A\vec{y} = A(\vec{x}+\vec{y})$, for this is just the Distributive Property. However, now we get to see this graphically.

In Section 3.4 we'll study in greater depth how matrix multiplication affects vectors and the whole Cartesian plane. For now, we'll settle for simple practice: given a matrix and some vectors, we'll multiply and graph. Let's do one more example.

Example 69 Sketching the effect of matrix multiplicaiton

Let A, $\vec{x}$, $\vec{y}$, and $\vec{z}$ be as given below.

$$A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \vec{z} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

Graph $\vec{x}$, $\vec{y}$ and $\vec{z}$, as well as $A\vec{x}$, $A\vec{y}$ and $A\vec{z}$.

SOLUTION

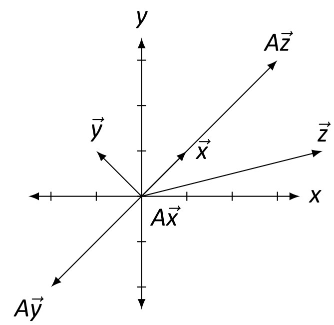


Figure 3.17: Multiplying vectors by a matrix in Example 68.

It is straightforward to compute:

$$A\vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad A\vec{y} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}, \quad \text{and} \quad A\vec{z} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

The vectors are sketched in Figure 3.17.

These results are interesting. While we won't explore them in great detail here, notice how $\vec{x}$ got sent to the zero vector. Notice also that $A\vec{x}$, $A\vec{y}$ and $A\vec{z}$ are all in a line (as well as $\vec{x}$!). Why is that? Are $\vec{x}$, $\vec{y}$ and $\vec{z}$ just special vectors, or would any other vector get sent to the same line when multiplied by A? (Don't just sit there, try it out!)

This section has focused on vectors in two dimensions. Later on in this book, we'll extend these ideas into three dimensions (3D).

In the next section we'll take a new idea (matrix multiplication) and apply it to an old idea (solving systems of linear equations). This will allow us to view an old idea in a new way – and we'll even get to "visualize" it.

Exercises 3.3

Problems

In Exercises 1 – 4, vectors $\vec{x}$ and $\vec{y}$ are given. Sketch $\vec{x}$, $\vec{y}$, $\vec{x}+\vec{y}$, and $\vec{x}-\vec{y}$ on the same Cartesian axes.

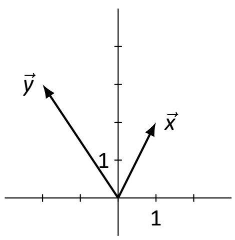
1. $\vec{x}=\begin{bmatrix}1\\1\end{bmatrix}$, $\vec{y}=\begin{bmatrix}-2\\3\end{bmatrix}$

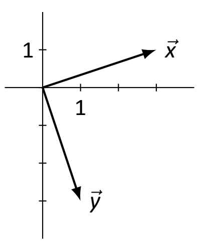
2. $\vec{x}=\begin{bmatrix}3\\1\end{bmatrix}$, $\vec{y}=\begin{bmatrix}1\\-2\end{bmatrix}$

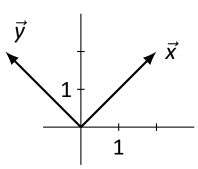
3. $\vec{x}=\begin{bmatrix}-1\\1\end{bmatrix}$, $\vec{y}=\begin{bmatrix}-2\\2\end{bmatrix}$

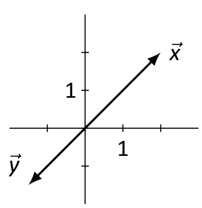
4. $\vec{x}=\begin{bmatrix}2\\0\end{bmatrix}$, $\vec{y}=\begin{bmatrix}1\\3\end{bmatrix}$

In Exercises 5 – 8, vectors $\vec{x}$ and $\vec{y}$ are drawn. Sketch $2\vec{x}$, $-\vec{y}$, $\vec{x}+\vec{y}$, and $\vec{x}-\vec{y}$ on the same Cartesian axes.

5. 

6. 

7. 

8. 

In Exercises 9 – 12, a vector $\vec{x}$ and a scalar a are given. Using Definition 31, compute the lengths of $\vec{x}$ and $a\vec{x}$, then compare these lengths.

9. $\vec{x}=\begin{bmatrix}2\\1\end{bmatrix}$, $a=3$.

10. $\vec{x}=\begin{bmatrix}4\\7\end{bmatrix}$, $a=-2$.

11. $\vec{x}=\begin{bmatrix}-3\\5\end{bmatrix}$, $a=-1$.

12. $\vec{x}=\begin{bmatrix}3\\-9\end{bmatrix}$, $a=\frac{1}{3}$.

13. Four pairs of vectors $\vec{x}$ and $\vec{y}$ are given below. For each pair, compute $||\vec{x}||$, $||\vec{y}||$, and $||\vec{x}+\vec{y}||$. Use this information to answer: Is it always, sometimes, or never true that $||\vec{x}||+||\vec{y}||=||\vec{x}+\vec{y}||$? If it always or never true, explain why. If it is sometimes true, explain when it is true.

    (a) $\vec{x}=\begin{bmatrix}1\\1\end{bmatrix}$, $\vec{y}=\begin{bmatrix}2\\3\end{bmatrix}$

    (b) $\vec{x}=\begin{bmatrix}1\\-2\end{bmatrix}$, $\vec{y}=\begin{bmatrix}3\\-6\end{bmatrix}$

    (c) $\vec{x}=\begin{bmatrix}-1\\3\end{bmatrix}$, $\vec{y}=\begin{bmatrix}2\\5\end{bmatrix}$

    (d) $\vec{x}=\begin{bmatrix}2\\1\end{bmatrix}$, $\vec{y}=\begin{bmatrix}-4\\-2\end{bmatrix}$

In Exercises 14 – 17, a matrix A is given. Sketch $\vec{x}$, $\vec{y}$, $A\vec{x}$ and $A\vec{y}$ on the same Cartesian axes, where

$$\vec{x}=\begin{bmatrix}1\\1\end{bmatrix}\quad\textbf{and}\quad\vec{y}=\begin{bmatrix}-1\\2\end{bmatrix}.$$

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

15. $A=\begin{bmatrix}2&0\\-1&3\end{bmatrix}$

16. $A=\begin{bmatrix}1&1\\1&1\end{bmatrix}$

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

We already looked at the basics of graphing vectors. In this chapter, we'll explore these ideas more fully. One often gains a better understanding of a concept by "seeing" it. For instance, one can study the function $f(x) = x^2$ and describe many properties of how the output relates to the input without producing a graph, but the graph can quickly bring meaning and insight to equations and formulae. Not only that, but the study of graphs of functions is in itself a wonderful mathematical world, worthy of exploration.

We've studied the graphing of vectors; in this chapter we'll take this a step further and study some fantastic graphical properties of vectors and matrix arithmetic. We mentioned earlier that these concepts form the basis of computer graphics; in this chapter, we'll see even better how that is true.

3.4 Transformations of the Cartesian Plane

AS YOU READ . . .

1. To understand how the Cartesian plane is affected by multiplication by a matrix, it helps to study how what is affected?

2. Transforming the Cartesian plane through matrix multiplication transforms straight lines into what kind of lines?

3. T/F: If one draws a picture of a sheep on the Cartesian plane, then transformed the plane using the matrix

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

one could say that the sheep was "sheared."

We studied in Section 3.3 how to visualize vectors and how certain matrix arithmetic operations can be graphically represented. We limited our visual understanding of matrix multiplication to graphing a vector, multiplying it by a matrix, then graphing the resulting vector. In this section we'll explore these multiplication ideas in greater depth. Instead of multiplying individual vectors by a matrix A, we'll study what happens when we multiply *every* vector in the Cartesian plans by A. (No, we won't do them one by one.)

Because of the Distributive Property, demonstrated way back in Example 67, we can say that the Cartesian plane will be *transformed* in a very nice, predictable way. Straight lines will be transformed into other straight lines (and they won't become curvy, or jagged, or broken). Curved lines will be transformed into other curved lines (perhaps the curve will become "straight," but it won't become jagged or broken).

One way of studying how the whole Cartesian plane is affected by multiplication by a matrix A is to study how the *unit square* is affected. The unit square is the square with corners at the points $(0, 0)$, $(1, 0)$, $(1, 1)$, and $(0, 1)$. Each corner can be represented by the vector that points to it; multiply each of these vectors by A and we can get an idea of how A affects the whole Cartesian plane.

Let's try an example.

Example 70 **Visualizing a matrix transformation using vectors**
Plot the vectors of the unit square before and after they have been multiplied by A, where

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

Solution The four corners of the unit square can be represented by the vectors

$$\begin{bmatrix}0\\0\end{bmatrix},\quad \begin{bmatrix}1\\0\end{bmatrix},\quad \begin{bmatrix}1\\1\end{bmatrix},\quad \begin{bmatrix}0\\1\end{bmatrix}.$$

Multiplying each by A gives the vectors

$$\begin{bmatrix}0\\0\end{bmatrix},\quad \begin{bmatrix}1\\2\end{bmatrix},\quad \begin{bmatrix}5\\5\end{bmatrix},\quad \begin{bmatrix}4\\3\end{bmatrix},$$

respectively.

(Hint: one way of using your calculator to do this for you quickly is to make a 2×4 matrix whose columns are each of these vectors. In this case, create a matrix

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

Then multiply B by A and read off the transformed vectors from the respective columns:

$$AB = \begin{bmatrix}0 & 1 & 5 & 4\\0 & 2 & 5 & 3\end{bmatrix}.$$

This saves time, especially if you do a similar procedure for multiple matrices A. Of course, we can save more time by skipping the first column; since it is the column of zeros, it will stay the column of zeros after multiplication by A.)

The unit square and its transformation are graphed in Figure 3.18, where the shaped vertices correspond to each other across the two graphs. Note how the square got turned into some sort of quadrilateral (it's actually a parallelogram). A really interesting thing is how the triangular and square vertices seem to have changed places – it is as though the square, in addition to being stretched out of shape, was flipped.

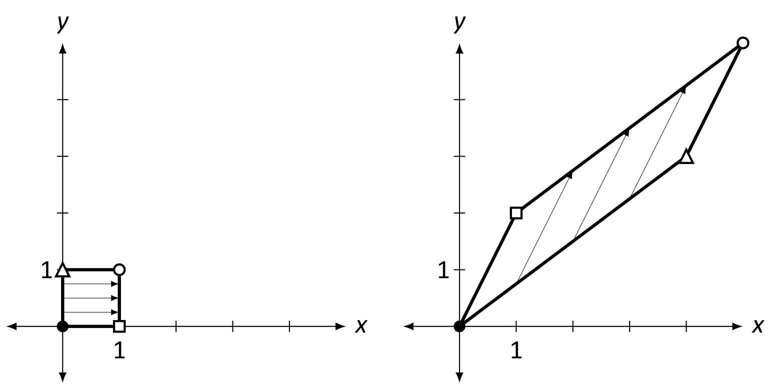


Figure 3.18: Transforming the unit square by matrix multiplication in Example 69.

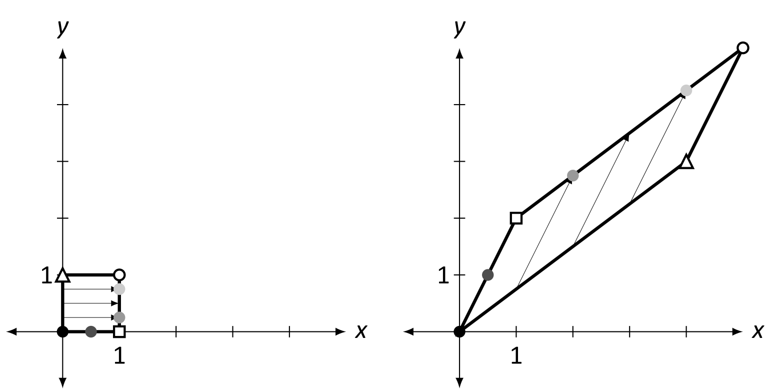


Figure 3.19: Emphasizing straight lines going to straight lines in Example 69.

To stress how "straight lines get transformed to straight lines," consider Figure 3.19. Here, the unit square has some additional points drawn on it which correspond to the shaded dots on the transformed parallelogram. Note how relative distances are also preserved; the dot halfway between the black and square dots is transformed to a position along the line, halfway between the black and square dots.

Much more can be said about this example. Before we delve into this, though, let's try one more example.

Example 71 **Visualizing a matrix transformation using a region**
Plot the transformed unit square after it has been transformed by A, where

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

Solution We'll put the vectors that correspond to each corner in a matrix B as before and then multiply it on the left by A. Doing so gives:

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

In Figure 3.20 the unit square is again drawn along with its transformation by A.

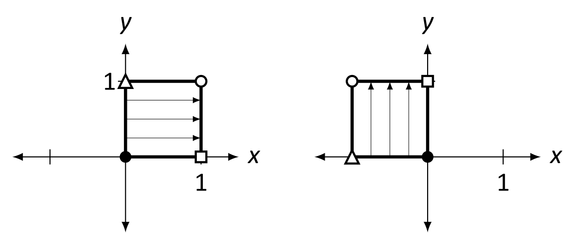


Figure 3.20: Transforming the unit square by matrix multiplication in Example 70.

Make note of how the square moved. It did not simply "slide" to the left; (mathematically, that is called a *translation*) nor did it "flip" across the y axis. Rather, it was *rotated* counterclockwise about the origin 90°. In a rotation, the shape of an object does not change; in our example, the square remained a square of the same size.

We have broached the topic of how the Cartesian plane can be transformed via multiplication by a 2×2 matrix A. We have seen two examples so far, and our intuition as to how the plane is changed has been informed only by seeing how the unit square changes. Let's explore this further by investigating two questions:

1. Suppose we want to transform the Cartesian plane in a known way (for instance, we may want to rotate the plane counterclockwise 180°). How do we find the matrix (if one even exists) which performs this transformation?

2. How does knowing how the unit square is transformed really help in understanding how the entire plane is transformed?

These questions are closely related, and as we answer one, we will help answer the other.

To get started with the first question, look back at Examples 69 and 70 and consider again how the unit square was transformed. In particular, is there any correlation between where the vertices ended up and the matrix A?

If you are just reading on, and haven't actually gone back and looked at the examples, go back now and try to make some sort of connection. Otherwise – you may have noted some of the following things:

1. The zero vector ($\vec{0}$, the "black" corner) never moved. That makes sense, though; $A\vec{0} = \vec{0}$.
2. The "square" corner, i.e., the corner corresponding to the vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, is always transformed to the vector in the first column of A!
3. Likewise, the "triangular" corner, i.e., the corner corresponding to the vector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, is always transformed to the vector in the second column of A![1]
4. The "white dot" corner is always transformed to the *sum* of the two column vectors of A.[2]

Let's now take the time to understand these four points. The first point should be clear; $\vec{0}$ will always be transformed to $\vec{0}$ via matrix multiplication. (Hence the hint in the middle of Example 69, where we are told that we can ignore entering in the column of zeros in the matrix B.)

We can understand the second and third points simultaneously. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \vec{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

What are $A\vec{e_1}$ and $A\vec{e_2}$?

$$\begin{aligned} A\vec{e_1} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} a \\ c \end{bmatrix} \end{aligned}$$

$$\begin{aligned} A\vec{e_2} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} b \\ d \end{bmatrix} \end{aligned}$$

So by mere mechanics of matrix multiplication, the square corner $\vec{e_1}$ is transformed to the first column of A, and the triangular corner $\vec{e_2}$ is transformed to the second column of A. A similar argument demonstrates why the white dot corner is transformed to the sum of the columns of A.[3]

[1] Although this is less of a surprise, given the result of the previous point.

[2] This observation is a bit more obscure than the first three. It follows from the fact that this corner of the unit square is the "sum" of the other two nonzero corners.

[3] Another way of looking at all of this is to consider what $A \cdot I$ is: of course, it is just A. What are the columns of I? Just $\vec{e_1}$ and $\vec{e_2}$.

Revisit now the question "How do we find the matrix that performs a given transformation on the Cartesian plane?" The answer follows from what we just did. Think about the given transformation and how it would transform the corners of the unit square. Make the first column of A the vector where $\vec{e_1}$ goes, and make the second column of A the vector where $\vec{e_2}$ goes.

Let's practice this in the context of an example.

Example 72 Determining a matrix transformation

Find the matrix A that flips the Cartesian plane about the x axis and then stretches the plane horizontally by a factor of two.

Solution We first consider $\vec{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Where does this corner go to under the given transformation? Flipping the plane across the x axis does not change $\vec{e_1}$ at all; stretching the plane sends $\vec{e_1}$ to $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$. Therefore, the first column of A is $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$.

Now consider $\vec{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Flipping the plane about the x axis sends $\vec{e_2}$ to the vector $\begin{bmatrix} 0 \\ -1 \end{bmatrix}$; subsequently stretching the plane horizontally does not affect this vector. Therefore the second column of A is $\begin{bmatrix} 0 \\ -1 \end{bmatrix}$.

Putting this together gives

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

To help visualize this, consider Figure 3.21 where a shape is transformed under this matrix. Notice how it is turned upside down and is stretched horizontally by a factor of two. (The gridlines are given as a visual aid.)

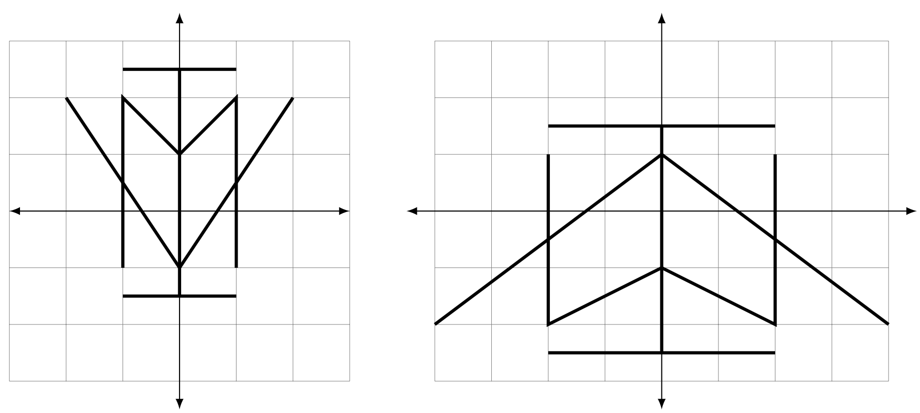


Figure 3.21: Transforming the Cartesian plane in Example 71

A while ago we asked two questions. The first was "How do we find the matrix that performs a given transformation?" We have just answered that question (although we will do more to explore it in the future). The second question was "How does knowing how the unit square is transformed really help us understand how the entire plane is transformed?"

Consider Figure 3.22 where the unit square (with vertices marked with shapes as before) is shown transformed under an unknown matrix. How does this help us understand how the whole Cartesian plane is transformed? For instance, how can we use this picture to figure out how the point $(2, 3)$ will be transformed?

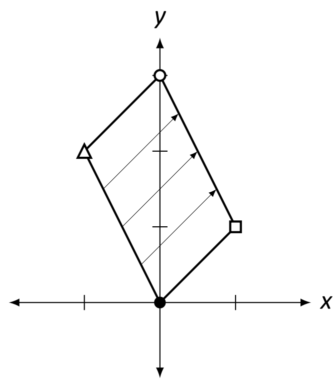


Figure 3.22: The unit square under an unknown transformation.

There are two ways to consider the solution to this question. First, we know now how to compute the transformation matrix; the new position of $\vec{e_1}$ is the first column of A, and the new position of $\vec{e_2}$ is the second column of A. Therefore, by looking at the figure, we can deduce that

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}.$$[4]

To find where the point $(2, 3)$ is sent, simply multiply

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

There is another way of doing this which isn't as computational – it doesn't involve computing the transformation matrix. Consider the following equalities:

$$\begin{aligned} \begin{bmatrix} 2 \\ 3 \end{bmatrix} &= \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix} \\ &= 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= 2\vec{e_1} + 3\vec{e_2} \end{aligned}$$

This last equality states something that is somewhat obvious: to arrive at the vector $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$, one needs to go 2 units in the $\vec{e_1}$ direction and 3 units in the $\vec{e_2}$ direction. To find where the point $(2, 3)$ is transformed, one needs to go 2 units in the *new* $\vec{e_1}$ direction and 3 units in the *new* $\vec{e_2}$ direction. This is demonstrated in Figure 3.23.

[4]At least, A is close to that. The square corner could actually be at the point $(1.01, .99)$.

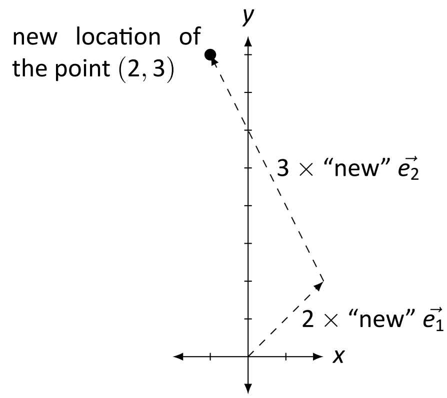


Figure 3.23: Finding the new location of the point $(2,3)$.

We are coming to grips with how matrix transformations work. We asked two basic questions: "How do we find the matrix for a given transformation?" and "How do we understand the transformation without the matrix?", and we've answered each accompanied by one example. Let's do another example that demonstrates both techniques at once.

Example 73 **Determining and analyzing a matrix transformation**
First, find the matrix A that transforms the Cartesian plane by stretching it vertically by a factor of 1.5, then stretches it horizontally by a factor of 0.5, then rotates it clockwise about the origin 90°. Secondly, using the new locations of $\vec{e_1}$ and $\vec{e_2}$, find the transformed location of the point $(-1,2)$.

Solution To find A, first consider the new location of $\vec{e_1}$. Stretching the plane vertically does not affect $\vec{e_1}$; stretching the plane horizontally by a factor of 0.5 changes $\vec{e_1}$ to $\begin{bmatrix} 1/2 \\ 0 \end{bmatrix}$, and then rotating it 90° about the origin moves it to $\begin{bmatrix} 0 \\ -1/2 \end{bmatrix}$. This is the first column of A.

Now consider the new location of $\vec{e_2}$. Stretching the plane vertically changes it to $\begin{bmatrix} 0 \\ 3/2 \end{bmatrix}$; stretching horizontally does not affect it, and rotating 90° moves it to $\begin{bmatrix} 3/2 \\ 0 \end{bmatrix}$. This is then the second column of A. This gives

$$A = \begin{bmatrix} 0 & 3/2 \\ -1/2 & 0 \end{bmatrix}.$$

Where does the point $(-1,2)$ get sent to? The corresponding vector $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ is found by going -1 units in the $\vec{e_1}$ direction and 2 units in the $\vec{e_2}$ direction. Therefore, the transformation will send the vector to -1 units in the new $\vec{e_1}$ direction and 2 units in the new $\vec{e_2}$ direction. This is sketched in Figure 3.24, along with the transformed unit square. We can also check this multiplicatively:

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

Figure 3.25 shows the effects of the transformation on another shape.

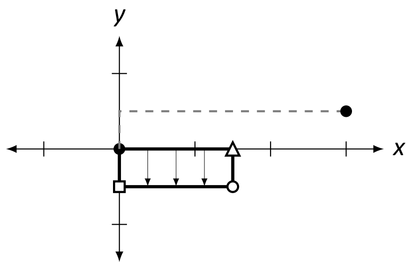


Figure 3.24: Understanding the transformation in Example 72.

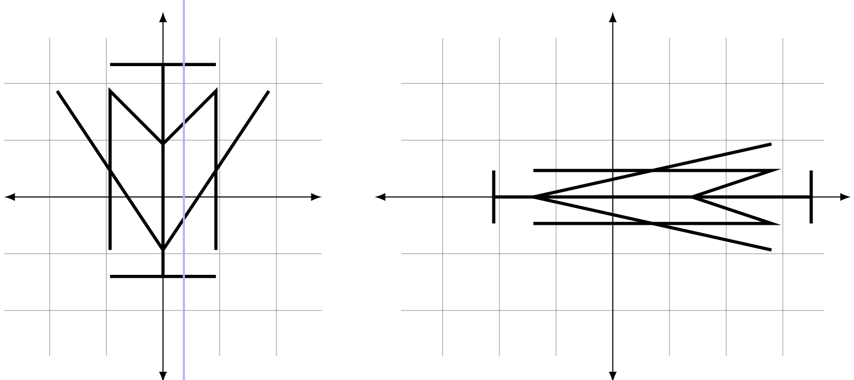


Figure 3.25: Transforming the Cartesian plane in Example 72

Right now we are focusing on transforming the Cartesian plane – we are making 2D transformations. Knowing how to do this provides a foundation for transforming 3D space,[5] which, among other things, is very important when producing 3D computer graphics. Basic shapes can be drawn and then rotated, stretched, and/or moved to other regions of space. This also allows for things like "moving the camera view."

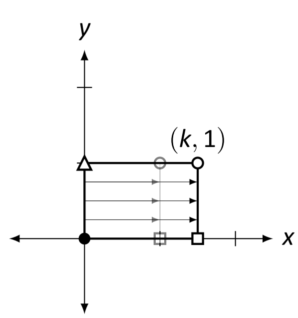
What kinds of transformations are possible? We have already seen some of the things that are possible: rotations, stretches, and flips. We have also mentioned some things that are not possible. For instance, we stated that straight lines always get transformed to straight lines. Therefore, we cannot transform the unit square into a circle using a matrix.

Let's look at some common transformations of the Cartesian plane and the matrices that perform these operations. In the following figures, a transformation matrix will be given alongside a picture of the transformed unit square. (The original unit square is drawn lightly as well to serve as a reference.)

2D Matrix Transformations

Horizontal stretch by a factor of k.

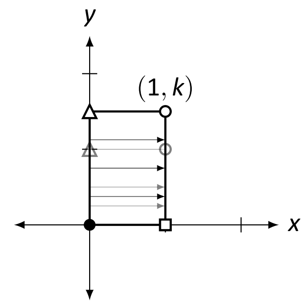
$$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$$



[5]Actually, it provides a foundation for doing it in 4D, 5D, . . ., 17D, etc. Those are just harder to visualize.

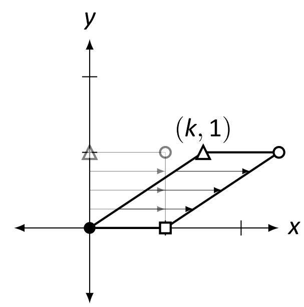
Vertical stretch by a factor of k.

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



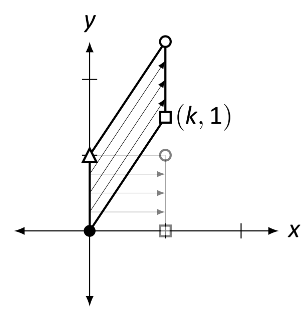
Horizontal shear by a factor of k.

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



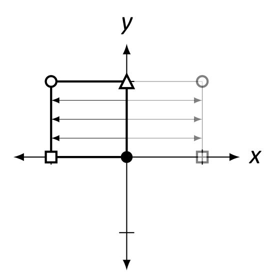
Vertical shear by a factor of k.

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



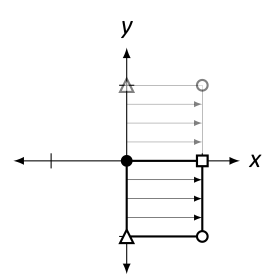
Horizontal reflection across the y axis.

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



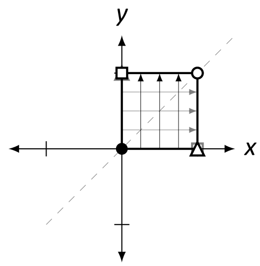
Vertical reflection across the x axis.

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



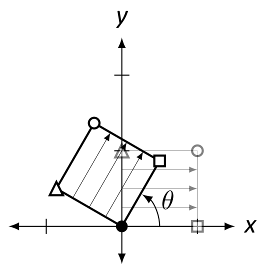
Diagonal reflection across the line $y = x$.

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



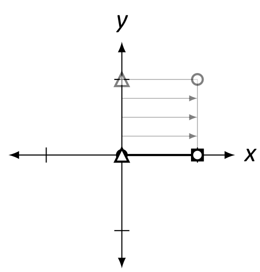
Rotation around the origin by an angle of θ.

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



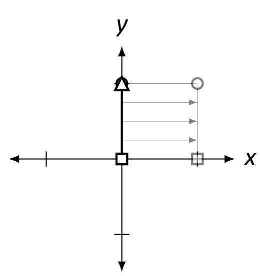
Projection onto the x axis.
(Note how the square is "squashed" down onto the x-axis.)

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



Projection onto the y axis.
(Note how the square is "squashed" over onto the y-axis.)

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



Now that we have seen a healthy list of transformations that we can perform on the Cartesian plane, let's practice a few more times creating the matrix that gives the desired transformation. In the following example, we develop our understanding one more critical step.

Example 74 **Determining the matrix of a transformation**
Find the matrix A that transforms the Cartesian plane by performing the following operations in order:

1. Vertical shear by a factor of 0.5
2. Counterclockwise rotation about the origin by an angle of $\theta = 30^\circ$
3. Horizontal stretch by a factor of 2
4. Diagonal reflection across the line $y = x$

Solution Wow! We already know how to do this – sort of. We know we can find the columns of A by tracing where $\vec{e_1}$ and $\vec{e_2}$ end up, but this also seems difficult. There is so much that is going on. Fortunately, we can accomplish what we need without much difficulty by being systematic.

First, let's perform the vertical shear. The matrix that performs this is

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix}.$$

After that, we want to rotate everything clockwise by 30°. To do this, we use

$$A_2 = \begin{bmatrix} \cos 30^\circ & -\sin 30^\circ \\ \sin 30^\circ & \cos 30^\circ \end{bmatrix} = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}.$$

In order to do both of these operations, in order, we multiply A_2A_1.[6]

To perform the final two operations, we note that

$$A_3 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A_4 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

perform the horizontal stretch and diagonal reflection, respectively. Thus to perform all of the operations "at once," we need to multiply by

$$\begin{aligned} A &= A_4A_3A_2A_1 \\ &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix} \\ &= \begin{bmatrix} (\sqrt{3}+2)/4 & \sqrt{3}/2 \\ (2\sqrt{3}-1)/2 & -1 \end{bmatrix} \\ &\approx \begin{bmatrix} 0.933 & 0.866 \\ 1.232 & -1 \end{bmatrix}. \end{aligned}$$

Let's consider this closely. Suppose I want to know where a vector $\vec{x}$ ends up. We claim we can find the answer by multiplying $A\vec{x}$. Why does this work? Consider:

$$\begin{aligned} A\vec{x} &= A_4A_3A_2A_1\vec{x} & \\ &= A_4A_3A_2(A_1\vec{x}) & \text{(performs the vertical shear)} \\ &= A_4A_3(A_2\vec{x}_1) & \text{(performs the rotation)} \\ &= A_4(A_3\vec{x}_2) & \text{(performs the horizontal stretch)} \\ &= A_4\vec{x}_3 & \text{(performs the diagonal reflection)} \\ &= \vec{x}_4 & \text{(the result of transforming } \vec{x}\text{)} \end{aligned}$$

Most readers are not able to visualize exactly what the given list of operations does to the Cartesian plane. In Figure 3.26 we sketch the transformed unit square; in Figure 3.27 we sketch a shape and its transformation.

[6] The reader might ask, "Is it important to do multiply these in that order? Could we have multiplied A_1A_2 instead?" Our answer starts with "Is matrix multiplication commutative?" The answer to our question is "No," so the answers to the reader's questions are "Yes" and "No," respectively.

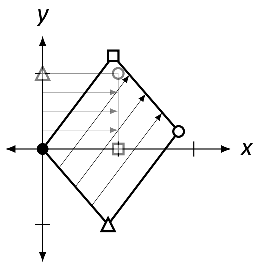


Figure 3.26: The transformed unit square in Example 73.

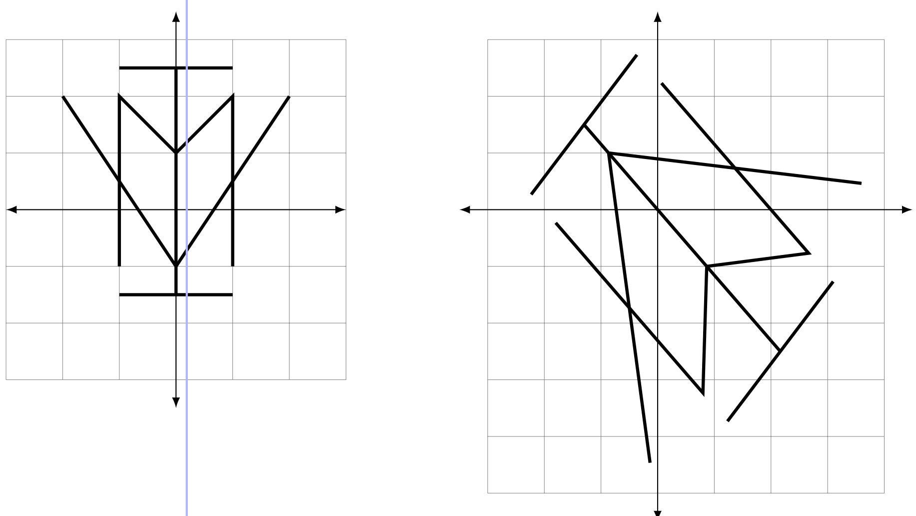


Figure 3.27: A transformed shape in Example 73.

Once we know what matrices perform the basic transformations,[7] performing complex transformations on the Cartesian plane really isn't that . . . complex. It boils down to multiplying by a series of matrices.

We've shown many examples of transformations that we can do, and we've mentioned just a few that we can't – for instance, we can't turn a square into a circle. Why not? Why is it that straight lines get sent to straight lines? We spent a lot of time within this text looking at invertible matrices; what connections, if any,[8] are there between invertible matrices and their transformations on the Cartesian plane?

All these questions require us to think like mathematicians – we are being asked to study the *properties* of an object we just learned about and their connections to things we've already learned. We'll do all this (and more!) in the following section.

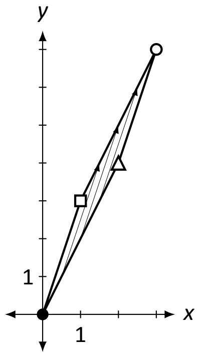
[7] or know where to find them

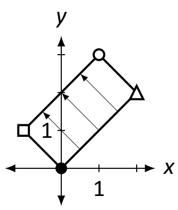
[8] By now, the reader should expect connections to exist.

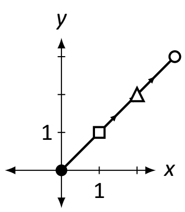
Exercises 3.4

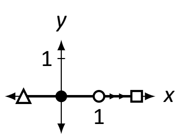
Problems

In Exercises 1 – 4, a sketch of transformed unit square is given. Find the matrix A that performs this transformation.

1. 

2. 

3. 

4. 

In Exercises 5 – 10, a list of transformations is given. Find the matrix A that performs those transformations, in order, on the Cartesian plane.

5. (a) vertical shear by a factor of 2
   (b) horizontal shear by a factor of 2

6. (a) horizontal shear by a factor of 2
   (b) vertical shear by a factor of 2

7. (a) horizontal stretch by a factor of 3
   (b) reflection across the line $y = x$

8. (a) counterclockwise rotation by an angle of 45°
   (b) vertical stretch by a factor of $1/2$

9. (a) clockwise rotation by an angle of 90°
   (b) horizontal reflection across the y axis
   (c) vertical shear by a factor of 1

10. (a) vertical reflection across the x axis
    (b) horizontal reflection across the y axis
    (c) diagonal reflection across the line $y = x$

In Exercises 11 – 14, two sets of transformations are given. Sketch the transformed unit square under each set of transformations. Are the transformations the same? Explain why/why not.

11. (a) a horizontal reflection across the y axis, followed by a vertical reflection across the x axis, compared to
    (b) a counterclockise rotation of 180°

12. (a) a horizontal stretch by a factor of 2 followed by a reflection across the line $y = x$, compared to
    (b) a vertical stretch by a factor of 2

13. (a) a horizontal stretch by a factor of 1/2 followed by a vertical stretch by a factor of 3, compared to
    (b) the same operations but in opposite order

14. (a) a reflection across the line $y = x$ followed by a reflection across the x axis, compared to
    (b) a reflection across the the y axis, followed by a reflection across the line $y = x$.

3.5 Properties of Linear Transformations

AS YOU READ . . .

1. T/F: Translating the Cartesian plane 2 units up is a linear transformation.
2. T/F: If T is a linear transformation, then $T(\vec{0}) = \vec{0}$.

In the previous section we discussed standard transformations of the Cartesian plane – rotations, reflections, etc. As a motivational example for this section's study, let's consider another transformation – let's find the matrix that moves the unit square one unit to the right (see Figure 3.28). This is called a *translation*.

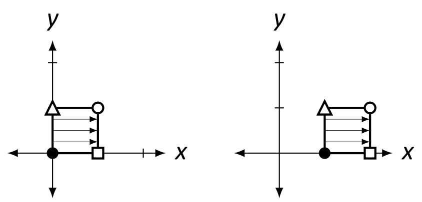


Figure 3.28: Translating the unit square one unit to the right.

Our work from the previous section allows us to find the matrix quickly. By looking at the picture, it is easy to see that $\vec{e_1}$ is moved to $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$ and $\vec{e_2}$ is moved to $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Therefore, the transformation matrix should be

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

However, look at Figure 3.29 where the unit square is drawn after being transformed by A. It is clear that we did not get the desired result; the unit square was not translated, but rather stretched/sheared in some way.

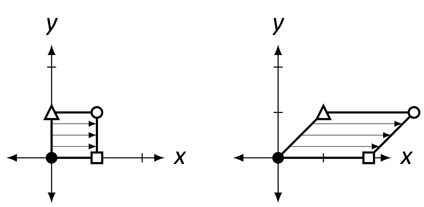


Figure 3.29: Actual transformation of the unit square by matrix A.

We used T instead of f to define the function $T(\vec{x}) = A\vec{x}$ to help differentiate it from "regular" functions. "Normally" functions are defined using lower case letters when the input is a number; when the input is a vector, we use upper case letters. (It also appears to be tradition to use the letter T to describe linear transformations, and mathematicians are suckers for tradition.)

What did we do wrong? We will answer this question, but first we need to develop a few thoughts and vocabulary terms.

We've been using the term "transformation" to describe how we've changed vectors. In fact, "transformation" is synonymous to "function." We are used to functions like $f(x) = x^2$, where the input is a number and the output is another number. In the previous section, we learned about transformations (functions) where the input was a vector and the output was another vector. If A is a "transformation matrix," then we could create a function of the form $T(\vec{x}) = A\vec{x}$. That is, a vector $\vec{x}$ is the input, and the output is $\vec{x}$ multiplied by A.

When we defined $f(x) = x^2$ above, we let the reader assume that the input was indeed a number. If we wanted to be complete, we should have stated

$$f : \mathbb{R} \to \mathbb{R} \quad \text{where} \quad f(x) = x^2.$$

The first part of that line told us that the input was a real number (that was the first $\mathbb{R}$) and the output was also a real number (the second $\mathbb{R}$).

To define a transformation where a 2D vector is transformed into another 2D vector via multiplication by a 2×2 matrix A, we should write

$$T : \mathbb{R}^2 \to \mathbb{R}^2 \quad \text{where} \quad T(\vec{x}) = A\vec{x}.$$

Here, the first $\mathbb{R}^2$ means that we are using 2D vectors as our input, and the second $\mathbb{R}^2$ means that a 2D vector is the output.

Consider a quick example:

$$T : \mathbb{R}^2 \to \mathbb{R}^3 \quad \text{where} \quad T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1^2 \\ 2x_1 \\ x_1x_2 \end{bmatrix}.$$

Notice that this takes 2D vectors as input and returns 3D vectors as output. For instance,

$$T\left(\begin{bmatrix} 3 \\ -2 \end{bmatrix}\right) = \begin{bmatrix} 9 \\ 6 \\ -6 \end{bmatrix}.$$

We now define a special type of transformation (function).

Definition 32 Linear Transformation

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transformation* if it satisfies the following two properties:

1. $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$ for all vectors $\vec{x}$ and $\vec{y}$, and
2. $T(k\vec{x}) = kT(\vec{x})$ for all vectors $\vec{x}$ and all scalars k.

If T is a linear transformation, it is often said that "T is *linear*."

Let's learn about this definition through some examples.

Example 75 Identifying linear transformations

Determine whether or not the transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation, where

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1^2 \\ 2x_1 \\ x_1x_2 \end{bmatrix}.$$

Solution We'll arbitrarily pick two vectors $\vec{x}$ and $\vec{y}$:

$$\vec{x} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

Let's check to see if T is linear by using the definition.

1. Is $T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$? First, compute $\vec{x}+\vec{y}$:

$$\vec{x}+\vec{y} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} + \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

Now compute $T(\vec{x})$, $T(\vec{y})$, and $T(\vec{x}+\vec{y})$:

$$T(\vec{x}) = T\left(\begin{bmatrix} 3 \\ -2 \end{bmatrix}\right) = \begin{bmatrix} 9 \\ 6 \\ -6 \end{bmatrix} \qquad T(\vec{y}) = T\left(\begin{bmatrix} 1 \\ 5 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} \qquad T(\vec{x}+\vec{y}) = T\left(\begin{bmatrix} 4 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 16 \\ 8 \\ 12 \end{bmatrix}$$

Is $T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$?

$$\begin{bmatrix} 9 \\ 6 \\ -6 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} \overset{!}{\neq} \begin{bmatrix} 16 \\ 8 \\ 12 \end{bmatrix}.$$

Therefore, T is *not* a linear transformation.

So we have an example of something that *doesn't* work. Let's try an example where things *do* work.

Example 76 Identifying linear transformations

Determine whether or not the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation, where $T(\vec{x}) = A\vec{x}$ and

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

It's important to remember the following principle of logic: to show that something doesn't work, we just need to show one case where it fails, which we did in Example 74. To show that something *always* works, we need to show it works for *all* cases – simply showing it works for a few cases isn't enough. (An example is not a proof.) However, doing so can be helpful in understanding the situation better.

Solution Let's start by again considering arbitrary $\vec{x}$ and $\vec{y}$. Let's choose the same $\vec{x}$ and $\vec{y}$ from Example 74.

$$\vec{x} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \quad \text{and} \quad \vec{y} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

If the lineararity properties hold for these vectors, then *maybe* it is actually linear (and we'll do more work).

1. Is $T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$? Recall:

$$\vec{x}+\vec{y} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

Now compute $T(\vec{x})$, $T(\vec{y})$, and $T(\vec{x}) + T(\vec{y})$:

$$T(\vec{x}) = T\left(\begin{bmatrix} 3 \\ -2 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad T(\vec{y}) = T\left(\begin{bmatrix} 1 \\ 5 \end{bmatrix}\right) = \begin{bmatrix} 11 \\ 23 \end{bmatrix} \qquad T(\vec{x}+\vec{y}) = T\left(\begin{bmatrix} 4 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 10 \\ 24 \end{bmatrix}$$

Is $T(\vec{x}+\vec{y}) = T(\vec{x}) + T(\vec{y})$?

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 11 \\ 23 \end{bmatrix} \overset{!}{=} \begin{bmatrix} 10 \\ 24 \end{bmatrix}.$$

So far, so good: $T(\vec{x}+\vec{y})$ is equal to $T(\vec{x}) + T(\vec{y})$.

2. Is $T(k\vec{x}) = kT(\vec{x})$? Let's arbitrarily pick $k = 7$, and use $\vec{x}$ as before.

$$\begin{aligned} T(7\vec{x}) &= T\left(\begin{bmatrix} 21 \\ -14 \end{bmatrix}\right) \\ &= \begin{bmatrix} -7 \\ 7 \end{bmatrix} \\ &= 7\begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= 7 \cdot T(\vec{x}) \quad ! \end{aligned}$$

So far it *seems* that T is indeed linear, for it worked in one example with arbitrarily chosen vectors and scalar. Now we need to try to show it is always true.

Consider $T(\vec{x}+\vec{y})$. By the definition of T, we have

$$T(\vec{x}+\vec{y}) = A(\vec{x}+\vec{y}).$$

By Theorem 9, part 2 (on page 95) we state that the Distributive Property holds for matrix multiplication. (Recall that a vector is just a special type of matrix, so this theorem applies to matrix–vector multiplication as well.) So $A(\vec{x}+\vec{y}) = A\vec{x}+A\vec{y}$. Recognize now that this last part is just $T(\vec{x}) + T(\vec{y})$! We repeat the above steps, all together:

$$\begin{aligned} T(\vec{x}+\vec{y}) &= A(\vec{x}+\vec{y}) && \text{(by the definition of } T \text{ in this example)} \\ &= A\vec{x}+A\vec{y} && \text{(by the Distributive Property)} \\ &= T(\vec{x})+T(\vec{y}) && \text{(again, by the definition of } T\text{)} \end{aligned}$$

Therefore, no matter what $\vec{x}$ and $\vec{y}$ are chosen, $T(\vec{x}+\vec{y}) = T(\vec{x})+T(\vec{y})$. Thus the first part of the linearity definition is satisfied.

The second part is satisfied in a similar fashion. Let k be a scalar, and consider:

$$\begin{aligned} T(k\vec{x}) &= A(k\vec{x}) && \text{(by the definition of } T \text{ is this example)} \\ &= kA\vec{x} && \text{(by Theorem 9 part 3)} \\ &= kT(\vec{x}) && \text{(again, by the definition of } T\text{)} \end{aligned}$$

Since T satisfies both parts of the definition, we conclude that T is a linear transformation.

We have seen two examples of transformations so far, one which was not linear and one that was. One might wonder "Why is linearity important?", which we'll address shortly.

First, consider how we proved the transformation in Example 75 was linear. We defined T by matrix multiplication, that is, $T(\vec{x}) = A\vec{x}$. We proved T was linear using properties of matrix multiplication – *we never considered the specific values of A!* That is, we didn't just choose a good matrix for T; *any* matrix A would have worked. This leads us to an important theorem. The first part we have essentially just proved; the second part we won't prove, although its truth is very powerful.

Theorem 11 Matrices and Linear Transformations

1. Define $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T(\vec{x}) = A\vec{x}$, where A is an $m \times n$ matrix. Then T is a linear transformation.
2. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be any linear transformation. Then there exists an unique $m \times n$ matrix A such that $T(\vec{x}) = A\vec{x}$.

The second part of the theorem says that *all* linear transformations can be described using matrix multiplication. Given *any* linear transformation, there is a matrix that completely defines that transformation. This important matrix gets its own name.

Definition 33 Standard Matrix of a Linear Transformation

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. By Theorem 11, there is a matrix A such that $T(\vec{x}) = A\vec{x}$. This matrix A is called the *standard matrix of the linear transformation T*, and is denoted $[\,T\,]$.

While exploring all of the ramifications of Theorem 11 is outside the scope of this text, let it suffice to say that since 1) linear transformations are very, very important in economics, science, engineering and mathematics, and 2) the theory of matrices is well developed and easy to implement by hand and on computers, then 3) it is great news that these two concepts go hand in hand.

The matrix–like brackets around T are intended to suggest that the standard matrix A is a matrix "with T inside."

We have already used the second part of this theorem in a small way. In the previous section we looked at transformations graphically and found the matrices that produced them. At the time, we didn't realize that these transformations were linear, but indeed they were.

This brings us back to the motivating example with which we started this section. We tried to find the matrix that translated the unit square one unit to the right. Our attempt failed, and we have yet to determine why. Given our link between matrices and linear transformations, the answer is likely "the translation transformation is not a linear transformation." While that is a true statement, it doesn't really explain things all that well. Is there some way we could have recognized that this transformation wasn't linear? (That is, apart from applying the definition directly?)

Yes, there is. Consider the second part of the linear transformation definition. It states that $T(k\vec{x}) = kT(\vec{x})$ for all scalars k. If we let $k = 0$, we have $T(0\vec{x}) = 0 \cdot T(\vec{x})$, or more simply, $T(\vec{0}) = \vec{0}$. That is, if T is to be a linear transformation, it must send the zero vector to the zero vector.

This is a quick way to see that the translation transformation fails to be linear. By shifting the unit square to the right one unit, the corner at the point $(0, 0)$ was sent to the point $(1, 0)$, i.e.,

$$\text{the vector } \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ was sent to the vector } \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

This property relating to $\vec{0}$ is important, so we highlight it here.

Key Idea 12 Linear Transformations and $\vec{0}$

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then:

$$T(\vec{0}_n) = \vec{0}_m.$$

That is, the zero vector in $\mathbb{R}^n$ gets sent to the zero vector in $\mathbb{R}^m$.

The Standard Matrix of a Linear Transformation

It is often the case that while one can describe a linear transformation, one doesn't know what matrix performs that transformation (i.e., one doesn't know the standard matrix of that linear transformation). How do we systematically find it? We'll need a new definition.

Definition 34 Standard Unit Vectors

In $\mathbb{R}^n$, the *standard unit vectors* $\vec{e_i}$ are the vectors with a 1 in the i^{th} entry and 0s everywhere else.

The idea that linear transformations "send zero to zero" has an interesting relation to terminology. The reader is likely familiar with functions like $f(x) = 2x + 3$ and would likely refer to this as a "linear function." However, $f(0) \neq 0$, so f is *not* "linear" by our new definition of linear. We erroneously call f "linear" since its graph produces a line, though we should be careful to instead state that "the graph of f is a line."

We've already seen these vectors in the previous section. In $\mathbb{R}^2$, we identified

$$\vec{e_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{e_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

In $\mathbb{R}^4$, there are 4 standard unit vectors:

$$\vec{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \vec{e_4} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

How do these vectors help us find the standard matrix of a linear transformation? Recall again our work in the previous section. There, we practised looking at the transformed unit square and deducing the standard transformation matrix A. We did this by making the first column of A the vector where $\vec{e_1}$ ended up and making the second column of A the vector where $\vec{e_2}$ ended up. One could represent this with:

$$A = \begin{bmatrix} T(\vec{e_1}) & T(\vec{e_2}) \end{bmatrix} = [\, T \,].$$

That is, $T(\vec{e_1})$ is the vector where $\vec{e_1}$ ends up, and $T(\vec{e_2})$ is the vector where $\vec{e_2}$ ends up.

The same holds true in general. Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, the standard matrix of T is the matrix whose i^{th} column is the vector where $\vec{e_i}$ ends up. While we won't prove this is true, it is, and it is very useful. Therefore we'll state it again as a theorem.

Theorem 12 The Standard Matrix of a Linear Transformation

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then $[\, T \,]$ is the $m \times n$ matrix:

$$[\, T \,] = \begin{bmatrix} T(\vec{e_1}) & T(\vec{e_2}) & \cdots & T(\vec{e_n}) \end{bmatrix}.$$

Let's practice this theorem in an example.

Example 77 **Computing the matrix of a linear transformation**
Define $T : \mathbb{R}^3 \to \mathbb{R}^4$ to be the linear transformation where

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ 3x_1 - x_3 \\ 2x_2 + 5x_3 \\ 4x_1 + 3x_2 + 2x_3 \end{bmatrix}.$$

Find $[\,T\,]$.

Solution T takes vectors from $\mathbb{R}^3$ into $\mathbb{R}^4$, so $[\,T\,]$ is going to be a 4×3 matrix. Note that

$$\vec{e_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{e_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

We find the columns of $[\,T\,]$ by finding where $\vec{e_1}$, $\vec{e_2}$ and $\vec{e_3}$ are sent, that is, we find $T(\vec{e_1})$, $T(\vec{e_2})$ and $T(\vec{e_3})$.

$$T(\vec{e_1}) = T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 4 \end{bmatrix} \qquad T(\vec{e_2}) = T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 3 \end{bmatrix} \qquad T(\vec{e_3}) = T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ -1 \\ 5 \\ 2 \end{bmatrix}$$

Thus

$$[\,T\,] = A = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 0 & -1 \\ 0 & 2 & 5 \\ 4 & 3 & 2 \end{bmatrix}.$$

Let's check this. Consider the vector

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Strictly from the original definition, we can compute that

$$T(\vec{x}) = T\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 1+2 \\ 3-3 \\ 4+15 \\ 4+6+6 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 19 \\ 16 \end{bmatrix}.$$

Now compute $T(\vec{x})$ by computing $[\,T\,]\vec{x}= A\vec{x}$.

$$A\vec{x} = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 0 & -1 \\ 0 & 2 & 5 \\ 4 & 3 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 19 \\ 16 \end{bmatrix}.$$

They match! (Of course they do. That was the whole point.)

Let's do another example, one that is more application oriented.

Example 78 **An application to baseball**

A baseball team manager has collected basic data concerning his hitters. He has the number of singles, doubles, triples, and home runs they have hit over the past year. For each player, he wants two more pieces of information: the total number of hits and the total number of bases.

Using the techniques developed in this section, devise a method for the manager to accomplish his goal.

Solution If the manager only wants to compute this for a few players, then he could do it by hand fairly easily. After all:

total # hits = # of singles + # of doubles + # of triples + # of home runs,

and

total # bases = # of singles + 2×# of doubles + 3×# of triples + 4×# of home runs.

However, if he has a lot of players to do this for, he would likely want a way to automate the work. One way of approaching the problem starts with recognizing that he wants to input four numbers into a function (i.e., the number of singles, doubles, etc.) and he wants two numbers as output (i.e., number of hits and bases). Thus he wants a transformation $T : \mathbb{R}^4 \to \mathbb{R}^2$ where each vector in $\mathbb{R}^4$ can be interpreted as

$$\begin{bmatrix} \text{\# of singles} \\ \text{\# of doubles} \\ \text{\# of triples} \\ \text{\# of home runs} \end{bmatrix},$$

and each vector in $\mathbb{R}^2$ can be interpreted as

$$\begin{bmatrix} \text{\# of hits} \\ \text{\# of bases} \end{bmatrix}.$$

To find $[\,T\,]$, he computes $T(\vec{e_1})$, $T(\vec{e_2})$, $T(\vec{e_3})$ and $T(\vec{e_4})$.

$$T(\vec{e_1}) = T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad T(\vec{e_2}) = T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$T(\vec{e_3}) = T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \qquad T(\vec{e_4}) = T\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

(What do these calculations mean? For example, finding $T(\vec{e_3}) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ means that one triple counts as 1 hit and 3 bases.)

Thus our transformation matrix $[T]$ is

$$[T] = A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}.$$

As an example, consider a player who had 102 singles, 30 doubles, 8 triples and 14 home runs. By using A, we find that

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 102 \\ 30 \\ 8 \\ 14 \end{bmatrix} = \begin{bmatrix} 154 \\ 242 \end{bmatrix},$$

meaning the player had 154 hits and 242 total bases.

A question that we should ask concerning the previous example is "How do we know that the function the manager used was actually a linear transformation? After all, we were wrong before – the translation example at the beginning of this section had us fooled at first."

This is a good point; the answer is fairly easy. Recall from Example 74 the transformation

$$T_{74}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1^2 \\ 2x_1 \\ x_1x_2 \end{bmatrix}$$

and from Example 76

$$T_{76}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ 3x_1 - x_3 \\ 2x_2 + 5x_3 \\ 4x_1 + 3x_2 + 2x_3 \end{bmatrix},$$

where we use the subscripts for T to remind us which example they came from.

We found that T_{74} was not a linear transformation, but stated that T_{76} was (although we didn't prove this). What made the difference?

Look at the entries of $T_{74}(\vec{x})$ and $T_{76}(\vec{x})$. T_{74} contains entries where a variable is squared and where 2 variables are multiplied together – these prevent T_{74} from being linear. On the other hand, the entries of T_{76} are all of the form $a_1x_1 + \cdots + a_nx_n$; that is, they are just sums of the variables multiplied by coefficients. T is linear if and only if the entries of $T(\vec{x})$ are of this form. (Hence linear transformations are related to linear equations, as defined in Section 4.1.) This idea is important.

Key Idea 13 Conditions on Linear Transformations

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a transformation and consider the entries of

$$T(\vec{x}) = T\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right).$$

T is linear if and only if each entry of $T(\vec{x})$ is of the form $a_1x_1 + a_2x_2 + \cdots a_nx_n$.

Going back to our baseball example, the manager could have defined his transformation as

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 + x_3 + x_4 \\ x_1 + 2x_2 + 3x_3 + 4x_4 \end{bmatrix}.$$

Since that fits the model shown in Key Idea 13, the transformation T is indeed linear and hence we can find a matrix $[\,T\,]$ that represents it.

Let's practice this concept further in an example.

Example 79 Using Key Idea 13 to identify linear transformations

Using Key Idea 13, determine whether or not each of the following transformations is linear.

$$T_1\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 1 \\ x_2 \end{bmatrix} \qquad T_2\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1/x_2 \\ \sqrt{x_2} \end{bmatrix}$$

$$T_3\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} \sqrt{7}x_1 - x_2 \\ \pi x_2 \end{bmatrix}$$

Solution T_1 is *not* linear! This may come as a surprise, but we are not allowed to add constants to the variables. By thinking about this, we can see that this transformation is trying to accomplish the translation that got us started in this section – it adds 1 to all the x values and leaves the y values alone, shifting everything to the right one unit. However, this is not linear; again, notice how $\vec{0}$ does not get mapped to $\vec{0}$.

T_2 is also not linear. We cannot divide variables, nor can we put variables inside the square root function (among other other things; again, see Section 4.1). This means that the baseball manager would not be able to use matrices to compute a batting average, which is (number of hits)/(number of at bats).

T_3 is linear. Recall that $\sqrt{7}$ and π are just numbers, just coefficients.

We've mentioned before that we can draw vectors other than 2D vectors, although the more dimensions one adds, the harder it gets to understand. In the next section we'll learn about graphing vectors in 3D – that is, how to draw on paper or a computer screen a 3D vector.

Exercises 3.5

Problems

In Exercises 1 – 5, a transformation T is given. Determine whether or not T is linear; if not, state why not.

1. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ 3x_1 - x_2 \end{bmatrix}$

2. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2^2 \\ x_1 - x_2 \end{bmatrix}$

3. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 1 \\ x_2 + 1 \end{bmatrix}$

4. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

5. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

In Exercises 6 – 11, a linear transformation T is given. Find $[\,T\,]$.

6. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ x_1 - x_2 \end{bmatrix}$

7. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ 3x_1 - 5x_2 \\ 2x_2 \end{bmatrix}$

8. $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 - 3x_3 \\ 0 \\ x_1 + 4x_3 \\ 5x_2 + x_3 \end{bmatrix}$

9. $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 3x_3 \\ x_1 - x_3 \\ x_1 + x_3 \end{bmatrix}$

10. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

11. $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 + 3x_3 + 4x_4 \end{bmatrix}$

4: Systems of Linear Equations

You have probably encountered systems of linear equations before; you can probably remember solving systems of equations where you had three equations, three unknowns, and you tried to find the value of the unknowns. In this chapter we will uncover some of the fundamental principles guiding the solution to such problems.

Solving such systems was a bit time consuming, but not terribly difficult. So why bother? We bother because linear equations have many, many, *many* applications, from business to engineering to computer graphics to understanding more mathematics. And not only are there many applications of systems of linear equations, on most occasions where these systems arise we are using far more than three variables. (Engineering applications, for instance, often require thousands of variables.) So getting a good understanding of how to solve these systems effectively is important.

But don't worry; we'll start at the beginning.

4.1 Introduction to Linear Equations

AS YOU READ . . .

1. What is one of the annoying habits of mathematicians?
2. What is the difference between constants and coefficients?
3. Can a coefficient in a linear equation be 0?

We'll begin this section by examining a problem you probably already know how to solve.

Example 80 Counting marbles in a jar

Suppose a jar contains red, blue and green marbles. You are told that there are a total of 30 marbles in the jar; there are twice as many red marbles as green ones; the number of blue marbles is the same as the sum of the red and green marbles. How many marbles of each colour are there?

Solution We could attempt to solve this with some trial and error, and we'd probably get the correct answer without too much work. However, this won't lend itself towards learning a good technique for solving larger problems, so let's be more mathematical about it.

Let's let r represent the number of red marbles, and let b and g denote the number of blue and green marbles, respectively. We can use the given statements about the marbles in the jar to create some equations.

Since we know there are 30 marbles in the jar, we know that

$$r + b + g = 30. \tag{4.1}$$

Also, we are told that there are twice as many red marbles as green ones, so we know that

$$r = 2g. \tag{4.2}$$

Finally, we know that the number of blue marbles is the same as the sum of the red and green marbles, so we have

$$b = r + g. \tag{4.3}$$

From this stage, there isn't one "right" way of proceeding. Rather, there are many ways to use this information to find the solution. One way is to combine ideas from equations 4.2 and 4.3; in 4.3 replace r with $2g$. This gives us

$$b = 2g + g = 3g. \tag{4.4}$$

We can then combine equations 4.1, 4.2 and 4.4 by replacing r in 4.1 with $2g$ as we did before, and replacing b with $3g$ to get

$$\begin{aligned} r + b + g &= 30 \\ 2g + 3g + g &= 30 \\ 6g &= 30 \\ g &= 5 \end{aligned} \tag{4.5}$$

We can now use equation 4.5 to find r and b; we know from 4.2 that $r = 2g = 10$ and then since $r + b + g = 30$, we easily find that $b = 15$.

Mathematicians often see solutions to given problems and then ask "What if. . .?" It's an annoying habit that we would do well to develop – we should learn to think like a mathematician. What are the right kinds of "what if" questions to ask? Here's another annoying habit of mathematicians: they often ask "wrong" questions. That is, they often ask questions and find that the answer isn't particularly interesting. But asking enough questions often leads to some good "right" questions. So don't be afraid of doing something "wrong;" we mathematicians do it all the time.

So what is a good question to ask after seeing Example 79? Here are two possible questions:

1. Did we really have to call the red balls "*r*"? Could we call them "*q*"?

2. What if we had 60 balls at the start instead of 30?

Let's look at the first question. Would the solution to our problem change if we called the red balls q? Of course not. At the end, we'd find that $q = 10$, and we would know that this meant that we had 10 red balls.

Now let's look at the second question. Suppose we had 60 balls, but the other relationships stayed the same. How would the situation and solution change? Let's compare the "original" equations to the "new" equations.

Original	New
$r + b + g = 30$	$r + b + g = 60$
$r = 2g$	$r = 2g$
$b = r + g$	$b = r + g$

By examining these equations, we see that nothing has changed except the first equation. It isn't too much of a stretch of the imagination to see that we would solve this new problem exactly the same way that we solved the original one, except that we'd have twice as many of each type of ball.

A conclusion from answering these two questions is this: it doesn't matter what we call our variables, and while changing constants in the equations changes the solution, they don't really change the *method* of how we solve these equations.

In fact, it is a great discovery to realize that all we care about are the *constants* and the *coefficients* of the equations. By systematically handling these, we can solve any set of linear equations in a very nice way. Before we go on, we must first define what a linear equation is.

Definition 35 Linear Equation

A *linear equation* is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = c$$

where the x_i are variables (the unknowns), the a_i are coefficients, and c is a constant.

A *system of linear equations* is a set of linear equations that involve the same variables.

A *solution* to a system of linear equations is a set of values for the variables x_i such that each equation in the system is satisfied.

So in Example 79, when we answered "how many marbles of each colour are there?," we were also answering "find a solution to a certain system of linear equations."

The following are examples of linear equations:

$$\begin{aligned} 2x + 3y - 7z &= 29 \\ x_1 + \frac{7}{2}x_2 + x_3 - x_4 + 17x_5 &= \sqrt[3]{-10} \\ y_1 + 14^2y_4 + 4 &= y_2 + 13 - y_1 \\ \sqrt{7}r + \pi s + \frac{3t}{5} &= \cos(45^\circ) \end{aligned}$$

Notice that the coefficients and constants can be fractions and irrational numbers (like π, $\sqrt[3]{-10}$ and $\cos(45^\circ)$). The variables only come in the form of a_ix_i; that is, just one variable multiplied by a coefficient. (Note that $\frac{3t}{5} = \frac{3}{5}t$, just a variable multiplied by a coefficient.) Also, it doesn't really matter what side of the equation we put the variables and the constants, although most of the time we write them with the variables on the left and the constants on the right.

We would not regard the above collection of equations to constitute a system of equations, since each equation uses differently named variables. An example of a system of linear equations is

$$\begin{aligned} x_1 - x_2 + x_3 + x_4 &= 1 \\ 2x_1 + 3x_2 + x_4 &= 25 \\ x_2 + x_3 &= 10 \end{aligned}$$

It is important to notice that not all equations used all of the variables (it is more accurate to say that the coefficients can be 0, so the last equation could have been written as $0x_1 + x_2 + x_3 + 0x_4 = 10$). Also, just because we have four unknowns does not mean we have to have four equations. We could have had fewer, even just one, and we could have had more.

To get a better feel for what a linear equation is, we point out some examples of what are *not* linear equations.

$$
\begin{aligned}
2xy + z &= 1\\
5x^2 + 2y^5 &= 100\\
\frac{1}{x} + \sqrt{y} + 24z &= 3\\
\sin^2 x_1 + \cos^2 x_2 &= 29\\
2^{x_1} + \ln x_2 &= 13
\end{aligned}
$$

The first example is not a linear equation since the variables x and y are multiplied together. The second is not a linear equation because the variables are raised to powers other than 1; that is also a problem in the third equation (remember that $1/x = x^{-1}$ and $\sqrt{x} = x^{1/2}$). Our variables cannot be the argument of function like sin, cos or ln, nor can our variables be raised as an exponent.

At this stage, we have yet to discuss how to efficiently find a solution to a system of linear equations. That is a goal for the upcoming sections. Right now we focus on identifying linear equations. It is also useful to “limber” up by solving a few systems of equations using any method we have at hand to refresh our memory about the basic process.

Exercises 4.1

Problems

In Exercises 1 – 10, state whether or not the given equation is linear.

1. $x + y + z = 10$

2. $xy + yz + xz = 1$

3. $-3x + 9 = 3y - 5z + x - 7$

4. $\sqrt{5}y + \pi x = -1$

5. $(x-1)(x+1) = 0$

6. $\sqrt{x_1^2 + x_2^2} = 25$

7. $x_1 + y + t = 1$

8. $\frac{1}{x} + 9 = 3\cos(y) - 5z$

9. $\cos(15)y + \frac{x}{4} = -1$

10. $2^x + 2^y = 16$

In Exercises 11 – 14, solve the system of linear equations.

11. $$\begin{array}{rcrcr} x & + & y & = & -1 \\ 2x & - & 3y & = & 8 \end{array}$$

12. $$\begin{array}{rcrcr} 2x & - & 3y & = & 3 \\ 3x & + & 6y & = & 8 \end{array}$$

13. $$\begin{array}{rcrcrcr} x & - & y & + & z & = & 1 \\ 2x & + & 6y & - & z & = & -4 \\ 4x & - & 5y & + & 2z & = & 0 \end{array}$$

14. $$\begin{array}{rcrcrcr} x & + & y & - & z & = & 1 \\ 2x & + & y & & & = & 2 \\ & & y & + & 2z & = & 0 \end{array}$$

15. A farmer looks out his window at his chickens and pigs. He tells his daughter that he sees 62 heads and 190 legs. How many chickens and pigs does the farmer have?

16. A lady buys 20 trinkets at a yard sale. The cost of each trinket is either \$0.30 or \$0.65. If she spends \$8.80, how many of each type of trinket does she buy?

4.2 Using Matrices To Solve Systems of Linear Equations

AS YOU READ . . .

1. What is remarkable about the definition of a matrix?
2. Vertical lines of numbers in a matrix are called what?
3. In a matrix A, the entry a_{53} refers to which entry?
4. What is an augmented matrix?

In Section 4.1 we solved a linear system using familiar techniques. Later, we commented that in the linear equations we formed, the most important information was the coefficients and the constants; the names of the variables really didn't matter. In Example 79 we had the following three equations:

$$\begin{aligned} r+b+g &= 30 \\ r &= 2g \\ b &= r+g \end{aligned}$$

Let's rewrite these equations so that all variables are on the left of the equal sign and all constants are on the right. Also, for a bit more consistency, let's list the variables in alphabetical order in each equation. Therefore we can write the equations as

$$\begin{array}{ccccccc} b & + & g & + & r & = & 30 \\ & - & 2g & + & r & = & 0 \\ -b & + & g & + & r & = & 0 \end{array}. \tag{4.6}$$

As we mentioned before, there isn't just one "right" way of finding the solution to this system of equations. Here is another way to do it, a way that is a bit different from our method in Section 4.1.

First, let's add the first and last equations together, and write the result as a new third equation. This gives us:

$$\begin{array}{ccccccc} b & + & g & + & r & = & 30 \\ & - & 2g & + & r & = & 0 \\ & & 2g & + & 2r & = & 30 \end{array}.$$

A nice feature of this is that the only equation with a b in it is the first equation.

Now let's multiply the second equation by $-\frac{1}{2}$. This gives

$$\begin{array}{ccccccc} b & + & g & + & r & = & 30 \\ & & g & - & 1/2r & = & 0 \\ & & 2g & + & 2r & = & 30 \end{array}.$$

Let's now do two steps in a row; our goal is to get rid of the g's in the first and third equations. In order to remove the g in the first equation, let's multiply the second equation by -1 and add that to the first equation, replacing the first equation with that sum. To remove the g in the third equation, let's multiply the second equation by -2 and add that to the third equation, replacing the third equation. Our new system of equations now becomes

$$\begin{array}{ccccccc} b & + & & & 3/2r & = & 30 \\ & & g & - & 1/2r & = & 0 \\ & & & & 3r & = & 30 \end{array}.$$

Clearly we can multiply the third equation by $\frac{1}{3}$ and find that $r = 10$; let's make this our new third equation, giving

$$\begin{array}{ccccccc} b & + & & & 3/2r & = & 30 \\ & & g & - & 1/2r & = & 0 \\ & & & & r & = & 10 \end{array}.$$

Now let's get rid of the r's in the first and second equation. To remove the r in the first equation, let's multiply the third equation by $-\frac{3}{2}$ and add the result to the first equation, replacing the first equation with that sum. To remove the r in the second equation, we can multiply the third equation by $\frac{1}{2}$ and add that to the second equation, replacing the second equation with that sum. This gives us:

$$\begin{array}{ccccc} b & & & = & 15 \\ & g & & = & 5 \\ & & r & = & 10 \end{array}.$$

Clearly we have discovered the same result as when we solved this problem in Section 4.1.

Now again revisit the idea that all that really matters are the coefficients and the constants. There is nothing special about the letters b, g and r; we could have used x, y and z or x_1, x_2 and x_3. And even then, since we wrote our equations so carefully, we really didn't need to write the variable names at all as long as we put things "in the right place."

Let's look again at our system of equations in (4.6) and write the coefficients and the constants in a rectangular array. This time we won't ignore the zeros, but rather write them out.

$$\begin{array}{ccccccc} b & + & g & + & r & = & 30 \\ & - & 2g & + & r & = & 0 \\ -b & + & g & + & r & = & 0 \end{array} \qquad \Leftrightarrow \qquad \begin{bmatrix} 1 & 1 & 1 & 30 \\ 0 & -2 & 1 & 0 \\ -1 & 1 & 1 & 0 \end{bmatrix}$$

Notice how even the equal signs are gone; we don't need them, for we know that the last *column* contains the coefficients.

We have just created a *matrix*. The definition of matrix is remarkable only in how unremarkable it seems.

Definition 36 Matrix

A *matrix* is a rectangular array of numbers.

The horizontal lines of numbers form *rows* and the vertical lines of numbers form *columns*. A matrix with m rows and n columns is said to be an $m \times n$ matrix ("an m by n matrix").

The entries of an $m \times n$ matrix are indexed as follows:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}.$$

That is, a_{32} means "the number in the third row and second column."

In the future, we'll want to create matrices with just the coefficients of a system of linear equations and leave out the constants. Therefore, when we include the constants, we often refer to the resulting matrix as an *augmented matrix*.

It is common (but not mandatory) to place a vertical line separating the final column of an augmented matrix (containing the constants) from the other columns (containing the coefficients). One advantage of doing so is that we can quickly recognize that we're dealing with an augmented matrix rather than a matrix of coefficients. For example, the augmented matrix for the system (4.6) would be written as seen below on the right:

$$\begin{bmatrix} 1 & 1 & 1 & 30 \\ 0 & -2 & 1 & 0 \\ -1 & 1 & 1 & 0 \end{bmatrix} \qquad \left[\begin{array}{ccc|c} 1 & 1 & 1 & 30 \\ 0 & -2 & 1 & 0 \\ -1 & 1 & 1 & 0 \end{array}\right]$$

Without the vertical line　　With the vertical line

Two ways of writing an augmented matrix

We can use augmented matrices to find solutions to linear equations by using essentially the same steps we used above. Every time we used the word "equation" above, substitute the word "row," as we show below. The comments explain how we get from the current set of equations (or matrix) to the one on the next line.

We can use a shorthand to describe matrix operations; let R_1, R_2 represent "row 1" and "row 2," respectively. We can write "add row 1 to row 3, and replace row 3 with that sum" as "$R_1 + R_3 \to R_3$." The expression "$R_1 \leftrightarrow R_2$" means "interchange row 1 and row 2."

$$\begin{array}{rcrcrcr} b & + & g & + & r & = & 30 \\ & - & 2g & + & r & = & 0 \\ -b & + & g & + & r & = & 0 \end{array} \qquad \begin{bmatrix} 1 & 1 & 1 & 30 \\ 0 & -2 & 1 & 0 \\ -1 & 1 & 1 & 0 \end{bmatrix}$$

Replace equation 3 with the sum of equations 1 and 3

Replace row 3 with the sum of rows 1 and 3. $(R_1 + R_3 \to R_3)$

$$\begin{array}{rcrcrcr} b & + & g & + & r & = & 30 \\ & - & 2g & + & r & = & 0 \\ & & 2g & + & 2r & = & 30 \end{array} \qquad \begin{bmatrix} 1 & 1 & 1 & 30 \\ 0 & -2 & 1 & 0 \\ 0 & 2 & 2 & 30 \end{bmatrix}$$

Multiply equation 2 by $-\frac{1}{2}$

Multiply row 2 by $-\frac{1}{2}$ $(-\frac{1}{2}R_2 \to R_2)$

$$\begin{array}{rcrcrcr} b & + & g & + & r & = & 30 \\ & & g & + & -1/2r & = & 0 \\ & & 2g & + & 2r & = & 30 \end{array} \qquad \begin{bmatrix} 1 & 1 & 1 & 30 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 2 & 2 & 30 \end{bmatrix}$$

Replace equation 1 with the sum of (-1) times equation 2 plus equation 1;
Replace equation 3 with the sum of (-2) times equation 2 plus equation 3

Replace row 1 with the sum of (-1) times row 2 plus row 1 $(-R_2 + R_1 \to R_1)$;
Replace row 3 with the sum of (-2) times row 2 plus row 3 $(-2R_2 + R_3 \to R_3)$

$$\begin{array}{ccccccc} b & + & & & 3/2r & = & 30 \\ & & g & - & 1/2r & = & 0 \\ & & & & 3r & = & 30 \end{array} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & 30 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 3 & 30 \end{bmatrix}$$

Multiply equation 3 by $\frac{1}{3}$

Multiply row 3 by $\frac{1}{3}$
($\frac{1}{3}R_3 \rightarrow R_3$)

$$\begin{array}{ccccccc} b & + & & & 3/2r & = & 30 \\ & & g & - & 1/2r & = & 0 \\ & & & & r & = & 10 \end{array} \qquad \begin{bmatrix} 1 & 0 & \frac{3}{2} & 30 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & 10 \end{bmatrix}$$

Replace equation 2 with the sum of $\frac{1}{2}$ times equation 3 plus equation 2;
Replace equation 1 with the sum of $-\frac{3}{2}$ times equation 3 plus equation 1

Replace row 2 with the sum of $\frac{1}{2}$ times row 3 plus row 2
($\frac{1}{2}R_3 + R_2 \rightarrow R_2$);
Replace row 1 with the sum of $-\frac{3}{2}$ times row 3 plus row 1
($-\frac{3}{2}R_3 + R_1 \rightarrow R_1$)

$$\begin{array}{cccc} b & & & = 15 \\ & g & & = 5 \\ & & r & = 10 \end{array} \qquad \begin{bmatrix} 1 & 0 & 0 & 15 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 10 \end{bmatrix}$$

The final matrix contains the same solution information as we have on the left in the form of equations. Recall that the first column of our matrices held the coefficients of the b variable; the second and third columns held the coefficients of the g and r variables, respectively. Therefore, the first row of the matrix can be interpreted as "$b + 0g + 0r = 15$," or more concisely, "$b = 15$."

Let's practice this manipulation again.

Example 81 Solving a system using augmented matrices

Find a solution to the following system of linear equations by simultaneously manipulating the equations and the corresponding augmented matrices.

$$\begin{array}{ccccccc} x_1 & + & x_2 & + & x_3 & = & 0 \\ 2x_1 & + & 2x_2 & + & x_3 & = & 0 \\ -1x_1 & + & x_2 & - & 2x_3 & = & 2 \end{array}$$

Solution We'll first convert this system of equations into a matrix, then we'll proceed by manipulating the system of equations (and hence the matrix) to find a solution. Again, there is not just one "right" way of proceeding; we'll choose a method that is pretty efficient, but other methods certainly exist (and may be "better"!). The method use here, though, is a good one, and it is the method that we will be learning in the future.

The given system and its corresponding augmented matrix are seen below.

Original system of equations

Corresponding matrix

$$\begin{array}{ccccccc} x_1 & + & x_2 & + & x_3 & = & 0 \\ 2x_1 & + & 2x_2 & + & x_3 & = & 0 \\ -1x_1 & + & x_2 & - & 2x_3 & = & 2 \end{array} \qquad \begin{bmatrix} 1 & 1 & 1 & 0 \\ 2 & 2 & 1 & 0 \\ -1 & 1 & -2 & 2 \end{bmatrix}$$

We'll proceed by trying to get the x_1 out of the second and third equation.

Replace equation 2 with the sum of (-2) times equation 1 plus equation 2;
Replace equation 3 with the sum of equation 1 and equation 3

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 0 \\ & & & & -x_3 & = & 0 \\ & & 2x_2 & - & x_3 & = & 2 \end{array}$$

Replace row 2 with the sum of (-2) times row 1 plus row 2 $(-2R_1 + R_2 \rightarrow R_2)$;
Replace row 3 with the sum of row 1 and row 3 $(R_1 + R_3 \rightarrow R_3)$

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

Notice that the second equation no longer contains x_2. We'll exchange the order of the equations so that we can follow the convention of solving for the second variable in the second equation.

Interchange equations 2 and 3

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 0 \\ & & 2x_2 & - & x_3 & = & 2 \\ & & & & -x_3 & = & 0 \end{array}$$

Interchange rows 2 and 3 $R_2 \leftrightarrow R_3$

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

Multiply equation 2 by $\frac{1}{2}$

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 0 \\ & & x_2 & - & \frac{1}{2}x_3 & = & 1 \\ & & & & -x_3 & = & 0 \end{array}$$

Multiply row 2 by $\frac{1}{2}$ $(\frac{1}{2}R_2 \rightarrow R_2)$

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

Multiply equation 3 by -1

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 0 \\ & & x_2 & - & \frac{1}{2}x_3 & = & 1 \\ & & & & x_3 & = & 0 \end{array}$$

Multiply row 3 by -1 $(-1R_3 \rightarrow R_3)$

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

Notice that the last equation (and also the last row of the matrix) show that $x_3 = 0$. Knowing this would allow us to simply eliminate the x_3 from the first two equations. However, we will formally do this by manipulating the equations (and rows) as we have previously.

Replace equation 1 with the sum of (-1) times equation 3 plus equation 1;
Replace equation 2 with the sum of $\frac{1}{2}$ times equation 3 plus equation 2

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & & & = & 0 \\ & & x_2 & & & = & 1 \\ & & & & x_3 & = & 0 \end{array}$$

Replace row 1 with the sum of (-1) times row 3 plus row 1 $(-R_3 + R_1 \rightarrow R_1)$;
Replace row 2 with the sum of $\frac{1}{2}$ times row 3 plus row 2 $(\frac{1}{2}R_3 + R_2 \rightarrow R_2)$

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

Notice how the second equation shows that $x_2 = 1$. All that remains to do is to solve for x_1.

Replace equation 1 with the sum of (-1) times equation 2 plus equation 1	Replace row 1 with the sum of (-1) times row 2 plus row 1 $(-R_2 + R_1 \to R_1)$
$\begin{array}{rcrcrcr} x_1 & & & & & = & -1 \\ & & x_2 & & & = & 1 \\ & & & & x_3 & = & 0 \end{array}$	$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

Obviously the equations on the left tell us that $x_1 = -1$, $x_2 = 1$ and $x_3 = 0$, and notice how the matrix on the right tells us the same information.

Exercises 4.2

Problems

In Exercises 1 – 4, convert the given system of linear equations into an augmented matrix.

1. $\begin{array}{rcrcrcr} 3x & + & 4y & + & 5z & = & 7 \\ -x & + & y & - & 3z & = & 1 \\ 2x & - & 2y & + & 3z & = & 5 \end{array}$

2. $\begin{array}{rcrcrcr} 2x & + & 5y & - & 6z & = & 2 \\ 9x & & & - & 8z & = & 10 \\ -2x & + & 4y & + & z & = & -7 \end{array}$

3. $\begin{array}{rl} x_1 + 3x_2 - 4x_3 + 5x_4 = & 17 \\ -x_1 + 4x_3 + 8x_4 = & 1 \\ 2x_1 + 3x_2 + 4x_3 + 5x_4 = & 6 \end{array}$

4. $\begin{array}{rl} 3x_1 - 2x_2 = & 4 \\ 2x_1 = & 3 \\ -x_1 + 9x_2 = & 8 \\ 5x_1 - 7x_2 = & 13 \end{array}$

In Exercises 5 – 9, convert the given augmented matrix into a system of linear equations. Use the variables x_1, x_2, etc.

5. $\begin{bmatrix} 1 & 2 & 3 \\ -1 & 3 & 9 \end{bmatrix}$

6. $\begin{bmatrix} -3 & 4 & 7 \\ 0 & 1 & -2 \end{bmatrix}$

7. $\begin{bmatrix} 1 & 1 & -1 & -1 & 2 \\ 2 & 1 & 3 & 5 & 7 \end{bmatrix}$

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

9. $\begin{bmatrix} 1 & 0 & 1 & 0 & 7 & 2 \\ 0 & 1 & 3 & 2 & 0 & 5 \end{bmatrix}$

In Exercises 10 – 15, perform the given row operations on A, where

$$A = \begin{bmatrix} 2 & -1 & 7 \\ 0 & 4 & -2 \\ 5 & 0 & 3 \end{bmatrix}.$$

10. $-1R_1 \to R_1$

11. $R_2 \leftrightarrow R_3$

12. $R_1 + R_2 \to R_2$

13. $2R_2 + R_3 \to R_3$

14. $\frac{1}{2}R_2 \to R_2$

15. $-\frac{5}{2}R_1 + R_3 \to R_3$

A matrix A is given below. In Exercises 16 – 20, a matrix B is given. Give the row operation that transforms A into B.

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

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

17. $B = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$

18. $B = \begin{bmatrix} 3 & 5 & 7 \\ 1 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}$

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

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

In Exercises 21 – 26, rewrite the system of equations in matrix form. Find the solution to the linear system by simultaneously manipulating the equations and the matrix.

21. $\begin{array}{rcrcr} x & + & y & = & 3 \\ 2x & - & 3y & = & 1 \end{array}$

22. $\begin{array}{rcrcr} 2x & + & 4y & = & 10 \\ -x & + & y & = & 4 \end{array}$

23. $\begin{array}{rcrcr} -2x & + & 3y & = & 2 \\ -x & + & y & = & 1 \end{array}$

24. $\begin{array}{rcrcr} 2x & + & 3y & = & 2 \\ -2x & + & 6y & = & 1 \end{array}$

25. $\begin{array}{rcrcrcr} -5x_1 & & & + & 2x_3 & = & 14 \\ & & x_2 & & & = & 1 \\ -3x_1 & & & + & x_3 & = & 8 \end{array}$

26. $\begin{array}{rcrcrcr} & - & 5x_2 & + & 2x_3 & = & -11 \\ x_1 & & & + & 2x_3 & = & 15 \\ & - & 3x_2 & + & x_3 & = & -8 \end{array}$

4.3 Elementary Row Operations and Gaussian Elimination

AS YOU READ . . .

1. Give two reasons why the Elementary Row Operations are called "Elementary."
2. T/F: Assuming a solution exists, all linear systems of equations can be solved using only elementary row operations.
3. Give one reason why one might not be interested in putting a matrix into reduced row echelon form.
4. Identify the leading 1s in the following matrix:

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

5. Using the "forward" and "backward" steps of Gaussian elimination creates lots of ________ making computations easier.

In our examples thus far, we have essentially used just three types of manipulations in order to find solutions to our systems of equations. These three manipulations are:

1. Add a scalar multiple of one equation to a second equation, and replace the second equation with that sum
2. Multiply one equation by a nonzero scalar
3. Swap the position of two equations in our list

We saw earlier how we could write all the information of a system of equations in a matrix, so it makes sense that we can perform similar operations on matrices (as we have done before). Again, simply replace the word "equation" above with the word "row."

We didn't justify our ability to manipulate our equations in the above three ways; it seems rather obvious that we should be able to do that. In that sense, these operations are "elementary." These operations are *elementary* in another sense; they are *fundamental* – they form the basis for much of what we will do in matrix algebra. Since these operations are so important, we list them again here in the context of matrices.

Key Idea 14 Elementary Row Operations

1. Add a scalar multiple of one row to another row, and replace the latter row with that sum
2. Multiply one row by a nonzero scalar
3. Swap the position of two rows

Given any system of linear equations, we can find a solution (if one exists) by using these three row operations. Elementary row operations give us a new linear system, but the solution to the new system is the same as the old. We can use these operations as much as we want and not change the solution. This brings to mind two good questions:

1. Since we can use these operations as much as we want, how do we know when to stop? (Where are we supposed to "go" with these operations?)

2. Is there an efficient way of using these operations? (How do we get "there" the fastest?)

We'll answer the first question first. Most of the time (unless one prefers obfuscation to clarification) we will want to take our original matrix and, using the elementary row operations, put it into something called **reduced row echelon form**. This is our "destination," for this form allows us to readily identify whether or not a solution exists, and in the case that it does, what that solution is.

In the previous section, when we manipulated matrices to find solutions, we were unwittingly putting the matrix into reduced row echelon form. However, not all solutions come in such a simple manner as we've seen so far. Putting a matrix into reduced row echelon form helps us identify all types of solutions. We'll explore the topic of understanding what the reduced row echelon form of a matrix tells us in the following sections; in this section we focus on finding it.

Definition 37 Reduced Row Echelon Form

A matrix is in *reduced row echelon form* if its entries satisfy the following conditions.

1. The first nonzero entry in each row is a 1 (called a *leading 1*).
2. Each leading 1 comes in a column to the right of the leading 1s in rows above it.
3. All rows of all 0s come at the bottom of the matrix.
4. If a column contains a leading 1, then all other entries in that column are 0.

A matrix that satisfies the first three conditions is said to be in *row echelon form*.

Example 82 Determining if a matrix is in reduced row echelon form
Which of the following matrices is in reduced row echelon form?

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

c) $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ d) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 \end{bmatrix}$ f) $\begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$

g) $\begin{bmatrix} 0 & 1 & 2 & 3 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ h) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Solution The matrices in a), b), c), d) and g) are all in reduced row echelon form. Check to see that each satisfies the necessary conditions. If your instincts were wrong on some of these, correct your thinking accordingly.

The matrix in e) is not in reduced row echelon form since the row of all zeros is not at the bottom. The matrix in f) is not in reduced row echelon form since the first nonzero entries in rows 2 and 3 are not 1. Finally, the matrix in h) is not in reduced row echelon form since the first entry in column 2 is not zero; the second 1 in column 2 is a leading one, hence all other entries in that column should be 0.

We end this example with a preview of what we'll learn in the future. Consider the matrix in b). If this matrix came from the augmented matrix of a system of linear equations, then we can readily recognize that the solution of the system is $x_1 = 1$ and $x_2 = 2$. Again, in previous examples, when we found the solution to a linear system, we were unwittingly putting our matrices into reduced row echelon form.

We began this section discussing how we can manipulate the entries in a matrix with elementary row operations. This led to two questions, "Where do we go?" and "How do we get there quickly?" We've just answered the first question: most of the time we are "going to" reduced row echelon form. We now address the second question.

There is no one "right" way of using these operations to transform a matrix into reduced row echelon form. However, there is a general technique that works very well in that it is very efficient (so we don't waste time on unnecessary steps). This technique is called *Gaussian elimination*. It is named in honour of the great mathematician Karl Friedrich Gauss.

While this technique isn't very difficult to use, it is one of those things that is easier understood by watching it being used than explained as a series of steps. With this in mind, we will go through one more example highlighting important steps and then we'll explain the procedure in detail.

Example 83 **Using row operations to simplify an augmented matrix**
Put the augmented matrix of the following system of linear equations into reduced row echelon form.

$$\begin{array}{rcrcrcr} -3x_1 & - & 3x_2 & + & 9x_3 & = & 12 \\ 2x_1 & + & 2x_2 & - & 4x_3 & = & -2 \\ & & -2x_2 & - & 4x_3 & = & -8 \end{array}$$

Solution We start by converting the linear system into an augmented matrix.

$$\left[\begin{array}{ccc|c} \boxed{-3} & -3 & 9 & 12 \\ 2 & 2 & -4 & -2 \\ 0 & -2 & -4 & -8 \end{array}\right]$$

Our next step is to change the entry in the box to a 1. To do this, let's multiply row 1 by $-\frac{1}{3}$.

$$-\tfrac{1}{3}R_1 \rightarrow R_1 \qquad \left[\begin{array}{ccc|c} 1 & 1 & -3 & -4 \\ 2 & 2 & -4 & -2 \\ 0 & -2 & -4 & -8 \end{array}\right]$$

We have now created a *leading 1*; that is, the first entry in the first row is a 1. Our next step is to put zeros under this 1. To do this, we'll use the elementary row operation given below.

$$-2R_1 + R_2 \rightarrow R_2 \qquad \left[\begin{array}{ccc|c} 1 & 1 & -3 & -4 \\ 0 & \boxed{0} & 2 & 6 \\ 0 & -2 & -4 & -8 \end{array}\right]$$

Once this is accomplished, we shift our focus from the leading one down one row, and to the right one column, to the position that is boxed. We again want to put a 1 in this position. We can use any elementary row operations, but we need to restrict ourselves to using only the second row and any rows below it. Probably the simplest thing we can do is interchange rows 2 and 3, and then scale the new second row so that there is a 1 in the desired position.

$$R_2 \leftrightarrow R_3 \qquad \left[\begin{array}{ccc|c} 1 & 1 & -3 & -4 \\ 0 & \boxed{-2} & -4 & -8 \\ 0 & 0 & 2 & 6 \end{array}\right]$$

$$-\tfrac{1}{2}R_2 \rightarrow R_2 \qquad \left[\begin{array}{ccc|c} 1 & 1 & -3 & -4 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & \boxed{2} & 6 \end{array}\right]$$

We have now created another leading 1, this time in the second row. Our next desire is to put zeros underneath it, but this has already been accomplished by our previous steps. Therefore we again shift our attention to the right one column and down one row, to the next position put in the box. We want that to be a 1. A simple scaling will accomplish this.

$$\tfrac{1}{2}R_3 \rightarrow R_3 \qquad \left[\begin{array}{ccc|c} 1 & 1 & -3 & -4 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

This ends what we will refer to as the *forward steps*. Our next task is to use the elementary row operations and go back and put zeros above our leading 1s. This is referred to as the *backward steps*. These steps are given below.

$$\begin{array}{c} 3R_3 + R_1 \rightarrow R_1 \\ -2R_3 + R_2 \rightarrow R_2 \end{array} \qquad \left[\begin{array}{ccc|c} 1 & 1 & 0 & 5 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

$$-R_2 + R_1 \rightarrow R_1 \qquad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

It is now easy to read off the solution as $x_1 = 7$, $x_2 = -2$ and $x_3 = 3$.

We now formally explain the procedure used to find the solution above. As you read through the procedure, follow along with the example above so that the explanation makes more sense.

Forward Steps

1. Working from left to right, consider the first column that isn't all zeros that hasn't already been worked on. Then working from top to bottom, consider the first row that hasn't been worked on.

2. If the entry in the row and column that we are considering is zero, interchange rows with a row below the current row so that that entry is nonzero. If all entries below are zero, we are done with this column; start again at step 1.

3. Multiply the current row by a scalar to make its first entry a 1 (a leading 1).

4. Repeatedly use Elementary Row Operation 1 to put zeros underneath the leading one.

5. Go back to step 1 and work on the new rows and columns until either all rows or columns have been worked on.

If the above steps have been followed properly, then the following should be true about the current state of the matrix:

1. The first nonzero entry in each row is a 1 (a leading 1).

2. Each leading 1 is in a column to the right of the leading 1s above it.

3. All rows of all zeros come at the bottom of the matrix.

Note that this means we have just put a matrix into row echelon form. The next steps finish the conversion into *reduced* row echelon form. These next steps are referred to as the *backward* steps. These are much easier to state.

Backward Steps

1. Starting from the right and working left, use Elementary Row Operation 1 repeatedly to put zeros above each leading 1.

The basic method of Gaussian elimination is this: create leading ones and then use elementary row operations to put zeros above and below these leading ones. We can do this in any order we please, but by following the "Forward Steps" and "Backward Steps," we make use of the presence of zeros to make the overall computations easier. This method is very efficient, so it gets its own name (which we've already been using).

Definition 38 Gaussian Elimination

Gaussian elimination is the technique for finding the reduced row echelon form of a matrix using the above procedure. It can be abbreviated to:

1. Create a leading 1.
2. Use this leading 1 to put zeros underneath it.
3. Repeat the above steps until all possible rows have leading 1s.
4. Put zeros above these leading 1s.

Let's practice some more.

Example 84 Using Gaussian elimination

Use Gaussian elimination to put the matrix A into reduced row echelon form, where

$$A = \begin{bmatrix} -2 & -4 & -2 & -10 & 0 \\ 2 & 4 & 1 & 9 & -2 \\ 3 & 6 & 1 & 13 & -4 \end{bmatrix}.$$

Solution We start by wanting to make the entry in the first column and first row a 1 (a leading 1). To do this we'll scale the first row by a factor of $-\frac{1}{2}$.

$$-\tfrac{1}{2}R_1 \rightarrow R_1 \qquad \begin{bmatrix} 1 & 2 & 1 & 5 & 0 \\ 2 & 4 & 1 & 9 & -2 \\ 3 & 6 & 1 & 13 & -4 \end{bmatrix}$$

Next we need to put zeros in the column below this newly formed leading 1.

$$\begin{array}{l} -2R_1 + R_2 \rightarrow R_2 \\ -3R_1 + R_3 \rightarrow R_3 \end{array} \qquad \begin{bmatrix} 1 & 2 & 1 & 5 & 0 \\ 0 & \boxed{0} & -1 & -1 & -2 \\ 0 & 0 & -2 & -2 & -4 \end{bmatrix}$$

Our attention now shifts to the right one column and down one row to the position indicated by the box. We want to put a 1 in that position. Our only options are to either scale the current row or to interchange rows with a row below it. However, in this case neither of these options will accomplish our goal. Therefore, we shift our attention to the right one more column.

We want to put a 1 where there is a –1. A simple scaling will accomplish this; once done, we will put a 0 underneath this leading one.

$$-R_2 \rightarrow R_2 \qquad \begin{bmatrix} 1 & 2 & 1 & 5 & 0 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & -2 & -2 & -4 \end{bmatrix}$$

$$2R_2 + R_3 \rightarrow R_3 \qquad \begin{bmatrix} 1 & 2 & 1 & 5 & 0 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & \boxed{0} & 0 \end{bmatrix}$$

Our attention now shifts over one more column and down one row to the position indicated by the box; we wish to make this a 1. Of course, there is no way to do this, so we are done with the forward steps.

Our next goal is to put a 0 above each of the leading 1s (in this case there is only one leading 1 to deal with).

$$-R_2 + R_1 \rightarrow R_1 \qquad \begin{bmatrix} 1 & 2 & 0 & 4 & -2 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This final matrix is in reduced row echelon form.

Example 85 **Gaussian elimination, again**

Put the matrix

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

into reduced row echelon form.

Solution Here we will show all steps without explaining each one.

$$\begin{array}{l} -2R_1 + R_2 \rightarrow R_2 \\ -3R_1 + R_3 \rightarrow R_3 \end{array} \qquad \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & -3 & -1 & -5 \\ 0 & -3 & -1 & -8 \end{bmatrix}$$

$$-\tfrac{1}{3}R_2 \rightarrow R_2 \qquad \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & 1/3 & 5/3 \\ 0 & -3 & -1 & -8 \end{bmatrix}$$

$$3R_2 + R_3 \rightarrow R_3 \qquad \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & 1/3 & 5/3 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

$$-\tfrac{1}{3}R_3 \rightarrow R_3 \qquad \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & 1/3 & 5/3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{array}{l} -3R_3 + R_1 \rightarrow R_1 \\ -\tfrac{5}{3}R_3 + R_2 \rightarrow R_2 \end{array} \qquad \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 1/3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$-2R_2 + R_1 \rightarrow R_1 \qquad \begin{bmatrix} 1 & 0 & 1/3 & 0 \\ 0 & 1 & 1/3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The last matrix in the above example is in reduced row echelon form. If one thinks of the original matrix as representing the augmented matrix of a system of linear equations, this final result is interesting. What does it mean to have a leading one in the last column? We'll figure this out in the next section.

Example 86 **Using back substitution**

Put the matrix A into reduced row echelon form, where

$$A = \begin{bmatrix} 2 & 1 & -1 & 4 \\ 1 & -1 & 2 & 12 \\ 2 & 2 & -1 & 9 \end{bmatrix}.$$

Solution We'll again show the steps without explanation, although we will stop at the end of the forward steps and make a comment.

$$\frac{1}{2}R_1 \rightarrow R_1 \qquad \begin{bmatrix} 1 & 1/2 & -1/2 & 2 \\ 1 & -1 & 2 & 12 \\ 2 & 2 & -1 & 9 \end{bmatrix}$$

$$\begin{array}{c} -R_1 + R_2 \rightarrow R_2 \\ -2R_1 + R_3 \rightarrow R_3 \end{array} \qquad \begin{bmatrix} 1 & 1/2 & -1/2 & 2 \\ 0 & -3/2 & 5/2 & 10 \\ 0 & 1 & 0 & 5 \end{bmatrix}$$

$$-\frac{2}{3}R_2 \rightarrow R_2 \qquad \begin{bmatrix} 1 & 1/2 & -1/2 & 2 \\ 0 & 1 & -5/3 & -20/3 \\ 0 & 1 & 0 & 5 \end{bmatrix}$$

$$-R_2 + R_3 \rightarrow R_3 \qquad \begin{bmatrix} 1 & 1/2 & -1/2 & 2 \\ 0 & 1 & -5/3 & -20/3 \\ 0 & 0 & 5/3 & 35/3 \end{bmatrix}$$

$$\frac{3}{5}R_3 \rightarrow R_3 \qquad \begin{bmatrix} 1 & 1/2 & -1/2 & 2 \\ 0 & 1 & -5/3 & -20/3 \\ 0 & 0 & 1 & 7 \end{bmatrix}$$

Let's take a break here and think about the state of our linear system at this moment. Converting back to linear equations, we now know

$$\begin{array}{rcl} x_1 + 1/2x_2 - 1/2x_3 & = & 2 \\ x_2 - 5/3x_3 & = & -20/3 \\ x_3 & = & 7 \end{array}.$$

Since we know that $x_3 = 7$, the second equation turns into

$$x_2 - (5/3)(7) = -20/3,$$

telling us that $x_2 = 5$.

Finally, knowing values for x_2 and x_3 lets us substitute in the first equation and find

$$x_1 + (1/2)(5) - (1/2)(7) = 2,$$

so $x_1 = 3$.

This process of substituting known values back into other equations is called *back substitution.* This process is essentially what happens when we perform the backward steps of Gaussian elimination. We make note of this below as we finish out finding the reduced row echelon form of our matrix.

$$\begin{array}{c} \frac{5}{3}R_3 + R_2 \rightarrow R_2 \\ \text{(knowing } x_3 = 7 \text{ allows us} \\ \text{to find } x_2 = 5) \end{array} \qquad \begin{bmatrix} 1 & 1/2 & -1/2 & 2 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 7 \end{bmatrix}$$

$$\begin{array}{c} \frac{1}{2}R_3 + R_1 \rightarrow R_1 \\ -\frac{1}{2}R_2 + R_1 \rightarrow R_1 \\ \text{(knowing } x_2 = 5 \text{ and } x_3 = 7 \\ \text{allows us to find } x_1 = 3) \end{array} \qquad \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 7 \end{bmatrix}$$

We did our operations slightly "out of order" in that we didn't put the zeros above our leading 1 in the third column in the same step, highlighting how back substitution works.

In all of our practice, we've only encountered systems of linear equations with exactly one solution. Is this always going to be the case? Could we ever have systems with more than one solution? If so, how many solutions could there be? Could we have systems without a solution? These are some of the questions we'll address in the next section.

Exercises 4.3

Problems

In Exercises 1 – 4, state whether or not the given matrices are in reduced row echelon form. If it is not, state why.

1. (a) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$

2. (a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

3. (a) $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

   (b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

   (c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

   (d) $\begin{bmatrix} 1 & 0 & 0 & -5 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & 3 \end{bmatrix}$

4. (a) $\begin{bmatrix} 2 & 0 & 0 & 2 \\ 0 & 2 & 0 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix}$

   (b) $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

   (c) $\begin{bmatrix} 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

   (d) $\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$

In Exercises 5 – 22, use Gaussian Elimination to put the given matrix into reduced row echelon form.

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

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

7. $\begin{bmatrix} 4 & 12 \\ -2 & -6 \end{bmatrix}$

8. $\begin{bmatrix} -5 & 7 \\ 10 & 14 \end{bmatrix}$

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

10. $\begin{bmatrix} 7 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}$

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

12. $\begin{bmatrix} 4 & 5 & -6 \\ -12 & -15 & 18 \end{bmatrix}$

13. $\begin{bmatrix} -2 & -4 & -8 \\ -2 & -3 & -5 \\ 2 & 3 & 6 \end{bmatrix}$

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

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

16. $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 1 & 6 & 9 \end{bmatrix}$

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

18. $\begin{bmatrix} 2 & -1 & 1 & 5 \\ 3 & 1 & 6 & -1 \\ 3 & 0 & 5 & 0 \end{bmatrix}$

19. $\begin{bmatrix} 1 & 1 & -1 & 7 \\ 2 & 1 & 0 & 10 \\ 3 & 2 & -1 & 17 \end{bmatrix}$

20. $\begin{bmatrix} 4 & 1 & 8 & 15 \\ 1 & 1 & 2 & 7 \\ 3 & 1 & 5 & 11 \end{bmatrix}$

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

22. $\begin{bmatrix} 1 & -1 & 3 & 1 & -2 & 9 \\ 2 & -2 & 6 & 1 & -2 & 13 \end{bmatrix}$

4.4 Existence and Uniqueness of Solutions

AS YOU READ . . .

1. T/F: It is possible for a linear system to have exactly 5 solutions.
2. T/F: A variable that corresponds to a leading 1 is "free."
3. How can one tell what kind of solution a linear system of equations has?
4. Give an example (different from those given in the text) of a 2 equation, 2 unknown linear system that is not consistent.
5. T/F: A particular solution for a linear system with infinite solutions can be found by arbitrarily picking values for the free variables.

So far, whenever we have solved a system of linear equations, we have always found exactly one solution. This is not always the case; we will find in this section that some systems do not have a solution, and others have more than one.

We start with a very simple example. Consider the following linear system:

$$x - y = 0.$$

There are obviously infinite solutions to this system; as long as $x = y$, we have a solution. We can picture all of these solutions by thinking of the graph of the equation $y = x$ on the traditional x, y coordinate plane.

Let's continue this visual aspect of considering solutions to linear systems. Consider the system

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

Each of these equations can be viewed as lines in the coordinate plane, and since their slopes are different, we know they will intersect somewhere (see Figure 4.1 (a)). In this example, they intersect at the point $(1, 1)$ – that is, when $x = 1$ and $y = 1$, both equations are satisfied and we have a solution to our linear system. Since this is the only place the two lines intersect, this is the only solution.

Now consider the linear system

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

It is clear that while we have two equations, they are essentially the same equation; the second is just a multiple of the first. Therefore, when we graph the two equations, we are graphing the same line twice (see Figure 4.1 (b); the thicker line is used to represent drawing the line twice). In this case, we have an infinite solution set, just as if we only had the one equation $x + y = 1$. We often write the solution as $x = 1 - y$ to demonstrate that y can be any real number, and x is determined once we pick a value for y.

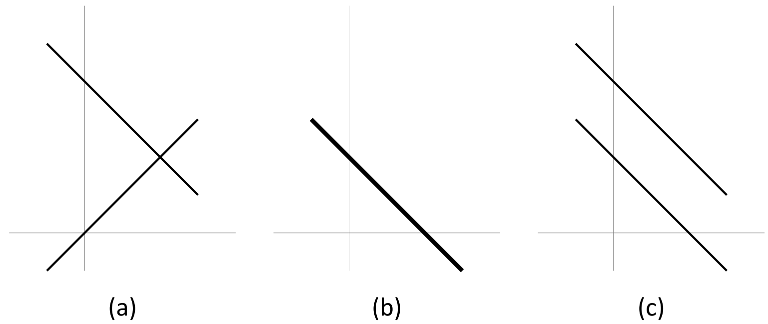


Figure 4.1: The three possibilities for two linear equations with two unknowns.

Finally, consider the linear system

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

We should immediately spot a problem with this system; if the sum of x and y is 1, how can it also be 2? There is no solution to such a problem; this linear system has no solution. We can visualize this situation in Figure 4.1 (c); the two lines are parallel and never intersect.

If we were to consider a linear system with three equations and two unknowns, we could visualize the solution by graphing the corresponding three lines. We can picture that perhaps all three lines would meet at one point, giving exactly 1 solution; perhaps all three equations describe the same line, giving an infinite number of solutions; perhaps we have different lines, but they do not all meet at the same point, giving no solution. We further visualize similar situations with, say, 20 equations with two variables.

While it becomes harder to visualize when we add variables, no matter how many equations and variables we have, solutions to linear equations always come in one of three forms: exactly one solution, infinite solutions, or no solution. This is a fact that we will not prove here, but it deserves to be stated.

Theorem 13 Solution Forms of Linear Systems

Every linear system of equations has exactly one solution, infinite solutions, or no solution.

This leads us to a definition. Here we don't differentiate between having one solution and infinite solutions, but rather just whether or not a solution exists.

Definition 39 Consistent and Inconsistent Linear Systems

A system of linear equations is *consistent* if it has a solution (perhaps more than one). A linear system is *inconsistent* if it does not have a solution.

How can we tell what kind of solution (if one exists) a given system of linear equations has? The answer to this question lies with properly understanding the reduced row echelon form of a matrix. To discover what the solution is to a linear system, we first put the matrix into reduced row echelon form and then interpret that form properly.

Before we start with a simple example, let us make a note about finding the reduced row echelon form of a matrix.

Technology Note: In the previous section, we learned how to find the reduced row echelon form of a matrix using Gaussian elimination – by hand. We need to know how to do this; understanding the process has benefits. However, actually executing the process by hand for every problem is not usually beneficial. In fact, with large systems, computing the reduced row echelon form by hand is effectively impossible. Our main concern is *what* "the rref" is, not what exact steps were used to arrive there. Therefore, the reader is encouraged to employ some form of technology to find the reduced row echelon form. Computer programs such as *Mathematica*, MATLAB, Maple, and Derive can be used; many handheld calculators (such as Texas Instruments calculators) will perform these calculations very quickly.

As a general rule, when we are learning a new technique, it is best to not use technology to aid us. This helps us learn not only the technique but some of its "inner workings." We can then use technology once we have mastered the technique and are now learning how to use it to solve problems.

From here on out, in our examples, when we need the reduced row echelon form of a matrix, we will not show the steps involved. Rather, we will give the initial matrix, then immediately give the reduced row echelon form of the matrix. We trust that the reader can verify the accuracy of this form by both performing the necessary steps by hand or utilizing some technology to do it for them.

Our first example explores officially a quick example used in the introduction of this section.

Example 87 **Solving a linear system**
Find the solution to the linear system

$$\begin{array}{ccccc} x_1 & + & x_2 & = & 1 \\ 2x_1 & + & 2x_2 & = & 2 \end{array}.$$

Solution Create the corresponding augmented matrix, and then put the matrix into reduced row echelon form.

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Now convert the reduced matrix back into equations. In this case, we only have one equation,

$$x_1 + x_2 = 1$$

or, equivalently,

$$\begin{aligned} &x_1 = 1 - x_2 \\ &x_2 \text{ is free.} \end{aligned}$$

We have just introduced a new term, the word *free*. It is used to stress that idea that x_2 can take on *any* value; we are "free" to choose any value for x_2. Once this value is chosen, the value of x_1 is determined. We have infinite choices for the value of x_2, so therefore we have infinite solutions.

For example, if we set $x_2 = 0$, then $x_1 = 1$; if we set $x_2 = 5$, then $x_1 = -4$.

Let's try another example, one that uses more variables.

Example 88 **Solving another linear system**
Find the solution to the linear system

$$\begin{array}{ccccccc} & & x_2 & - & x_3 & = & 3 \\ x_1 & & & + & 2x_3 & = & 2 \\ & & -3x_2 & + & 3x_3 & = & -9 \end{array}.$$

Solution To find the solution, put the corresponding matrix into reduced row echelon form.

$$\begin{bmatrix} 0 & 1 & -1 & 3 \\ 1 & 0 & 2 & 2 \\ 0 & -3 & 3 & -9 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Now convert this reduced matrix back into equations. We have

$$\begin{aligned} x_1 + 2x_3 &= 2 \\ x_2 - x_3 &= 3 \end{aligned}$$

or, equivalently,

$$\begin{aligned} &x_1 = 2 - 2x_3 \\ &x_2 = 3 + x_3 \\ &x_3 \text{ is free.} \end{aligned}$$

These two equations tell us that the values of x_1 and x_2 depend on what x_3 is. As we saw before, there is no restriction on what x_3 must be; it is "free" to take on the value of any real number. Once x_3 is chosen, we have a solution. Since we have infinite choices for the value of x_3, we have infinitely many solutions.

As examples, $x_1 = 2$, $x_2 = 3$, $x_3 = 0$ is one solution; $x_1 = -2$, $x_2 = 5$, $x_3 = 2$ is another solution. Try plugging these values back into the original equations to verify that these indeed are solutions. (By the way, since infinitely many solutions exist, this system of equations is consistent.)

In the two previous examples we have used the word "free" to describe certain variables. What exactly is a free variable? How do we recognize which variables are free and which are not?

Look back to the reduced matrix in Example 86. Notice that there is only one leading 1 in that matrix, and that leading 1 corresponded to the x_1 variable. That told us that x_1 was *not* a free variable; since x_2 *did not* correspond to a leading 1, it was a free variable.

Look also at the reduced matrix in Example 87. There were two leading 1s in that matrix; one corresponded to x_1 and the other to x_2. This meant that x_1 and x_2 were not free variables; since there was not a leading 1 that corresponded to x_3, it was a free variable.

We formally define this and a few other terms in this following definition.

Definition 40 Dependent and Independent Variables

Consider the reduced row echelon form of an augmented matrix of a linear system of equations. Then:

a variable that corresponds to a leading 1 is a *basic*, or *dependent*, variable, and

a variable that does not correspond to a leading 1 is a *free*, or *independent*, variable.

One can probably see that "free" and "independent" are relatively synonymous. It follows that if a variable is not independent, it must be dependent; the word "basic" comes from connections to other areas of mathematics that we won't explore here.

These definitions help us understand when a consistent system of linear equations will have infinite solutions. If there are no free variables, then there is exactly one solution; if there are any free variables, there are infinite solutions.

Key Idea 15 Consistent Solution Types

A consistent linear system of equations will have exactly one solution if and only if there is a leading 1 for each variable in the system.

If a consistent linear system of equations has a free variable, it has infinite solutions.

If a consistent linear system has more variables than leading 1s, then the system will have infinite solutions.

A consistent linear system with more variables than equations will always have infinite solutions.

Note: Key Idea 15 applies only to *consistent* systems. If a system is *inconsistent*, then no solution exists and talking about free and basic variables is meaningless.

When a consistent system has only one solution, each equation that comes from the reduced row echelon form of the corresponding augmented matrix will contain exactly one variable. If the consistent system has infinite solutions, then there will be at least one equation coming from the reduced row echelon form that contains more than one variable. The "first" variable will be the basic (or dependent) variable; all others will be free variables.

We have now seen examples of consistent systems with exactly one solution and others with infinite solutions. How will we recognize that a system is inconsistent? Let's find out through an example.

Example 89 An inconsistent system

Find the solution to the linear system

$$\begin{array}{ccccccc} x_1 & + & x_2 & + & x_3 & = & 1 \\ x_1 & + & 2x_2 & + & x_3 & = & 2 \\ 2x_1 & + & 3x_2 & + & 2x_3 & = & 0 \end{array}.$$

Solution We start by putting the corresponding matrix into reduced row echelon form.

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 \\ 2 & 3 & 2 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Now let us take the reduced matrix and write out the corresponding equations. The first two rows give us the equations

$$\begin{aligned} x_1 + x_3 &= 0 \\ x_2 &= 0. \end{aligned}$$

So far, so good. However the last row gives us the equation

$$0x_1 + 0x_2 + 0x_3 = 1$$

or, more concisely, $0 = 1$. Obviously, this is not true; we have reached a contradiction. Therefore, no solution exists; this system is inconsistent.

In previous sections we have only encountered linear systems with unique solutions (exactly one solution). Now we have seen three more examples with different solution types. The first two examples in this section had infinite solutions, and the third had no solution. How can we tell if a system is inconsistent?

A linear system will be inconsistent only when it implies that 0 equals 1. We can tell if a linear system implies this by putting its corresponding augmented matrix into reduced row echelon form. If we have any row where all entries are 0 except for the entry in the last column, then the system implies 0=1. More succinctly, if we have a leading 1 in the last column of an augmented matrix, then the linear system has no solution.

Key Idea 16 Inconsistent Systems of Linear Equations

A system of linear equations is inconsistent if the reduced row echelon form of its corresponding augmented matrix has a leading 1 in the last column.

Example 90 Verifying that a system is inconsistent

Confirm that the linear system

$$\begin{array}{ccccc} x & + & y & = & 0 \\ 2x & + & 2y & = & 4 \end{array}$$

has no solution.

Solution We can verify that this system has no solution in two ways. First, let's just think about it. If $x+y=0$, then it stands to reason, by multiplying both sides of this equation by 2, that $2x+2y=0$. However, the second equation of our system says that $2x+2y=4$. Since $0 \neq 4$, we have a contradiction and hence our system has no solution. (We cannot possibly pick values for x and y so that $2x+2y$ equals both 0 and 4.)

Now let us confirm this using the prescribed technique from above. The reduced row echelon form of the corresponding augmented matrix is

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

We have a leading 1 in the last column, so therefore the system is inconsistent.

Let's summarize what we have learned up to this point. Consider the reduced row echelon form of the augmented matrix of a system of linear equations. (That sure seems like a mouthful in and of itself. However, it boils down to "look at the reduced form of the usual matrix.") If there is a leading 1 in the last column, the system has no solution. Otherwise, if there is a leading 1 for each variable, then there is exactly one solution; otherwise (i.e., there are free variables) there are infinite solutions.

Systems with exactly one solution or no solution are the easiest to deal with; systems with infinite solutions are a bit harder to deal with. Therefore, we'll do a little more practice. First, a definition: if there are infinite solutions, what do we call one of those infinite solutions?

Definition 41 Particular Solution

Consider a linear system of equations with infinite solutions. A *particular solution* is one solution out of the infinite set of possible solutions.

The easiest way to find a particular solution is to pick values for the free variables which then determines the values of the dependent variables. Again, more practice is called for.

Why include the row of zeros in Example 90? Rows of zeros sometimes appear "unexpectedly" in matrices after they have been put in reduced row echelon form. When this happens, we do learn *something*; it means that at least one equation was a combination of some of the others.

Example 91 Finding general and particular solutions

Give the solution to a linear system whose augmented matrix in reduced row echelon form is

$$\begin{bmatrix} 1 & -1 & 0 & 2 & 4 \\ 0 & 0 & 1 & -3 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and give two particular solutions.

Solution We can essentially ignore the third row; it does not divulge any information about the solution. The first and second rows can be rewritten as the following equations:

$$\begin{aligned} x_1 - x_2 + 2x_4 &= 4 \\ x_3 - 3x_4 &= 7. \end{aligned}$$

Notice how the variables x_1 and x_3 correspond to the leading 1s of the given matrix. Therefore x_1 and x_3 are dependent variables; all other variables (in this case, x_2 and x_4) are free variables.

We generally write our solution with the dependent variables on the left and independent variables and constants on the right. It is also a good practice to acknowledge the fact that our free variables are, in fact, free. So our final solution would look something like

$$\begin{aligned} &x_1 = 4 + x_2 - 2x_4 \\ &x_2 \text{ is free} \\ &x_3 = 7 + 3x_4 \\ &x_4 \text{ is free.} \end{aligned}$$

To find particular solutions, choose values for our free variables. There is no "right" way of doing this; we are "free" to choose whatever we wish.

By setting $x_2 = 0 = x_4$, we have the solution $x_1 = 4$, $x_2 = 0$, $x_3 = 7$, $x_4 = 0$. By setting $x_2 = 1$ and $x_4 = -5$, we have the solution $x_1 = 15$, $x_2 = 1$, $x_3 = -8$, $x_4 = -5$. It is easier to read this when are variables are listed vertically, so we repeat these solutions:

One particular solution is:	Another particular solution is:
$x_1 = 4$	$x_1 = 15$
$x_2 = 0$	$x_2 = 1$
$x_3 = 7$	$x_3 = -8$
$x_4 = 0.$	$x_4 = -5.$

Example 92 **Finding general and particular solutions**

Find the solution to a linear system whose augmented matrix in reduced row echelon form is

$$\begin{bmatrix} 1 & 0 & 0 & 2 & 3 \\ 0 & 1 & 0 & 4 & 5 \end{bmatrix}$$

and give two particular solutions.

Solution Converting the two rows into equations we have

$$\begin{aligned} x_1 + 2x_4 &= 3 \\ x_2 + 4x_4 &= 5. \end{aligned}$$

We see that x_1 and x_2 are our dependent variables, for they correspond to the leading 1s. Therefore, x_3 and x_4 are independent variables. This situation feels a little unusual, for x_3 doesn't appear in any of the equations above, but cannot overlook it; it is still a free variable since there is not a leading 1 that corresponds to it. We write our solution as:

$$\begin{aligned} x_1 &= 3 - 2x_4 \\ x_2 &= 5 - 4x_4 \\ x_3 &\text{ is free} \\ x_4 &\text{ is free.} \end{aligned}$$

What kind of situation would lead to a column of all zeros? To have such a column, the original matrix needed to have a column of all zeros, meaning that while we acknowledged the existence of a certain variable, we never actually used it in any equation. In practical terms, we could respond by removing the corresponding column from the matrix and just keep in mind that that variable is free. In very large systems, it might be hard to determine whether or not a variable is actually used and one would not worry about it.
When we learn about eigenvectors and eigenvalues, we will see that under certain circumstances this situation arises. In those cases we leave the variable in the system just to remind ourselves that it is there.

To find two particular solutions, we pick values for our free variables. Again, there is no "right" way of doing this (in fact, there are . . . infinite ways of doing this) so we give only an example here.

One particular solution is:	Another particular solution is:
$x_1 = 3$	$x_1 = 3 - 2\pi$
$x_2 = 5$	$x_2 = 5 - 4\pi$
$x_3 = 1000$	$x_3 = e^2$
$x_4 = 0.$	$x_4 = \pi.$

(In the second particular solution we picked "unusual" values for x_3 and x_4 just to highlight the fact that we can.)

Example 93 **Finding general and particular solutions**

Find the solution to the linear system

$$\begin{array}{ccccccc} x_1 & + & x_2 & + & x_3 & = & 5 \\ x_1 & - & x_2 & + & x_3 & = & 3 \end{array}$$

and give two particular solutions.

Solution The corresponding augmented matrix and its reduced row echelon form are given below.

$$\begin{bmatrix} 1 & 1 & 1 & 5 \\ 1 & -1 & 1 & 3 \end{bmatrix} \quad \xrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 1 & 4 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Converting these two rows into equations, we have

$$\begin{aligned} x_1 + x_3 &= 4 \\ x_2 &= 1 \end{aligned}$$

giving us the solution

$$\begin{aligned} &x_1 = 4 - x_3 \\ &x_2 = 1 \\ &x_3 \text{ is free.} \end{aligned}$$

Once again, we get a bit of an "unusual" solution; while x_2 is a dependent variable, it does not depend on any free variable; instead, it is always 1. (We can think of it as depending on the value of 1.) By picking two values for x_3, we get two particular solutions.

One particular solution is:	Another particular solution is:
$x_1 = 4$	$x_1 = 3$
$x_2 = 1$	$x_2 = 1$
$x_3 = 0$.	$x_3 = 1$.

The constants and coefficients of a matrix work together to determine whether a given system of linear equations has one, infinite, or no solution. The concept will be fleshed out more in later chapters, but in short, the coefficients determine whether a matrix will have exactly one solution or not. In the "or not" case, the constants determine whether or not infinite solutions or no solution exists. (So if a given linear system has exactly one solution, it will always have exactly one solution even if the constants are changed.) Let's look at an example to get an idea of how the values of constants and coefficients work together to determine the solution type.

Example 94 Solving a system with a variable coefficient

For what values of k will the given system have exactly one solution, infinite solutions, or no solution?

$$\begin{array}{ccccc} x_1 & + & 2x_2 & = & 3 \\ 3x_1 & + & kx_2 & = & 9 \end{array}$$

Solution We answer this question by forming the augmented matrix and starting the process of putting it into reduced row echelon form. Below we see the augmented matrix and one elementary row operation that starts the Gaussian elimination process.

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & k & 9 \end{bmatrix} \quad \xrightarrow{-3R_1 + R_2 \to R_2} \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & k-6 & 0 \end{bmatrix}$$

This is as far as we need to go. In looking at the second row, we see that if $k = 6$, then that row contains only zeros and x_2 is a free variable; we have infinite solutions. If $k \neq 6$, then our next step would be to make that second row, second column entry a leading one. We don't particularly care about the solution, only that we would have exactly one as both x_1 and x_2 would correspond to a leading one and hence be dependent variables.

Our final analysis is then this. If $k \neq 6$, there is exactly one solution; if $k = 6$, there are infinite solutions. In this example, it is not possible to have no solutions.

As an extension of the previous example, consider the similar augmented matrix where the constant 9 is replaced with a 10. Performing the same elementary row operation gives

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & k & 10 \end{bmatrix} \quad \xrightarrow[-3R_1 + R_2 \rightarrow R_2]{} \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & k-6 & 1 \end{bmatrix}.$$

As in the previous example, if $k \neq 6$, we can make the second row, second column entry a leading one and hence we have one solution. However, if $k = 6$, then our last row is $[0\ 0\ 1]$, meaning we have no solution.

We have been studying the solutions to linear systems mostly in an "academic" setting; we have been solving systems for the sake of solving systems. In the next section, we'll look at situations which create linear systems that need solving (i.e., "word problems").

Exercises 4.4

Problems

In Exercises 1 – 14, find the solution to the given linear system. If the system has infinite solutions, give 2 particular solutions.

1. $\begin{aligned} 2x_1 + 4x_2 &= 2 \\ x_1 + 2x_2 &= 1 \end{aligned}$

2. $\begin{aligned} -x_1 + 5x_2 &= 3 \\ 2x_1 - 10x_2 &= -6 \end{aligned}$

3. $\begin{aligned} x_1 + x_2 &= 3 \\ 2x_1 + x_2 &= 4 \end{aligned}$

4. $\begin{aligned} -3x_1 + 7x_2 &= -7 \\ 2x_1 - 8x_2 &= 8 \end{aligned}$

5. $\begin{aligned} 2x_1 + 3x_2 &= 1 \\ -2x_1 - 3x_2 &= 1 \end{aligned}$

6. $\begin{aligned} x_1 + 2x_2 &= 1 \\ -x_1 - 2x_2 &= 5 \end{aligned}$

7. $\begin{aligned} -2x_1 + 4x_2 + 4x_3 &= 6 \\ x_1 - 3x_2 + 2x_3 &= 1 \end{aligned}$

8. $\begin{aligned} -x_1 + 2x_2 + 2x_3 &= 2 \\ 2x_1 + 5x_2 + x_3 &= 2 \end{aligned}$

9. $\begin{aligned} -x_1 - x_2 + x_3 + x_4 &= 0 \\ -2x_1 - 2x_2 + x_3 &= -1 \end{aligned}$

10. $\begin{aligned} x_1 + x_2 + 6x_3 + 9x_4 &= 0 \\ -x_1 - x_3 - 2x_4 &= -3 \end{aligned}$

11. $\begin{aligned} 2x_1 + x_2 + 2x_3 &= 0 \\ x_1 + x_2 + 3x_3 &= 1 \\ 3x_1 + 2x_2 + 5x_3 &= 3 \end{aligned}$

12. $\begin{aligned} x_1 + 3x_2 + 3x_3 &= 1 \\ 2x_1 - x_2 + 2x_3 &= -1 \\ 4x_1 + 5x_2 + 8x_3 &= 2 \end{aligned}$

13. $\begin{aligned} x_1 + 2x_2 + 2x_3 &= 1 \\ 2x_1 + x_2 + 3x_3 &= 1 \\ 3x_1 + 3x_2 + 5x_3 &= 2 \end{aligned}$

14. $\begin{aligned} 2x_1 + 4x_2 + 6x_3 &= 2 \\ 1x_1 + 2x_2 + 3x_3 &= 1 \\ -3x_1 - 6x_2 - 9x_3 &= -3 \end{aligned}$

In Exercises 15 – 18, state for which values of *k* the given system will have exactly 1 solution, infinite solutions, or no solution.

15. $\begin{aligned} x_1 + 2x_2 &= 1 \\ 2x_1 + 4x_2 &= k \end{aligned}$

16. $\begin{aligned} x_1 + 2x_2 &= 1 \\ x_1 + kx_2 &= 1 \end{aligned}$

17. $\begin{aligned} x_1 + 2x_2 &= 1 \\ x_1 + kx_2 &= 2 \end{aligned}$

18. $\begin{aligned} x_1 + 2x_2 &= 1 \\ x_1 + 3x_2 &= k \end{aligned}$

4.5 Applications of Linear Systems

AS YOU READ . . .

1. How do most problems appear "in the real world?"
2. The unknowns in a problem are also called what?
3. How many points are needed to determine the coefficients of a 5^{th} degree polynomial?

We've started this chapter by addressing the issue of finding the solution to a system of linear equations. In subsequent sections, we defined matrices to store linear equation information; we described how we can manipulate matrices without changing the solutions; we described how to efficiently manipulate matrices so that a working solution can be easily found.

We shouldn't lose sight of the fact that our work in the previous sections was aimed at finding solutions to systems of linear equations. In this section, we'll learn how to apply what we've learned to actually solve some problems.

Many, many, *many* problems that are addressed by engineers, businesspeople, scientists and mathematicians can be solved by properly setting up systems of linear equations. In this section we highlight only a few of the wide variety of problems that matrix algebra can help us solve.

We start with a simple example.

Example 95 **Counting marbles, again**
A jar contains 100 blue, green, red and yellow marbles. There are twice as many yellow marbles as blue; there are 10 more blue marbles than red; the sum of the red and yellow marbles is the same as the sum of the blue and green. How many marbles of each color are there?

Solution Let's call the number of blue balls b, and the number of the other balls g, r and y, each representing the obvious. Since we know that we have 100 marbles, we have the equation

$$b + g + r + y = 100.$$

The next sentence in our problem statement allows us to create three more equations.

We are told that there are twice as many yellow marbles as blue. One of the following two equations is correct, based on this statement; which one is it?

$$2y = b \qquad \text{or} \qquad 2b = y$$

The first equation says that if we take the number of yellow marbles, then double it, we'll have the number of blue marbles. That is not what we were told. The second equation states that if we take the number of blue marbles, then double it, we'll have the number of yellow marbles. This *is* what we were told.

The next statement of "there are 10 more blue marbles as red" can be written as either

$$b = r + 10 \qquad \text{or} \qquad r = b + 10.$$

Which is it?

The first equation says that if we take the number of red marbles, then add 10, we'll have the number of blue marbles. This is what we were told. The next equation is wrong; it implies there are more red marbles than blue.

The final statement tells us that the sum of the red and yellow marbles is the same as the sum of the blue and green marbles, giving us the equation

$$r + y = b + g.$$

We have four equations; altogether, they are

$$\begin{aligned} b + g + r + y &= 100 \\ 2b &= y \\ b &= r + 10 \\ r + y &= b + g. \end{aligned}$$

We want to write these equations in a standard way, with all the unknowns on the left and the constants on the right. Let us also write them so that the variables appear in the same order in each equation (we'll use alphabetical order to make it simple). We now have

$$\begin{aligned} b + g + r + y &= 100 \\ 2b - y &= 0 \\ b - r &= 10 \\ -b - g + r + y &= 0 \end{aligned}$$

To find the solution, let's form the appropriate augmented matrix and put it into reduced row echelon form. We do so here, without showing the steps.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 100 \\ 2 & 0 & 0 & -1 & 0 \\ 1 & 0 & -1 & 0 & 10 \\ -1 & -1 & 1 & 1 & 0 \end{bmatrix} \quad \xrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 20 \\ 0 & 1 & 0 & 0 & 30 \\ 0 & 0 & 1 & 0 & 10 \\ 0 & 0 & 0 & 1 & 40 \end{bmatrix}$$

We interpret from the reduced row echelon form of the matrix that we have 20 blue, 30 green, 10 red and 40 yellow marbles.

Even if you had a bit of difficulty with the previous example, in reality, this type of problem is pretty simple. The unknowns were easy to identify, the equations were pretty straightforward to write (maybe a bit tricky for some), and only the necessary information was given.

Most problems that we face in the world do not approach us in this way; most problems do not approach us in the form of "Here is an equation. Solve it." Rather, most problems come in the form of:

> Here is a problem. I want the solution. To help, here is lots of information. It may be just enough; it may be too much; it may not be enough. You figure out what you need; just give me the solution.

Faced with this type of problem, how do we proceed? Like much of what we've done in the past, there isn't just one "right" way. However, there are a few steps that can guide us. You don't have to follow these steps, "step by step," but if you find that you are having difficulty solving a problem, working through these steps may help. (Note: while the principles outlined here will help one solve any type of problem, these steps are written specifically for solving problems that involve only linear equations.)

Key Idea 17 Mathematical Problem Solving

1. Understand the problem. What exactly is being asked?
2. Identify the unknowns. What are you trying to find? What units are involved?
3. Give names to your unknowns (these are your *variables*).
4. Use the information given to write as many equations as you can that involve these variables.
5. Use the equations to form an augmented matrix; use Gaussian elimination to put the matrix into reduced row echelon form.
6. Interpret the reduced row echelon form of the matrix to identify the solution.
7. Ensure the solution makes sense in the context of the problem.

Having identified some steps, let us put them into practice with some examples.

Example 96 Arranging seating
A concert hall has seating arranged in three sections. As part of a special promotion, guests will receive two of three prizes. Guests seated in the first and second sections will receive Prize A, guests seated in the second and third sections will receive Prize B, and guests seated in the first and third sections will receive Prize C. Concert promoters told the concert hall managers of their plans, and asked how many seats were in each section. (The promoters want to store prizes for each section separately for easier distribution.) The managers, thinking they were being helpful, told the promoters they would need 105 A prizes, 103 B prizes, and 88 C prizes, and have since been unavailable for further help. How many seats are in each section?

Solution Before we rush in and start making equations, we should be clear about what is being asked. The final sentence asks: "How many seats are in each section?" This tells us what our unknowns should be: we should name our unknowns for the number of seats in each section. Let x_1, x_2 and x_3 denote the number of seats in the first, second and third sections, respectively. This covers the first two steps of our general problem solving technique.

(It is tempting, perhaps, to name our variables for the number of prizes given away. However, when we think more about this, we realize that we already know this – that information is given to us. Rather, we should name our variables for the things we don't know.)

Having our unknowns identified and variables named, we now proceed to forming equations from the information given. Knowing that Prize A goes to guests in the first and second sections and that we'll need 105 of these prizes tells us

$$x_1 + x_2 = 105.$$

Proceeding in a similar fashion, we get two more equations,

$$x_2 + x_3 = 103 \quad \text{and} \quad x_1 + x_3 = 88.$$

Thus our linear system is

$$\begin{aligned} x_1 + x_2 &= 105 \\ x_2 + x_3 &= 103 \\ x_1 + x_3 &= 88 \end{aligned}$$

and the corresponding augmented matrix is

$$\begin{bmatrix} 1 & 1 & 0 & 105 \\ 0 & 1 & 1 & 103 \\ 1 & 0 & 1 & 88 \end{bmatrix}.$$

To solve our system, let's put this matrix into reduced row echelon form.

$$\begin{bmatrix} 1 & 1 & 0 & 105 \\ 0 & 1 & 1 & 103 \\ 1 & 0 & 1 & 88 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & 45 \\ 0 & 1 & 0 & 60 \\ 0 & 0 & 1 & 43 \end{bmatrix}$$

We can now read off our solution. The first section has 45 seats, the second has 60 seats, and the third has 43 seats.

Example 97 Determining river speed

A lady takes a 2-mile motorized boat trip down the Highwater River, knowing the trip will take 30 minutes. She asks the boat pilot "How fast does this river flow?" He replies "I have no idea, lady. I just drive the boat."

She thinks for a moment, then asks "How long does the return trip take?" He replies "The same; half an hour." She follows up with the statement, "Since both legs take the same time, you must not drive the boat at the same speed."

"Naw," the pilot said. "While I really don't know exactly how fast I go, I do know that since we don't carry any tourists, I drive the boat twice as fast."

The lady walks away satisfied; she knows how fast the river flows.

(How fast *does* it flow?)

Solution This problem forces us to think about what information is given and how to use it to find what we want to know. In fact, to find the solution, we'll find out extra information that we weren't asked for!

We are asked to find how fast the river is moving (step 1). To find this, we should recognize that, in some sense, there are three speeds at work in the boat trips: the speed of the river (which we want to find), the speed of the boat, and the speed that they actually travel at.

We know that each leg of the trip takes half an hour; if it takes half an hour to cover 2 miles, then they must be travelling at 4 mph, each way.

The other two speeds are unknowns, but they are related to the overall speeds. Let's call the speed of the river r and the speed of the boat b. (And we should be careful. From the conversation, we know that the boat travels at two different speeds. So we'll say that b represents the speed of the boat when it travels downstream, so $2b$ represents the speed of the boat when it travels upstream.) Let's let our speed be measured in the units of miles/hour (mph) as we used above (steps 2 and 3).

What is the rate of the people on the boat? When they are travelling downstream, their rate is the sum of the water speed and the boat speed. Since their overall speed is 4 mph, we have the equation $r + b = 4$.

When the boat returns going against the current, its overall speed is the rate of the boat minus the rate of the river (since the river is working against the boat). The overall trip is still taken at 4 mph, so we have the equation $2b - r = 4$. (Recall: the boat is travelling twice as fast as before.)

The corresponding augmented matrix is

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

Note that we decided to let the first column hold the coefficients of b.

Putting this matrix in reduced row echelon form gives us:

$$\begin{bmatrix} 1 & 1 & 4 \\ 2 & -1 & 4 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 8/3 \\ 0 & 1 & 4/3 \end{bmatrix}.$$

We finish by interpreting this solution: the speed of the boat (going downstream) is 8/3 mph, or $2.\overline{6}$ mph, and the speed of the river is 4/3 mph, or $1.\overline{3}$ mph. All we really wanted to know was the speed of the river, at about 1.3 mph.

Example 98 Fitting a quadratic curve

Find the equation of the quadratic function that goes through the points $(-1, 6)$, $(1, 2)$ and $(2, 3)$.

SOLUTION This may not seem like a "linear" problem since we are talking about a quadratic function, but closer examination will show that it really is.

We normally write quadratic functions as $y = ax^2 + bx + c$ where a, b and c are the coefficients; in this case, they are our unknowns. We have three points; consider the point $(-1, 6)$. This tells us directly that if $x = -1$, then $y = 6$. Therefore we know that $6 = a(-1)^2+b(-1)+c$. Writing this in a more standard form, we have the linear equation

$$a - b + c = 6.$$

The second point tells us that $a(1)^2 + b(1) + c = 2$, which we can simplify as $a+b+c = 2$, and the last point tells us $a(2)^2 + b(2) + c = 3$, or $4a+2b+c = 3$. Thus our linear system is

$$\begin{aligned} a - b + c &= 6 \\ a + b + c &= 2 \\ 4a + 2b + c &= 3. \end{aligned}$$

Again, to solve our system, we find the reduced row echelon form of the corresponding augmented matrix. We don't show the steps here, just the final result.

$$\begin{bmatrix} 1 & -1 & 1 & 6 \\ 1 & 1 & 1 & 2 \\ 4 & 2 & 1 & 3 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

This tells us that $a = 1$, $b = -2$ and $c = 3$, giving us the quadratic function $y = x^2 - 2x + 3$.

One thing interesting about the previous example is that it confirms for us something that we may have known for a while (but didn't know *why* it was true). Why do we need two points to find the equation of the line? Because in the equation of the a line, we have two unknowns, and hence we'll need two equations to find values for these unknowns.

A quadratic has three unknowns (the coefficients of the x^2 term and the x term, and the constant). Therefore we'll need three equations, and therefore we'll need three points.

What happens if we try to find the quadratic function that goes through 3 points that are all on the same line? The fast answer is that you'll get the equation of a line; there isn't a quadratic function that goes through 3 colinear points. Try it and see! (Pick easy points, like $(0, 0)$, $(1, 1)$ and $(2, 2)$. You'll find that the coefficient of the x^2 term is 0.)

Of course, we can do the same type of thing to find polynomials that go through 4, 5, etc., points. In general, if you are given $n + 1$ points, a polynomial that goes through all $n + 1$ points will have degree at most n.

Example 99 A money counting problem

A woman has 32 \$1, \$5 and \$10 bills in her purse, giving her a total of \$100. How many bills of each denomination does she have?

Solution Let's name our unknowns x, y and z for our ones, fives and tens, respectively (it is tempting to call them o, f and t, but o looks too much like 0). We know that there are a total of 32 bills, so we have the equation

$$x + y + z = 32.$$

We also know that we have \$100, so we have the equation

$$x + 5y + 10z = 100.$$

We have three unknowns but only two equations, so we know that we cannot expect a unique solution. Let's try to solve this system anyway and see what we get.

Putting the system into a matrix and then finding the reduced row echelon form, we have

$$\begin{bmatrix} 1 & 1 & 1 & 32 \\ 1 & 5 & 10 & 100 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & -\frac{5}{4} & 15 \\ 0 & 1 & \frac{9}{4} & 17 \end{bmatrix}.$$

Reading from our reduced matrix, we have the infinite solution set

$$\begin{aligned} x &= 15 + \frac{5}{4}z \\ y &= 17 - \frac{9}{4}z \\ z &\text{ is free.} \end{aligned}$$

While we do have infinite solutions, most of these solutions really don't make sense in the context of this problem. (Setting $z = \frac{1}{2}$ doesn't make sense, for having half a ten dollar bill doesn't give us \$5. Likewise, having $z = 8$ doesn't make sense, for then we'd have "-1" \$5 bills.) So we must make sure that our choice of z doesn't give us fractions of bills or negative amounts of bills.

To avoid fractions, z must be a multiple of 4 $(-4, 0, 4, 8, \ldots)$. Of course, $z \geq 0$ for a negative number wouldn't make sense. If $z = 0$, then we have 15 one dollar bills and 17 five dollar bills, giving us \$100. If $z = 4$, then we have $x = 20$ and $y = 8$. We already mentioned that $z = 8$ doesn't make sense, nor does any value of z where $z \geq 8$.

So it seems that we have two answers; one with $z = 0$ and one with $z = 4$. Of course, by the statement of the problem, we are led to believe that the lady has at least one \$10 bill, so probably the "best" answer is that we have 20 \$1 bills, 8 \$5 bills and 4 \$10 bills. The real point of this example, though, is to address how infinite solutions may appear in a real world situation, and how surprising things may result.

Example 100 Recreating a football score

In a football game, teams can score points through touchdowns worth 6 points, extra points (that follow touchdowns) worth 1 point, two point conversions (that also follow touchdowns) worth 2 points and field goals, worth 3 points. You are told that in a football game, the two competing teams scored on 7 occasions, giving a total score of 24 points. Each touchdown was followed by either a successful extra point or two point conversion. In what ways were these points scored?

Solution The question asks how the points were scored; we can interpret this as asking how many touchdowns, extra points, two point conversions and field goals were scored. We'll need to assign variable names to our unknowns; let t represent the number of **t**ouchdowns scored; let x represent the number of e**x**tra points scored, let w represent the number of t**w**o point conversions, and let f represent the number of **f**ield goals scored.

Now we address the issue of writing equations with these variables using the given information. Since we have a total of 7 scoring occasions, we know that

$$t + x + w + f = 7.$$

The total points scored is 24; considering the value of each type of scoring opportunity, we can write the equation

$$6t + x + 2w + 3f = 24.$$

Finally, we know that each touchdown was followed by a successful extra point or two point conversion. This is subtle, but it tells us that the number of touchdowns is equal to the sum of extra points and two point conversions. In other words,

$$t = x + w.$$

To solve our problem, we put these equations into a matrix and put the matrix into reduced row echelon form. Doing so, we find

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 7 \\ 6 & 1 & 2 & 3 & 24 \\ 1 & -1 & -1 & 0 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & 0.5 & 3.5 \\ 0 & 1 & 0 & 1 & 4 \\ 0 & 0 & 1 & -0.5 & -0.5 \end{bmatrix}.$$

Therefore, we know that

$$\begin{aligned} t &= 3.5 - 0.5f \\ x &= 4 - f \\ w &= -0.5 + 0.5f. \end{aligned}$$

We recognize that this means there are "infinite solutions," but of course most of these will not make sense in the context of a real football game. We must apply some logic to make sense of the situation.

Progressing in no particular order, consider the second equation, $x = 4 - f$. In order for us to have a positive number of extra points, we must have $f \leq 4$. (And of course, we need $f \geq 0$, too.) Therefore, right away we know we have a total of only 5 possibilities, where $f = 0, 1, 2, 3$ or 4.

From the first and third equations, we see that if f is an even number, then t and w will both be fractions (for instance, if $f = 0$, then $t = 3.5$) which does not make sense. Therefore, we are down to two possible solutions, $f = 1$ and $f = 3$.

If $f = 1$, we have 3 touchdowns, 3 extra points, no two point conversions, and (of course), 1 field goal. (Check to make sure that gives 24 points!) If $f = 3$, then we 2 touchdowns, 1 extra point, 1 two point conversion, and (of course) 3 field goals. Again, check to make sure this gives us 24 points. Also, we should check each solution to make sure that we have a total of 7 scoring occasions and that each touchdown could be followed by an extra point or a two point conversion.

We have seen a variety of applications of systems of linear equations. We would do well to remind ourselves of the ways in which solutions to linear systems come: there can be exactly one solution, infinite solutions, or no solutions. While we did see a few examples where it seemed like we had only 2 solutions, this was because we were restricting our solutions to "make sense" within a certain context.

We should also remind ourselves that linear equations are immensely important. The examples we considered here ask fundamentally simple questions like "How fast is the water moving?" or "What is the quadratic function that goes through these three points?" or "How were points in a football game scored?" The real "important" situations ask much more difficult questions that often require *thousands* of equations! (Gauss began the systematic study of solving systems of linear equations while trying to predict the next sighting of a comet; he needed to solve a system of linear equations that had 17 unknowns. Today, this a relatively easy situation to handle with the help of computers, but to do it by hand is a real pain.) Once we understand the fundamentals of solving systems of equations, we can move on to looking at solving bigger systems of equations; this text focuses on getting us to understand the fundamentals.

Exercises 4.5

Problems

In Exercises 1 – 5, find the solution of the given problem by:

(a) creating an appropriate system of linear equations

(b) forming the augmented matrix that corresponds to this system

(c) putting the augmented matrix into reduced row echelon form

(d) interpreting the reduced row echelon form of the matrix as a solution

1. A farmer looks out his window at his chickens and pigs. He tells his daughter that he sees 62 heads and 190 legs. How many chickens and pigs does the farmer have?

2. A lady buys 20 trinkets at a yard sale. The cost of each trinket is either \$0.30 or \$0.65. If she spends \$8.80, how many of each type of trinket does she buy?

3. A carpenter can make two sizes of table, grande and venti. The grande table requires 4 table legs and 1 table top; the venti requires 6 table legs and 2 table tops. After doing work, he counts up spare parts in his warehouse and realizes that he has 86 table tops left over, and 300 legs. How many tables of each kind can he build and use up exactly all of his materials?

4. A jar contains 100 marbles. We know there are twice as many green marbles as red; that the number of blue and yellow marbles together is the same as the number of green; and that three times the number of yellow marbles together with the red marbles gives the same numbers as the blue marbles. How many of each color of marble are in the jar?

5. A rescue mission has 85 sandwiches, 65 bags of chips and 210 cookies. They know from experience that men will eat 2 sandwiches, 1 bag of chips and 4 cookies; women will eat 1 sandwich, a bag of chips and 2 cookies; kids will eat half a sandwhich, a bag of chips and 3 cookies. If they want to use all their food up, how many men, women and kids can they feed?

In Exercises 6 – 15, find the polynomial with the smallest degree that goes through the given points.

6. $(1,3)$ and $(3,15)$

7. $(-2,14)$ and $(3,4)$

8. $(1,5)$, $(-1,3)$ and $(3,-1)$

9. $(-4,-3)$, $(0,1)$ and $(1,4.5)$

10. $(-1,-8)$, $(1,-2)$ and $(3,4)$

11. $(-3,3)$, $(1,3)$ and $(2,3)$

12. $(-2,15)$, $(-1,4)$, $(1,0)$ and $(2,-5)$

13. $(-2,-7)$, $(1,2)$, $(2,9)$ and $(3,28)$

14. $(-3,10)$, $(-1,2)$, $(1,2)$ and $(2,5)$

15. $(0,1)$, $(-3,-3.5)$, $(-2,-2)$ and $(4,7)$

16. The general exponential function has the form $f(x) = ae^{bx}$, where a and b are constants and e is Euler's constant ($\approx$ 2.718). We want to find the equation of the exponential function that goes through the points $(1,2)$ and $(2,4)$.
    (a) Show why we cannot simply subsitute in values for x and y in $y = ae^{bx}$ and solve using the techniques we used for polynomials.
    (b) Show how the equality $y = ae^{bx}$ leads us to the linear equation $\ln y = \ln a + bx$.
    (c) Use the techniques we developed to solve for the unknowns $\ln a$ and b.
    (d) Knowing $\ln a$, find a; find the exponential function $f(x) = ae^{bx}$ that goes through the points $(1,2)$ and $(2,4)$.

17. In a football game, 24 points are scored from 8 scoring occasions. The number of successful extra point kicks is equal to the number of successful two point conversions. Find all ways in which the points may have been scored in this game.

18. In a football game, 29 points are scored from 8 scoring occasions. There are 2 more successful extra point kicks than successful two point conversions. Find all ways in which the points may have been scored in this game.

19. In a basketball game, where points are scored either by a 3 point shot, a 2 point shot or a 1 point free throw, 80 points were scored from 30 successful shots. Find all ways in which the points may have been scored in this game.

20. In a basketball game, where points are scored either by a 3 point shot, a 2 point shot or a 1 point free throw, 110 points were scored from 70 successful shots. Find all ways in which the points may have been scored in this game.

21. Describe the equations of the linear functions that go through the point (1,3). Give 2 examples.

22. Describe the equations of the linear functions that go through the point (2,5). Give 2 examples.

23. Describe the equations of the quadratic functions that go through the points $(2,-1)$ and (1,0). Give 2 examples.

24. Describe the equations of the quadratic functions that go through the points $(-1,3)$ and (2,6). Give 2 examples.

5: Matrix Algebra

5.1 Vector Solutions to Linear Systems

AS YOU READ . . .

1. T/F: The equation $A\vec{x} = \vec{b}$ is just another way of writing a system of linear equations.
2. T/F: In solving $A\vec{x} = \vec{0}$, if there are 3 free variables, then the solution will be "pulled apart" into 3 vectors.
3. T/F: A homogeneous system of linear equations is one in which all of the coefficients are 0.
4. Whether or not the equation $A\vec{x} = \vec{b}$ has a solution depends on an intrinsic property of ________.

The first chapter of this text was spent finding solutions to systems of linear equations. We have spent the first two sections of this chapter learning operations that can be performed with matrices. One may have wondered "Are the ideas of the first chapter related to what we have been doing recently?" The answer is yes, these ideas are related. This section begins to show that relationship.

We have often hearkened back to previous algebra experience to help understand matrix algebra concepts. We do that again here. Consider the equation $ax = b$, where $a = 3$ and $b = 6$. If we asked one to "solve for x," what exactly would we be asking? We would want to find a number, which we call x, where a times x gives b; in this case, it is a number, when multiplied by 3, returns 6.

Now we consider matrix algebra expressions. We'll eventually consider solving equations like $AX = B$, where we know what the matrices A and B are and we want to find the matrix X. For now, we'll only consider equations of the type $A\vec{x} = \vec{b}$, where we know the matrix A and the vector $\vec{b}$. We will want to find what vector $\vec{x}$ satisfies this equation; we want to "solve for $\vec{x}$."

To help understand what this is asking, we'll consider an example. Let

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 2 & 0 & 1 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} \quad \text{and} \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

(We don't know what $\vec{x}$ is, so we have to represent it's entries with the variables x_1, x_2 and x_3.) Let's "solve for $\vec{x}$," given the equation $A\vec{x} = \vec{b}$.

We can multiply out the left hand side of this equation. We find that

$$A\vec{x} = \begin{bmatrix} x_1 + x_2 + x_3 \\ x_1 - x_2 + 2x_3 \\ 2x_1 + x_3 \end{bmatrix}.$$

Be sure to note that the product is just a vector; it has just one column.

Since $A\vec{x}$ is equal to $\vec{b}$, we have

$$\begin{bmatrix} x_1 + x_2 + x_3 \\ x_1 - x_2 + 2x_3 \\ 2x_1 + x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}.$$

Knowing that two vectors are equal only when their corresponding entries are equal, we know

$$\begin{aligned} x_1 + x_2 + x_3 &= 2 \\ x_1 - x_2 + 2x_3 &= -3 \\ 2x_1 + x_3 &= 1. \end{aligned}$$

This should look familiar; it is a system of linear equations! Given the matrix-vector equation $A\vec{x} = \vec{b}$, we can recognize A as the coefficient matrix from a linear system and $\vec{b}$ as the vector of the constants from the linear system. To solve a matrix–vector equation (and the corresponding linear system), we simply augment the matrix A with the vector $\vec{b}$, put this matrix into reduced row echelon form, and interpret the results.

We convert the above linear system into an augmented matrix and find the reduced row echelon form:

$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & -1 & 2 & -3 \\ 2 & 0 & 1 & 1 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

This tells us that $x_1 = 1$, $x_2 = 2$ and $x_3 = -1$, so

$$\vec{x} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

We should check our work; multiply out $A\vec{x}$ and verify that we indeed get $\vec{b}$:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \quad \text{does equal} \quad \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}.$$

Example 101 **Solving a matrix equation**

Solve the equation $A\vec{x} = \vec{b}$ for $\vec{x}$ where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}.$$

Solution The solution is rather straightforward, even though we did a lot of work before to find the answer. Form the augmented matrix $\begin{bmatrix} A & \vec{b} \end{bmatrix}$ and interpret its reduced row echelon form.

$$\begin{bmatrix} 1 & 2 & 3 & 5 \\ -1 & 2 & 1 & -1 \\ 1 & 1 & 0 & 2 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

In previous sections we were fine stating that the result as

$$x_1 = 2, \quad x_2 = 0, \quad x_3 = 1,$$

but we were asked to find $\vec{x}$; therefore, we state the solution as

$$\vec{x} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}.$$

This probably seems all well and good. While asking one to solve the equation $A\vec{x} = \vec{b}$ for $\vec{x}$ seems like a new problem, in reality it is just asking that we solve a system of linear equations. Our variables x_1, etc., appear not individually but as the entries of our vector $\vec{x}$. We are simply writing an old problem in a new way.

In line with this new way of writing the problem, we have a new way of writing the solution. Instead of listing, individually, the values of the unknowns, we simply list them as the elements of our vector $\vec{x}$.

These are important ideas, so we state the basic principle once more: solving the equation $A\vec{x} = \vec{b}$ for $\vec{x}$ is the same thing as solving a linear system of equations. Equivalently, any system of linear equations can be written in the form $A\vec{x} = \vec{b}$ for some matrix A and vector $\vec{b}$.

Since these ideas are equivalent, we'll refer to $A\vec{x} = \vec{b}$ both as a matrix–vector equation and as a system of linear equations: they are the same thing.

We've seen two examples illustrating this idea so far, and in both cases the linear system had exactly one solution. We know from Theorem 13 that any linear system has either one solution, infinite solutions, or no solution. So how does our new method of writing a solution work with infinite solutions and no solutions?

Certainly, if $A\vec{x} = \vec{b}$ has no solution, we simply say that the linear system has no solution. There isn't anything special to write. So the only other option to consider is the case where we have infinite solutions. We'll learn how to handle these situations through examples.

Example 102 **Finding the vector solution to a linear system**

Solve the linear system $A\vec{x} = \vec{0}$ for $\vec{x}$ and write the solution in vector form, where

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad \vec{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Solution (Note: we didn't really need to specify that $\vec{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, but we did just to eliminate any uncertainty.)

To solve this system, put the augmented matrix into reduced row echelon form, which we do below.

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 0 \end{bmatrix} \quad \xrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We interpret the reduced row echelon form of this matrix to write the solution as

$$\begin{aligned} x_1 &= -2x_2 \\ x_2 &\text{ is free.} \end{aligned}$$

We are not done; we need to write the solution in vector form, for our solution is the vector $\vec{x}$. Recall that

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

From above we know that $x_1 = -2x_2$, so we replace the x_1 in $\vec{x}$ with $-2x_2$. This gives our solution as

$$\vec{x} = \begin{bmatrix} -2x_2 \\ x_2 \end{bmatrix}.$$

Now we pull the x_2 out of the vector (it is just a scalar) and write $\vec{x}$ as

$$\vec{x} = x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

For reasons that will become more clear later, set

$$\vec{v} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

Thus our solution can be written as

$$\vec{x} = x_2\vec{v}.$$

Recall that since our system was consistent and had a free variable, we have infinite solutions. This form of the solution highlights this fact; pick any value for x_2 and we get a different solution.

For instance, by setting $x_2 = -1$, 0, and 5, we get the solutions

$$\vec{x} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} -10 \\ 5 \end{bmatrix},$$

respectively.

We should check our work; multiply each of the above vectors by A to see if we indeed get $\vec{0}$.

We have officially solved this problem; we have found the solution to $A\vec{x} = \vec{0}$ and written it properly. One final thing we will do here is *graph* the solution, using our skills learned in the previous section.

Our solution is

$$\vec{x} = x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

This means that any scalar multiply of the vector $\vec{v} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ is a solution; we know how to sketch the scalar multiples of $\vec{v}$. This is done in Figure 5.1.

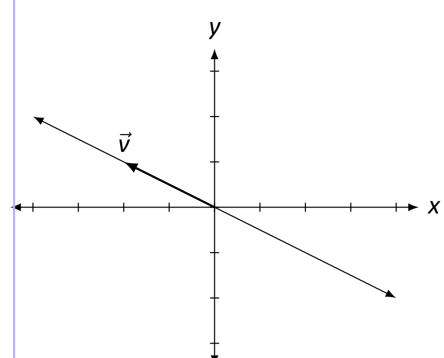


Figure 5.1: The solution, as a line, to $A\vec{x} = \vec{0}$ in Example 101.

Here vector $\vec{v}$ is drawn as well as the line that goes through the origin in the direction of $\vec{v}$. Any vector along this line is a solution. So in some sense, we can say that the solution to $A\vec{x} = \vec{0}$ is *a line*.

Let's practice this again.

Example 103 **Another matrix equation**
Solve the linear system $A\vec{x} = \vec{0}$ and write the solution in vector form, where

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

Solution Again, to solve this problem, we form the proper augmented matrix and we put it into reduced row echelon form, which we do below.

$$\begin{bmatrix} 2 & -3 & 0 \\ -2 & 3 & 0 \end{bmatrix} \quad \xrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & -3/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We interpret the reduced row echelon form of this matrix to find that

$$\begin{aligned} & x_1 = 3/2x_2 \\ & x_2 \text{ is free.} \end{aligned}$$

As before,

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Since $x_1 = 3/2x_2$, we replace x_1 in $\vec{x}$ with $3/2x_2$:

$$\vec{x} = \begin{bmatrix} 3/2x_2 \\ x_2 \end{bmatrix}.$$

Now we pull out the x_2 and write the solution as

$$\vec{x} = x_2 \begin{bmatrix} 3/2 \\ 1 \end{bmatrix}.$$

As before, let's set

$$\vec{v} = \begin{bmatrix} 3/2 \\ 1 \end{bmatrix}$$

so we can write our solution as

$$\vec{x} = x_2\vec{v}.$$

Again, we have infinite solutions; any choice of x_2 gives us one of these solutions. For instance, picking $x_2 = 2$ gives the solution

$$\vec{x} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

(This is a particularly nice solution, since there are no fractions. . .)

As in the previous example, our solutions are multiples of a vector, and hence we can graph this, as done in Figure 5.2.

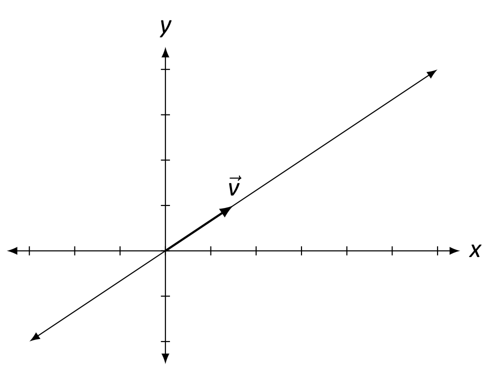


Figure 5.2: The solution, as a line, to $A\vec{x} = \vec{0}$ in Example 102.

Let's practice some more; this time, we won't solve a system of the form $A\vec{x} = \vec{0}$, but instead $A\vec{x} = \vec{b}$, for some vector $\vec{b}$.

Example 104 **A matrix equation with non-zero right-hand side**
Solve the linear system $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}.$$

Solution (Note that this is the same matrix A that we used in Example 101. This will be important later.)

Our methodology is the same as before; we form the augmented matrix and put it into reduced row echelon form.

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Interpreting this reduced row echelon form, we find that

$$x_1 = 3 - 2x_2$$
$$x_2 \text{ is free.}$$

Again,

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

and we replace x_1 with $3 - 2x_2$, giving

$$\vec{x} = \begin{bmatrix} 3 - 2x_2 \\ x_2 \end{bmatrix}.$$

This solution is different than what we've seen in the past two examples; we can't simply pull out a x_2 since there is a 3 in the first entry. Using the properties of matrix addition, we can "pull apart" this vector and write it as the sum of two vectors: one which contains only constants, and one that contains only "x_2 stuff." We do this below.

$$\begin{aligned} \vec{x} &= \begin{bmatrix} 3 - 2x_2 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \begin{bmatrix} -2x_2 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix}. \end{aligned}$$

Once again, let's give names to the different component vectors of this solution (we are getting near the explanation of why we are doing this). Let

$$\vec{x_p} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \quad \text{and} \quad \vec{v} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

We can then write our solution in the form

$$\vec{x} = \vec{x_p} + x_2\vec{v}.$$

We still have infinite solutions; by picking a value for x_2 we get one of these solutions. For instance, by letting $x_2 = -1$, 0, or 2, we get the solutions

$$\begin{bmatrix} 5 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

We have officially solved the problem; we have solved the equation $A\vec{x} = \vec{b}$ for $\vec{x}$ and have written the solution in vector form. As an additional visual aid, we will graph this solution.

Each vector in the solution can be written as the sum of two vectors: $\vec{x_p}$ and a multiple of $\vec{v}$. In Figure 5.3, $\vec{x_p}$ is graphed and $\vec{v}$ is graphed with its origin starting at the tip of $\vec{x_p}$. Finally, a line is drawn in the direction of $\vec{v}$ from the tip of $\vec{x_p}$; any vector pointing to any point on this line is a solution to $A\vec{x} = \vec{b}$.

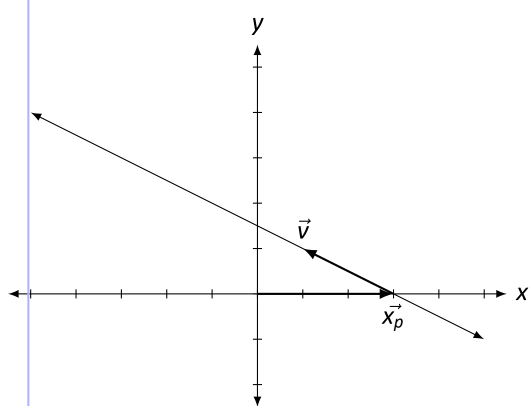


Figure 5.3: The solution, as a line, to $A\vec{x} = \vec{b}$ in Example 103.

The previous examples illustrate some important concepts. One is that we can "see" the solution to a system of linear equations in a new way. Before, when we had infinite solutions, we knew we could arbitrarily pick values for our free variables and get different solutions. We knew this to be true, and we even practiced it, but the result was not very "tangible." Now, we can view our solution as a vector; by picking different values for our free variables, we see this as multiplying certain important vectors by a scalar which gives a different solution.

Another important concept that these examples demonstrate comes from the fact that Examples 101 and 103 were only "slightly different" and hence had only "slightly different" answers. Both solutions had

$$x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

in them; in Example 103 the solution also had another vector added to this. Was this coincidence, or is there a definite pattern here?

Of course there is a pattern! Now . . . what exactly is it? First, we define a term.

Definition 42 Homogeneous Linear System of Equations

A system of linear equations is *homogeneous* if the constants in each equation are zero.

Note: a homogeneous system of equations can be written in vector form as $A\vec{x} = \vec{0}$.

The term *homogeneous* comes from two Greek words; *homo* meaning "same" and *genus* meaning "type." A homogeneous system of equations is a system in which each equation is of the same type – all constants are 0. Notice that the system of equations in Examples 101 and 103 are homogeneous.

Note that $A\vec{0} = \vec{0}$; that is, if we set $\vec{x} = \vec{0}$, we have a solution to a homogeneous set of equations. This fact is important; the zero vector is *always* a solution to a homogeneous linear system. Therefore a homogeneous system is always consistent; we need only to determine whether we have exactly one solution (just $\vec{0}$) or infinite solutions. This idea is important so we give it it's own box.

Key Idea 18 Homogeneous Systems and Consistency

All homogeneous linear systems are consistent.

How do we determine if we have exactly one or infinite solutions? Recall Key Idea 15: if the solution has any free variables, then it will have infinite solutions. How can we tell if the system has free variables? Form the augmented matrix $\begin{bmatrix} A & \vec{0} \end{bmatrix}$, put it into reduced row echelon form, and interpret the result.

It may seem that we've brought up a new question, "When does $A\vec{x} = \vec{0}$ have exactly one or infinite solutions?" only to answer with "Look at the reduced row echelon form of A and interpret the results, just as always." Why bring up a new question if the answer is an old one?

While the new question has an old solution, it does lead to a great idea. Let's refresh our memory; earlier we solved two linear systems,

$$A\vec{x} = \vec{0} \quad \text{and} \quad A\vec{x} = \vec{b}$$

where

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}.$$

The solution to the first system of equations, $A\vec{x} = \vec{0}$, is

$$\vec{x} = x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

and the solution to the second set of equations, $A\vec{x} = \vec{b}$, is

$$\vec{x} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix},$$

for all values of x_2.

Recalling our notation used earlier, set

$$\vec{x_p} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \quad \text{and let} \quad \vec{v} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

Thus our solution to the linear system $A\vec{x} = \vec{b}$ is

$$\vec{x} = \vec{x_p} + x_2\vec{v}.$$

Let us see how exactly this solution works; let's see why $A\vec{x}$ equals $\vec{b}$. Multiply $A\vec{x}$:

$$\begin{aligned} A\vec{x} &= A(\vec{x_p} + x_2\vec{v}) \\ &= A\vec{x_p} + A(x_2\vec{v}) \\ &= A\vec{x_p} + x_2(A\vec{v}) \\ &= A\vec{x_p} + x_2\vec{0} \\ &= A\vec{x_p} + \vec{0} \\ &= A\vec{x_p} \\ &= \vec{b} \end{aligned}$$

We know that the last line is true, that $A\vec{x_p} = \vec{b}$, since we know that $\vec{x}$ was a solution to $A\vec{x} = \vec{b}$. The whole point is that $\vec{x_p}$ itself is a solution to $A\vec{x} = \vec{b}$, and we could find more solutions by adding vectors "that go to zero" when multiplied by A. (The subscript p of "$\vec{x_p}$" is used to denote that this vector is a "particular" solution.)

Stated in a different way, let's say that we know two things: that $A\vec{x_p} = \vec{b}$ and $A\vec{v} = \vec{0}$. What is $A(\vec{x_p} + \vec{v})$? We can multiply it out:

$$\begin{aligned} A(\vec{x_p} + \vec{v}) &= A\vec{x_p} + A\vec{v} \\ &= \vec{b} + \vec{0} \\ &= \vec{b} \end{aligned}$$

and see that $A(\vec{x_p} + \vec{v})$ also equals $\vec{b}$.

So we wonder: does this mean that $A\vec{x} = \vec{b}$ will have infinite solutions? After all, if $\vec{x_p}$ and $\vec{x_p} + \vec{v}$ are both solutions, don't we have infinite solutions?

No. If $A\vec{x} = \vec{0}$ has exactly one solution, then $\vec{v} = \vec{0}$, and $\vec{x_p} = \vec{x_p} + \vec{v}$; we only have one solution.

So here is the culmination of all of our fun that started a few pages back. If $\vec{v}$ is a solution to $A\vec{x} = \vec{0}$ and $\vec{x_p}$ is a solution to $A\vec{x} = \vec{b}$, then $\vec{x_p} + \vec{v}$ is also a solution to $A\vec{x} = \vec{b}$. If $A\vec{x} = \vec{0}$ has infinite solutions, so does $A\vec{x} = \vec{b}$; if $A\vec{x} = \vec{0}$ has only one solution, so does $A\vec{x} = \vec{b}$. This culminating idea is of course important enough to be stated again.

Key Idea 19 Solutions of Consistent Systems

Let $A\vec{x} = \vec{b}$ be a consistent system of linear equations.

1. If $A\vec{x} = \vec{0}$ has exactly one solution ($\vec{x} = \vec{0}$), then $A\vec{x} = \vec{b}$ has exactly one solution.

2. If $A\vec{x} = \vec{0}$ has infinite solutions, then $A\vec{x} = \vec{b}$ has infinite solutions.

A key word in the above statement is *consistent*. If $A\vec{x} = \vec{b}$ is inconsistent (the linear system has no solution), then it doesn't matter how many solutions $A\vec{x} = \vec{0}$ has; $A\vec{x} = \vec{b}$ has no solution.

Example 105 Solving a homogeneous and non-homogeneous system

Let

$$A = \begin{bmatrix} 1 & -1 & 1 & 3 \\ 4 & 2 & 4 & 6 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}.$$

Solve the linear systems $A\vec{x} = \vec{0}$ and $A\vec{x} = \vec{b}$ for $\vec{x}$, and write the solutions in vector form.

Solution We'll tackle $A\vec{x} = \vec{0}$ first. We form the associated augmented matrix, put it into reduced row echelon form, and interpret the result.

$$\begin{bmatrix} 1 & -1 & 1 & 3 & 0 \\ 4 & 2 & 4 & 6 & 0 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix}$$

$$\begin{aligned} x_1 &= -x_3 - 2x_4 \\ x_2 &= x_4 \\ x_3 &\text{ is free} \\ x_4 &\text{ is free} \end{aligned}$$

To write our solution in vector form, we rewrite x_1 and x_2 in $\vec{x}$ in terms of x_3 and x_4.

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -x_3 - 2x_4 \\ x_4 \\ x_3 \\ x_4 \end{bmatrix}$$

Finally, we "pull apart" this vector into two vectors, one with the "x_3 stuff" and one with the "x_4 stuff."

$$\begin{aligned} \vec{x} &= \begin{bmatrix} -x_3 - 2x_4 \\ x_4 \\ x_3 \\ x_4 \end{bmatrix} \\ &= \begin{bmatrix} -x_3 \\ 0 \\ x_3 \\ 0 \end{bmatrix} + \begin{bmatrix} -2x_4 \\ x_4 \\ 0 \\ x_4 \end{bmatrix} \\ &= x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix} \\ &= x_3\vec{u} + x_4\vec{v} \end{aligned}$$

We use $\vec{u}$ and $\vec{v}$ simply to give these vectors names (and save some space).

It is easy to confirm that both $\vec{u}$ and $\vec{v}$ are solutions to the linear system $A\vec{x} = \vec{0}$. (Just multiply $A\vec{u}$ and $A\vec{v}$ and see that both are $\vec{0}$.) Since both are solutions to a homogeneous system of linear equations, any linear combination of $\vec{u}$ and $\vec{v}$ will be a solution, too.

Now let's tackle $A\vec{x} = \vec{b}$. Once again we put the associated augmented matrix into reduced row echelon form and interpret the results.

$$\begin{bmatrix} 1 & -1 & 1 & 3 & 1 \\ 4 & 2 & 4 & 6 & 10 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 1 & 2 & 2 \\ 0 & 1 & 0 & -1 & 1 \end{bmatrix}$$

$$\begin{aligned} x_1 &= 2 - x_3 - 2x_4 \\ x_2 &= 1 + x_4 \\ x_3 &\text{ is free} \\ x_4 &\text{ is free} \end{aligned}$$

Writing this solution in vector form gives

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 - x_3 - 2x_4 \\ 1 + x_4 \\ x_3 \\ x_4 \end{bmatrix}.$$

Again, we pull apart this vector, but this time we break it into three vectors: one with "x_3" stuff, one with "x_4" stuff, and one with just constants.

$$\begin{aligned} \vec{x} &= \begin{bmatrix} 2 - x_3 - 2x_4 \\ 1 + x_4 \\ x_3 \\ x_4 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -x_3 \\ 0 \\ x_3 \\ 0 \end{bmatrix} + \begin{bmatrix} -2x_4 \\ x_4 \\ 0 \\ x_4 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix} \\ &= \underbrace{\vec{x_p}}_{\substack{\text{particular} \\ \text{solution}}} + \underbrace{x_3\vec{u} + x_4\vec{v}}_{\substack{\text{solution to} \\ \text{homogeneous} \\ \text{equations } A\vec{x}=\vec{0}}} \end{aligned}$$

Note that $A\vec{x_p} = \vec{b}$; by itself, $\vec{x_p}$ is a solution. To get infinite solutions, we add a bunch of stuff that "goes to zero" when we multiply by A; we add the solution to the homogeneous equations.

Why don't we graph this solution as we did in the past? Before we had only two variables, meaning the solution could be graphed in 2D. Here we have four variables, meaning that our solution "lives" in 4D. You *can* draw this on paper, but it is *very* confusing.

Example 106 **Using matrices and vectors to solve a system of equations**

Rewrite the linear system

$$\begin{array}{rcrcrcrcrcr} x_1 & + & 2x_2 & - & 3x_3 & + & 2x_4 & + & 7x_5 & = & 2 \\ 3x_1 & + & 4x_2 & + & 5x_3 & + & 2x_4 & + & 3x_5 & = & -4 \end{array}$$

as a matrix–vector equation, solve the system using vector notation, and give the solution to the related homogeneous equations.

SOLUTION Rewriting the linear system in the form of $A\vec{x} = \vec{b}$, we have that

$$A = \begin{bmatrix} 1 & 2 & -3 & 2 & 7 \\ 3 & 4 & 5 & 2 & 3 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}.$$

To solve the system, we put the associated augmented matrix into reduced row echelon form and interpret the results.

$$\begin{bmatrix} 1 & 2 & -3 & 2 & 7 & 2 \\ 3 & 4 & 5 & 2 & 3 & -4 \end{bmatrix} \overrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 11 & -2 & -11 & -8 \\ 0 & 1 & -7 & 2 & 9 & 5 \end{bmatrix}$$

$$\begin{aligned} x_1 &= -8 - 11x_3 + 2x_4 + 11x_5 \\ x_2 &= 5 + 7x_3 - 2x_4 - 9x_5 \\ x_3 &\text{ is free} \\ x_4 &\text{ is free} \\ x_5 &\text{ is free} \end{aligned}$$

We use this information to write $\vec{x}$, again pulling it apart. Since we have three free variables and also constants, we'll need to pull $\vec{x}$ apart into four separate vectors.

$$\begin{aligned} \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} &= \begin{bmatrix} -8 - 11x_3 + 2x_4 + 11x_5 \\ 5 + 7x_3 - 2x_4 - 9x_5 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \\ &= \begin{bmatrix} -8 \\ 5 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -11x_3 \\ 7x_3 \\ x_3 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 2x_4 \\ -2x_4 \\ 0 \\ x_4 \\ 0 \end{bmatrix} + \begin{bmatrix} 11x_5 \\ -9x_5 \\ 0 \\ 0 \\ x_5 \end{bmatrix} \\ &= \begin{bmatrix} -8 \\ 5 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -11 \\ 7 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 11 \\ -9 \\ 0 \\ 0 \\ 1 \end{bmatrix} \\ &= \underbrace{\vec{x_p}}_{\substack{\text{particular} \\ \text{solution}}} + \underbrace{x_3\vec{u} + x_4\vec{v} + x_5\vec{w}}_{\substack{\text{solution to homogeneous} \\ \text{equations } A\vec{x} = \vec{0}}} \end{aligned}$$

So $\vec{x_p}$ is a particular solution; $A\vec{x_p} = \vec{b}$. (Multiply it out to verify that this is true.) The other vectors, $\vec{u}$, $\vec{v}$ and $\vec{w}$, that are multiplied by our free variables x_3, x_4 and x_5, are each solutions to the homogeneous equations, $A\vec{x} = \vec{0}$. Any linear combination of these three vectors, i.e., any vector found by choosing values for x_3, x_4 and x_5 in $x_3\vec{u} + x_4\vec{v} + x_5\vec{w}$ is a solution to $A\vec{x} = \vec{0}$.

Example 107 **Finding vector solutions**

Let

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}.$$

Find the solutions to $A\vec{x} = \vec{b}$ and $A\vec{x} = \vec{0}$.

Solution We go through the familiar work of finding the reduced row echelon form of the appropriate augmented matrix and interpreting the solution.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \overrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

$$x_1 = -1$$
$$x_2 = 2$$

Thus

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

This may strike us as a bit odd; we are used to having lots of different vectors in the solution. However, in this case, the linear system $A\vec{x} = \vec{b}$ has exactly one solution, and we've found it. What is the solution to $A\vec{x} = \vec{0}$? Since we've only found one solution to $A\vec{x} = \vec{b}$, we can conclude from Key Idea 19 the related homogeneous equations $A\vec{x} = \vec{0}$ have only one solution, namely $\vec{x} = \vec{0}$. We can write our solution vector $\vec{x}$ in a form similar to our previous examples to highlight this:

$$\begin{aligned} \vec{x} &= \begin{bmatrix} -1 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &= \underbrace{\vec{x_p}}_{\substack{\text{particular} \\ \text{solution}}} + \underbrace{\vec{0}}_{\substack{\text{solution to} \\ A\vec{x}=\vec{0}}} . \end{aligned}$$

Example 108 **Further vector solutions**

Let

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Find the solutions to $A\vec{x} = \vec{b}$ and $A\vec{x} = \vec{0}$.

SOLUTION To solve $A\vec{x} = \vec{b}$, we put the appropriate augmented matrix into reduced row echelon form and interpret the results.

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 1 \end{bmatrix} \quad \xrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

We immediately have a problem; we see that the second row tells us that $0x_1 + 0x_2 = 1$, the sign that our system does not have a solution. Thus $A\vec{x} = \vec{b}$ has no solution. Of course, this does not mean that $A\vec{x} = \vec{0}$ has no solution; it always has a solution.

To find the solution to $A\vec{x} = \vec{0}$, we interpret the reduced row echelon form of the appropriate augmented matrix.

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 0 \end{bmatrix} \quad \xrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$x_1 = -x_2$$
$$x_2 \text{ is free}$$

Thus

$$\begin{aligned} \vec{x} &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} -x_2 \\ x_2 \end{bmatrix} \\ &= x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= x_2\vec{u}. \end{aligned}$$

We have no solution to $A\vec{x} = \vec{b}$, but infinite solutions to $A\vec{x} = \vec{0}$.

The previous example may seem to violate the principle of Key Idea 19. After all, it seems that having infinite solutions to $A\vec{x} = \vec{0}$ should imply infinite solutions to $A\vec{x} = \vec{b}$. However, we remind ourselves of the key word in the idea that we observed before: *consistent*. If $A\vec{x} = \vec{b}$ is consistent and $A\vec{x} = \vec{0}$ has infinite solutions, then so will $A\vec{x} = \vec{b}$. But if $A\vec{x} = \vec{b}$ is not consistent, it does not matter how many solutions $A\vec{x} = \vec{0}$ has; $A\vec{x} = \vec{b}$ is still inconsistent.

This whole section is highlighting a very important concept that we won't fully understand until after two sections, but we get a glimpse of it here. When solving any system of linear equations (which we can write as $A\vec{x} = \vec{b}$), whether we have exactly one solution, infinite solutions, or no solution depends on an intrinsic property of A. We'll find out what that property is soon; in the next section we solve a problem we introduced at the beginning of this section, how to solve matrix equations $AX = B$.

Exercises 5.1

Problems

In Exercises 1 – 6, a matrix A and vectors $\vec{b}$, $\vec{u}$ and $\vec{v}$ are given. Verify that $\vec{u}$ and $\vec{v}$ are both solutions to the equation $A\vec{x} = \vec{b}$; that is, show that $A\vec{u} = A\vec{v} = \vec{b}$.

1. $A = \begin{bmatrix} 1 & -2 \\ -3 & 6 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \vec{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \vec{v} = \begin{bmatrix} -10 \\ -5 \end{bmatrix}$

2. $A = \begin{bmatrix} 1 & -2 \\ -3 & 6 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 2 \\ -6 \end{bmatrix}, \vec{u} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \vec{v} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

3. $A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \vec{u} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \vec{v} = \begin{bmatrix} 0 \\ 59 \end{bmatrix}$

4. $A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} -3 \\ -6 \end{bmatrix}, \vec{u} = \begin{bmatrix} -3 \\ -1 \end{bmatrix}, \vec{v} = \begin{bmatrix} -3 \\ 59 \end{bmatrix}$

5. $A = \begin{bmatrix} 0 & -3 & -1 & -3 \\ -4 & 2 & -3 & 5 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \vec{u} = \begin{bmatrix} 11 \\ 4 \\ -12 \\ 0 \end{bmatrix}$,
$\vec{v} = \begin{bmatrix} 9 \\ -12 \\ 0 \\ 12 \end{bmatrix}$

6. $A = \begin{bmatrix} 0 & -3 & -1 & -3 \\ -4 & 2 & -3 & 5 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 48 \\ 36 \end{bmatrix}, \vec{u} = \begin{bmatrix} -17 \\ -16 \\ 0 \\ 0 \end{bmatrix}$,
$\vec{v} = \begin{bmatrix} -8 \\ -28 \\ 0 \\ 12 \end{bmatrix}$

In Exercises 7 – 9, a matrix A and vectors $\vec{b}$, $\vec{u}$ and $\vec{v}$ are given. Verify that $A\vec{u} = \vec{0}$, $A\vec{v} = \vec{b}$ and $A(\vec{u} + \vec{v}) = \vec{b}$.

7. $A = \begin{bmatrix} 2 & -2 & -1 \\ -1 & 1 & -1 \\ -2 & 2 & -1 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \vec{u} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \vec{v} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$

8. $A = \begin{bmatrix} 1 & -1 & 3 \\ 3 & -3 & -3 \\ -1 & 1 & 1 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} -1 \\ -3 \\ 1 \end{bmatrix}, \vec{u} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \vec{v} = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$

9. $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & -3 \\ 3 & 1 & -3 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 2 \\ -4 \\ -1 \end{bmatrix}, \vec{u} = \begin{bmatrix} 0 \\ 6 \\ 2 \end{bmatrix}, \vec{v} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

In Exercises 10 – 24, a matrix A and vector $\vec{b}$ are given.

(a) Solve the equation $A\vec{x} = \vec{0}$.

(b) Solve the equation $A\vec{x} = \vec{b}$.

In each of the above, be sure to write your answer in vector format. Also, when possible, give 2 particular solutions to each equation.

10. $A = \begin{bmatrix} 0 & 2 \\ -1 & 3 \end{bmatrix}, \vec{b} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$

11. $A = \begin{bmatrix} -4 & -1 \\ -3 & -2 \end{bmatrix}, \vec{b} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}, \vec{b} = \begin{bmatrix} 0 \\ -5 \end{bmatrix}$

13. $A = \begin{bmatrix} 1 & 0 \\ 5 & -4 \end{bmatrix}, \vec{b} = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$

14. $A = \begin{bmatrix} 2 & -3 \\ -4 & 6 \end{bmatrix}, \vec{b} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

15. $A = \begin{bmatrix} -4 & 3 & 2 \\ -4 & 5 & 0 \end{bmatrix}, \vec{b} = \begin{bmatrix} -4 \\ -4 \end{bmatrix}$

16. $A = \begin{bmatrix} 1 & 5 & -2 \\ 1 & 4 & 5 \end{bmatrix}, \vec{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

17. $A = \begin{bmatrix} -1 & -2 & -2 \\ 3 & 4 & -2 \end{bmatrix}, \vec{b} = \begin{bmatrix} -4 \\ -4 \end{bmatrix}$

18. $A = \begin{bmatrix} 2 & 2 & 2 \\ 5 & 5 & -3 \end{bmatrix}, \vec{b} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$

19. $A = \begin{bmatrix} 1 & 5 & -4 & -1 \\ 1 & 0 & -2 & 1 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$

20. $A = \begin{bmatrix} -4 & 2 & -5 & 4 \\ 0 & 1 & -1 & 5 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$

21. $A = \begin{bmatrix} 0 & 0 & 2 & 1 & 4 \\ -2 & -1 & -4 & -1 & 5 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

22. $A = \begin{bmatrix} 3 & 0 & -2 & -4 & 5 \\ 2 & 3 & 2 & 0 & 2 \\ -5 & 0 & 4 & 0 & 5 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} -1 \\ -5 \\ 4 \end{bmatrix}$

23. $A = \begin{bmatrix} -1 & 3 & 1 & -3 & 4 \\ 3 & -3 & -1 & 1 & -4 \\ -2 & 3 & -2 & -3 & 1 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 1 \\ 1 \\ -5 \end{bmatrix}$

24. $A = \begin{bmatrix} -4 & -2 & -1 & 4 & 0 \\ 5 & -4 & 3 & -1 & 1 \\ 4 & -5 & 3 & 1 & -4 \end{bmatrix}$,
$\vec{b} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$

In Exercises 25 – 28, a matrix A and vector $\vec{b}$ are given. Solve the equation $A\vec{x} = \vec{b}$, write the solution in vector format, and sketch the solution as the appropriate line on the Cartesian plane.

25. $A = \begin{bmatrix} 2 & 4 \\ -1 & -2 \end{bmatrix}, \vec{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

26. $A = \begin{bmatrix} 2 & 4 \\ -1 & -2 \end{bmatrix}, \vec{b} = \begin{bmatrix} -6 \\ 3 \end{bmatrix}$

27. $A = \begin{bmatrix} 2 & -5 \\ -4 & -10 \end{bmatrix}, \vec{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

28. $A = \begin{bmatrix} 2 & -5 \\ -4 & -10 \end{bmatrix}, \vec{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

5.2 Solving Matrix Equations $AX = B$

AS YOU READ . . .

1. T/F: To solve the matrix equation $AX = B$, put the matrix $\begin{bmatrix} A & X \end{bmatrix}$ into reduced row echelon form and interpret the result properly.

2. T/F: The first column of a matrix product AB is A times the first column of B.

3. Give two reasons why one might solve for the columns of X in the equation AX=B separately.

We began last section talking about solving numerical equations like $ax = b$ for x. We mentioned that solving matrix equations of the form $AX = B$ is of interest, but we first learned how to solve the related, but simpler, equations $A\vec{x} = \vec{b}$. In this section we will learn how to solve the general matrix equation $AX = B$ for X.

We will start by considering the best case scenario when solving $A\vec{x} = \vec{b}$; that is, when A is square and we have exactly one solution. For instance, suppose we want to solve $A\vec{x} = \vec{b}$ where

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

We know how to solve this; put the appropriate matrix into reduced row echelon form and interpret the result.

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

We read from this that

$$\vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Written in a more general form, we found our solution by forming the augmented matrix

$$\begin{bmatrix} A & \vec{b} \end{bmatrix}$$

and interpreting its reduced row echelon form:

$$\begin{bmatrix} A & \vec{b} \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} I & \vec{x} \end{bmatrix}$$

Notice that when the reduced row echelon form of A is the identity matrix I we have exactly one solution. This, again, is the best case scenario.

We apply the same general technique to solving the matrix equation $AX = B$ for X. We'll assume that A is a square matrix (B need not be) and we'll form the augmented matrix

$$\begin{bmatrix} A & B \end{bmatrix}.$$

Putting this matrix into reduced row echelon form will give us X, much like we found $\vec{x}$ before.

$$\begin{bmatrix} A & B \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} I & X \end{bmatrix}$$

As long as the reduced row echelon form of A is the identity matrix, this technique works great. After a few examples, we'll discuss why this technique works, and we'll also talk just a little bit about what happens when the reduced row echelon form of A is not the identity matrix.

First, some examples.

Example 109 **Solving a matrix equation**
Solve the matrix equation $AX = B$ where

$$A = \begin{bmatrix} 1 & -1 \\ 5 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -8 & -13 & 1 \\ 32 & -17 & 21 \end{bmatrix}.$$

Solution To solve $AX = B$ for X, we form the proper augmented matrix, put it into reduced row echelon form, and interpret the result.

$$\begin{bmatrix} 1 & -1 & -8 & -13 & 1 \\ 5 & 3 & 32 & -17 & 21 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 1 & -7 & 3 \\ 0 & 1 & 9 & 6 & 2 \end{bmatrix}$$

We read from the reduced row echelon form of the matrix that

$$X = \begin{bmatrix} 1 & -7 & 3 \\ 9 & 6 & 2 \end{bmatrix}.$$

We can easily check to see if our answer is correct by multiplying $ttaX$.

Example 110 **Another matrix equation**
Solve the matrix equation $AX = B$ where

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & -2 \\ 2 & -1 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1 & 2 \\ 2 & -6 \\ 2 & -4 \end{bmatrix}.$$

Solution To solve, let's again form the augmented matrix

$$\begin{bmatrix} A & B \end{bmatrix},$$

put it into reduced row echelon form, and interpret the result.

$$\begin{bmatrix} 1 & 0 & 2 & -1 & 2 \\ 0 & -1 & -2 & 2 & -6 \\ 2 & -1 & 0 & 2 & -4 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & -1 & 1 \end{bmatrix}$$

We see from this that

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

Why does this work? To see the answer, let's define five matrices.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \vec{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \vec{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \vec{w} = \begin{bmatrix} 5 \\ 6 \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} 1 & -1 & 5 \\ 1 & 1 & 6 \end{bmatrix}$$

Notice that $\vec{u}$, $\vec{v}$ and $\vec{w}$ are the first, second and third columns of X, respectively. Now consider this list of matrix products: $A\vec{u}$, $A\vec{v}$, $A\vec{w}$ and AX.

$$\begin{aligned} A\vec{u} &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ 7 \end{bmatrix} \end{aligned} \qquad \begin{aligned} A\vec{v} &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} A\vec{w} &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} \\ &= \begin{bmatrix} 17 \\ 39 \end{bmatrix} \end{aligned} \qquad \begin{aligned} AX &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 & 5 \\ 1 & 1 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 1 & 17 \\ 7 & 1 & 39 \end{bmatrix} \end{aligned}$$

So again note that the columns of X are $\vec{u}$, $\vec{v}$ and $\vec{w}$; that is, we can write

$$X = \begin{bmatrix} \vec{u} & \vec{v} & \vec{w} \end{bmatrix}.$$

Notice also that the columns of AX are $A\vec{u}$, $A\vec{v}$ and $A\vec{w}$, respectively. Thus we can write

$$\begin{aligned} AX &= A\begin{bmatrix} \vec{u} & \vec{v} & \vec{w} \end{bmatrix} \\ &= \begin{bmatrix} A\vec{u} & A\vec{v} & A\vec{w} \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 3 \\ 7 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \end{bmatrix} & \begin{bmatrix} 17 \\ 39 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 3 & 1 & 17 \\ 7 & 1 & 39 \end{bmatrix} \end{aligned}$$

We summarize what we saw above in the following statement:

The columns of a matrix product AX are A times the columns of X.

How does this help us solve the matrix equation $AX = B$ for X? Assume that A is a square matrix (that forces X and B to be the same size). We'll let $\vec{x_1}, \vec{x_2}, \cdots \vec{x_n}$ denote the columns of the (unknown) matrix X, and we'll let $\vec{b_1}, \vec{b_2}, \cdots \vec{b_n}$ denote the columns of B. We want to solve $AX = B$ for X. That is, we want X where

$$\begin{aligned} AX &= B \\ A\begin{bmatrix} \vec{x_1} & \vec{x_2} & \cdots & \vec{x_n} \end{bmatrix} &= \begin{bmatrix} \vec{b_1} & \vec{b_2} & \cdots & \vec{b_n} \end{bmatrix} \\ \begin{bmatrix} A\vec{x_1} & A\vec{x_2} & \cdots & A\vec{x_n} \end{bmatrix} &= \begin{bmatrix} \vec{b_1} & \vec{b_2} & \cdots & \vec{b_n} \end{bmatrix} \end{aligned}$$

If the matrix on the left hand side is equal to the matrix on the right, then their respective columns must be equal. This means we need to solve n equations:

$$\begin{aligned} A\vec{x_1} &= \vec{b_1} \\ A\vec{x_2} &= \vec{b_2} \\ \vdots &= \vdots \\ A\vec{x_n} &= \vec{b_n} \end{aligned}$$

We already know how to do this; this is what we learned in the previous section. Let's do this in a concrete example. In our above work we defined matrices A and X, and looked at the product AX. Let's call the product B; that is, set B= AX. Now, let's pretend that we don't know what X is, and let's try to find the matrix X that satisfies the equation $AX = B$. As a refresher, recall that

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 1 & 17 \\ 7 & 1 & 39 \end{bmatrix}.$$

Since A is a 2×2 matrix and B is a 2×3 matrix, what dimensions must X be in the equation $AX = B$? The number of rows of X must match the number of columns of A; the number of columns of X must match the number of columns of B. Therefore we know that X must be a 2×3 matrix.

We'll call the three columns of X $\vec{x_1}$, $\vec{x_2}$ and $\vec{x_3}$. Our previous explanation tells us that if $AX = B$, then:

$$\begin{aligned} AX &= B \\ A\begin{bmatrix} \vec{x_1} & \vec{x_2} & \vec{x_3} \end{bmatrix} &= \begin{bmatrix} 3 & 1 & 17 \\ 7 & 1 & 39 \end{bmatrix} \\ \begin{bmatrix} A\vec{x_1} & A\vec{x_2} & A\vec{x_3} \end{bmatrix} &= \begin{bmatrix} 3 & 1 & 17 \\ 7 & 1 & 39 \end{bmatrix}. \end{aligned}$$

Hence

$$A\vec{x_1} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$$
$$A\vec{x_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$
$$A\vec{x_3} = \begin{bmatrix} 17 \\ 39 \end{bmatrix}$$

To find $\vec{x_1}$, we form the proper augmented matrix and put it into reduced row echelon form and interpret the results.

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 7 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

This shows us that

$$\vec{x_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

To find $\vec{x_2}$, we again form an augmented matrix and interpret its reduced row echelon form.

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & 4 & 1 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$$

Thus

$$\vec{x_2} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

which matches with what we already knew from above.

Before continuing on in this manner to find $\vec{x_3}$, we should stop and think. If the matrix vector equation $A\vec{x} = \vec{b}$ is consistent, then the steps involved in putting

$$\begin{bmatrix} A & \vec{b} \end{bmatrix}$$

into reduced row echelon form depend only on A; it does not matter what $\vec{b}$ is. So when we put the two matrices

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 7 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 1 \\ 3 & 4 & 1 \end{bmatrix}$$

from above into reduced row echelon form, we performed exactly the same steps! (In fact, those steps are: $-3R_1 + R_2 \to R_2$; $-\frac{1}{2}R_2 \to R_2$; $-2R_2 + R_1 \to R_1$.)

Instead of solving for each column of X separately, performing the same steps to put the necessary matrices into reduced row echelon form three different times, why don't we just do it all at once? (Unless you enjoy doing unnecessary work.) Instead of individually putting

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 7 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 1 \\ 3 & 4 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 17 \\ 3 & 4 & 39 \end{bmatrix}$$

into reduced row echelon form, let's just put

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 17 \\ 3 & 4 & 7 & 1 & 39 \end{bmatrix}$$

into reduced row echelon form.

$$\begin{bmatrix} 1 & 2 & 3 & 1 & 17 \\ 3 & 4 & 7 & 1 & 39 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 1 & -1 & 5 \\ 0 & 1 & 1 & 1 & 6 \end{bmatrix}$$

By looking at the last three columns, we see X:

$$X = \begin{bmatrix} 1 & -1 & 5 \\ 1 & 1 & 6 \end{bmatrix}.$$

Now that we've justified the technique we've been using in this section to solve $AX = B$ for X, we reinforce its importance by restating it as a Key Idea.

Key Idea 20 **Solving** $AX = B$

Let A be an $n \times n$ matrix, where the reduced row echelon form of A is I. To solve the matrix equation $AX = B$ for X,

1. Form the augmented matrix $\begin{bmatrix} A & B \end{bmatrix}$.
2. Put this matrix into reduced row echelon form. It will be of the form $\begin{bmatrix} I & X \end{bmatrix}$, where X appears in the columns where B once was.

These simple steps cause us to ask certain questions. First, we specify above that A should be a square matrix. What happens if A isn't square? Is a solution still possible? Secondly, we only considered cases where the reduced row echelon form of A was I (and stated that as a requirement in our Key Idea). What if the reduced row echelon form of A isn't I? Would we still be able to find a solution? (Instead of having exactly one solution, could we have no solution? Infinite solutions? How would we be able to tell?)

These questions are good to ask, and we leave it to the reader to discover their answers. Instead of tackling these questions, we instead tackle the problem of "Why do we care about solving $AX = B$?" The simple answer is that, for now, we only care about the special case when $B = I$. By solving $AX = I$ for X, we find a matrix X that, when multiplied by A, gives the identity I. That will be very useful.

Exercises 5.2

Problems

In Exercises 1 – 12, matrices A and B are given. Solve the matrix equation $AX = B$.

1. $A = \begin{bmatrix} 4 & -1 \\ -7 & 5 \end{bmatrix}$,
$B = \begin{bmatrix} 8 & -31 \\ -27 & 38 \end{bmatrix}$

2. $A = \begin{bmatrix} 1 & -3 \\ -3 & 6 \end{bmatrix}$,
$B = \begin{bmatrix} 12 & -10 \\ -27 & 27 \end{bmatrix}$

3. $A = \begin{bmatrix} 3 & 3 \\ 6 & 4 \end{bmatrix}$,
$B = \begin{bmatrix} 15 & -39 \\ 16 & -66 \end{bmatrix}$

4. $A = \begin{bmatrix} -3 & -6 \\ 4 & 0 \end{bmatrix}$,
$B = \begin{bmatrix} 48 & -30 \\ 0 & -8 \end{bmatrix}$

5. $A = \begin{bmatrix} -1 & -2 \\ -2 & -3 \end{bmatrix}$,
$B = \begin{bmatrix} 13 & 4 & 7 \\ 22 & 5 & 12 \end{bmatrix}$

6. $A = \begin{bmatrix} -4 & 1 \\ -1 & -2 \end{bmatrix}$,
$B = \begin{bmatrix} -2 & -10 & 19 \\ 13 & 2 & -2 \end{bmatrix}$

7. $A = \begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix}$, $B = I_2$

8. $A = \begin{bmatrix} 2 & 2 \\ 3 & 1 \end{bmatrix}$, $B = I_2$

9. $A = \begin{bmatrix} -2 & 0 & 4 \\ -5 & -4 & 5 \\ -3 & 5 & -3 \end{bmatrix}$,
$B = \begin{bmatrix} -18 & 2 & -14 \\ -38 & 18 & -13 \\ 10 & 2 & -18 \end{bmatrix}$

10. $A = \begin{bmatrix} -5 & -4 & -1 \\ 8 & -2 & -3 \\ 6 & 1 & -8 \end{bmatrix}$,
$B = \begin{bmatrix} -21 & -8 & -19 \\ 65 & -11 & -10 \\ 75 & -51 & 33 \end{bmatrix}$

11. $A = \begin{bmatrix} 0 & -2 & 1 \\ 0 & 2 & 2 \\ 1 & 2 & -3 \end{bmatrix}$, $B = I_3$

12. $A = \begin{bmatrix} -3 & 3 & -2 \\ 1 & -3 & 2 \\ -1 & -1 & 2 \end{bmatrix}$, $B = I_3$

5.3 The Matrix Inverse

AS YOU READ . . .

1. T/F: If A and B are square matrices where $AB = I$, then $BA = I$.
2. T/F: A matrix A has exactly one inverse, infinite inverses, or no inverse.
3. T/F: Everyone is special.
4. T/F: If A is invertible, then $A\vec{x} = \vec{0}$ has exactly 1 solution.
5. What is a corollary?
6. Fill in the blanks: ________ a matrix is invertible is useful; computing the inverse is ________.

Once again we visit the old algebra equation, $ax = b$. How do we solve for x? We know that, as long as $a \neq 0$,

$$x = \frac{b}{a}, \text{ or, stated in another way, } x = a^{-1}b.$$

What is a^{-1}? It is the number that, when multiplied by a, returns 1. That is,

$$a^{-1}a = 1.$$

Let us now think in terms of matrices. We have learned of the identity matrix I that "acts like the number 1." That is, if A is a square matrix, then

$$IA = AI = A.$$

If we had a matrix, which we'll call A^{-1}, where $A^{-1}A = I$, then by analogy to our algebra example above it seems like we might be able to solve the linear system $A\vec{x} = \vec{b}$ for $\vec{x}$ by multiplying both sides of the equation by A^{-1}. That is, perhaps

$$\vec{x} = A^{-1}\vec{b}.$$

Of course, there is a lot of speculation here. We don't know that such a matrix like A^{-1} exists. However, we do know how to solve the matrix equation $AX = B$, so we can use that technique to solve the equation $AX = I$ for X. This seems like it will get us close to what we want. Let's practice this once and then study our results.

Example 111 **Solving $AX = I$**

Let

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

Find a matrix X such that $AX = I$.

Solution We know how to solve this from the previous section: we form the proper augmented matrix, put it into reduced row echelon form and interpret the results.

$$\begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & -1 & 2 \end{bmatrix}$$

We read from our matrix that

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

Let's check our work:

$$\begin{aligned} AX &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= I \end{aligned}$$

Sure enough, it works.

Looking at our previous example, we are tempted to jump in and call the matrix X that we found "A^{-1}." However, there are two obstacles in the way of us doing this.

First, we know that in general $AB \neq BA$. So while we found that $AX = I$, we can't automatically assume that $XA = I$.

Secondly, we have seen examples of matrices where $AB = AC$, but $B \neq C$. So just because $AX = I$, it is possible that another matrix Y exists where $AY = I$. If this is the case, using the notation A^{-1} would be misleading, since it could refer to more than one matrix.

These obstacles that we face are not insurmountable. The first obstacle was that we know that $AX = I$ but didn't know that $XA = I$. That's easy enough to check, though. Let's look at A and X from our previous example.

$$\begin{aligned} XA &= \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= I \end{aligned}$$

Perhaps this first obstacle isn't much of an obstacle after all. Of course, we only have one example where it worked, so this doesn't mean that it always works. We have good news, though: it always does work. The only "bad" news to come with this is that this is a bit harder to prove. We won't worry about proving it always works, but state formally that it does in the following theorem.

Theorem 14 Special Commuting Matrix Products

Let A be an $n \times n$ matrix.

1. If there is a matrix X such that $AX = I_n$, then $XA = I_n$.
2. If there is a matrix X such that $XA = I_n$, then $AX = I_n$.

The second obstacle is easier to address. We want to know if another matrix Y exists where $AY = I = YA$. Let's suppose that it does. Consider the expression XAY. Since matrix multiplication is associative, we can group this any way we choose. We could group this as $(XA)Y$; this results in

$$\begin{aligned} (XA)Y &= IY \\ &= Y. \end{aligned}$$

We could also group XAY as $X(AY)$. This tells us

$$\begin{aligned} X(AY) &= XI \\ &= X \end{aligned}$$

Combining the two ideas above, we see that $X = XAY = Y$; that is, $X = Y$. We conclude that there is only one matrix X where $XA = I = AX$. (Even if we think we have two, we can do the above exercise and see that we really just have one.)

We have just proved the following theorem.

Theorem 15 Uniqueness of Solutions to $AX = I_n$

Let A be an $n \times n$ matrix and let X be a matrix where $AX = I_n$. Then X is unique; it is the only matrix that satisfies this equation.

So given a square matrix A, if we can find a matrix X where $AX = I$, then we know that $XA = I$ and that X is the only matrix that does this. This makes X special, so we give it a special name.

Definition 43 Invertible Matrices and the Inverse of A

Let A and X be $n \times n$ matrices where $AX = I = XA$. Then:

1. A is *invertible*.
2. X is the *inverse* of A, denoted by A^{-1}.

Example 111 shows that not all square matrices (or even non-zero square matrices) are invertible, hence Definition 43 is necessary: why bother calling A "invertible" if every square matrix is? If everyone is special, then no one is. Then again, everyone *is* special.

Let's do an example.

Example 112 A non-invertible matrix

Find the inverse of $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$.

Solution By solving the equation $AX = I$ for X will give us the inverse of A. Forming the appropriate augmented matrix and finding its reduced row echelon form gives us

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 2 & 0 & 1/2 \\ 0 & 0 & 1 & -1/2 \end{bmatrix}$$

Yikes! We were expecting to find that the reduced row echelon form of this matrix would look like

$$\begin{bmatrix} I & A^{-1} \end{bmatrix}.$$

However, we don't have the identity on the left hand side. Our conclusion: A is not invertible.

We have just seen that not all matrices are invertible.

With this thought in mind, let's complete the array of boxes we started before the example. We've discovered that if a matrix has an inverse, it has only one. Therefore, we gave that special matrix a name, "*the* inverse." Finally, we describe the most general way to find the inverse of a matrix, and a way to tell if it does not have one.

Key Idea 21 **Finding A^{-1}**

Let A be an $n \times n$ matrix. To find A^{-1}, put the augmented matrix

$$\begin{bmatrix} A & I_n \end{bmatrix}$$

into reduced row echelon form. If the result is of the form

$$\begin{bmatrix} I_n & X \end{bmatrix},$$

then $A^{-1} = X$. If not, (that is, if the first n columns of the reduced row echelon form are not I_n), then A is not invertible.

Let's try again.

Example 113 **Computing the inverse of a matrix**

Find the inverse, if it exists, of $A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$.

Solution We'll try to solve $AX = I$ for X and see what happens.

$$\begin{bmatrix} 1 & 1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 3 & 0 & 0 & 1 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 1 & 0 & 0.2 & -0.4 & 0.2 \\ 0 & 0 & 1 & -0.3 & 0.1 & 0.2 \end{bmatrix}$$

We have a solution, so

$$A = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.2 & -0.4 & 0.2 \\ -0.3 & 0.1 & 0.2 \end{bmatrix}.$$

Multiply AA^{-1} to verify that it is indeed the inverse of A.

We don't prove Theorem 16 here, but it really isn't hard to do. Put the matrix

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

into reduced row echelon form and you'll discover the result of the theorem. Alternatively, multiply A by what we propose is the inverse and see that we indeed get I.

In general, given a matrix A, to find A^{-1} we need to form the augmented matrix $\begin{bmatrix} A & I \end{bmatrix}$ and put it into reduced row echelon form and interpret the result. In the case of a 2×2 matrix, though, there is a shortcut. We give the shortcut in terms of a theorem.

Theorem 16 **The Inverse of a 2×2 Matrix**

Let

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

A is invertible if and only if $ad - bc \neq 0$.

If $ad - bc \neq 0$, then

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

We can't divide by 0, so if $ad - bc = 0$, we don't have an inverse. Recall Example 111, where

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

Here, $ad - bc = 1(4) - 2(2) = 0$, which is why A didn't have an inverse.

Although this idea is simple, we should practice it.

Example 114 **Computing a 2×2 inverse using Theorem 16**

Use Theorem 16 to find the inverse of

$$A = \begin{bmatrix} 3 & 2 \\ -1 & 9 \end{bmatrix}$$

if it exists.

SOLUTION Since $ad - bc = 29 \neq 0$, A^{-1} exists. By the Theorem,

$$\begin{aligned} A^{-1} &= \frac{1}{3(9) - 2(-1)} \begin{bmatrix} 9 & -2 \\ 1 & 3 \end{bmatrix} \\ &= \frac{1}{29} \begin{bmatrix} 9 & -2 \\ 1 & 3 \end{bmatrix} \end{aligned}$$

We can leave our answer in this form, or we could "simplify" it as

$$A^{-1} = \frac{1}{29} \begin{bmatrix} 9 & -2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 9/29 & -2/29 \\ 1/29 & 3/29 \end{bmatrix}.$$

We started this section out by speculating that just as we solved algebraic equations of the form $ax = b$ by computing $x = a^{-1}b$, we might be able to solve matrix equations of the form $A\vec{x} = \vec{b}$ by computing $\vec{x} = A^{-1}\vec{b}$. If A^{-1} does exist, then we *can* solve the equation $A\vec{x} = \vec{b}$ this way. Consider:

$$\begin{aligned} A\vec{x} &= \vec{b} && \text{(original equation)} \\ A^{-1}A\vec{x} &= A^{-1}\vec{b} && \text{(multiply both sides \textit{on the left} by } A^{-1}) \\ I\vec{x} &= A^{-1}\vec{b} && (\text{since } A^{-1}A = I) \\ \vec{x} &= A^{-1}\vec{b} && (\text{since } I\vec{x} = \vec{x}) \end{aligned}$$

Let's step back and think about this for a moment. The only thing we know about the equation $A\vec{x} = \vec{b}$ is that A is invertible. We also know that solutions to $A\vec{x} = \vec{b}$ come in three forms: exactly one solution, infinite solutions, and no solution. We just showed that if A is invertible, then $A\vec{x} = \vec{b}$ has *at least* one solution. We showed that by setting $\vec{x}$ equal to $A^{-1}\vec{b}$, we have a solution. Is it possible that more solutions exist?

No. Suppose we are told that a known vector $\vec{v}$ is a solution to the equation $A\vec{x} = \vec{b}$; that is, we know that $A\vec{v} = \vec{b}$. We can repeat the above steps:

$$\begin{aligned} A\vec{v} &= \vec{b} \\ A^{-1}A\vec{v} &= A^{-1}\vec{b} \\ I\vec{v} &= A^{-1}\vec{b} \\ \vec{v} &= A^{-1}\vec{b}. \end{aligned}$$

This shows that *all* solutions to $A\vec{x} = \vec{b}$ are exactly $\vec{x} = A^{-1}\vec{b}$ when A is invertible. We have just proved the following theorem.

Theorem 17 Invertible Matrices and Solutions to $A\vec{x}=\vec{b}$

Let A be an invertible $n\times n$ matrix, and let $\vec{b}$ be any $n\times 1$ column vector. Then the equation $A\vec{x}=\vec{b}$ has exactly one solution, namely

$$\vec{x}=A^{-1}\vec{b}.$$

A corollary to this theorem is: If A is not invertible, then $A\vec{x}=\vec{b}$ does not have exactly one solution. It may have infinite solutions and it may have no solution, and we would need to examine the reduced row echelon form of the augmented matrix $\begin{bmatrix} A & \vec{b}\end{bmatrix}$ to see which case applies.

A *corollary* is an idea that follows directly from a theorem, typically all the hard work is done in proving the theorem, and any resulting corollaries are easy consequences.

We demonstrate our theorem with an example.

Example 115 Using a matrix inverse to solve a system

Solve $A\vec{x}=\vec{b}$ by computing $\vec{x}=A^{-1}\vec{b}$, where

$$A=\begin{bmatrix} 1 & 0 & -3\\ -3 & -4 & 10\\ 4 & -5 & -11\end{bmatrix}\quad\text{and}\quad \vec{b}=\begin{bmatrix}-15\\ 57\\ -46\end{bmatrix}.$$

The method employed in Example 114 is useful in theory, but not in practice: the amount of work required to solve a system this way is significantly greater than the amount of work involved in Gaussian Elimination. The only scenario where you should consider using the inverse to solve a system (aside from being asked to do so on a test!) is when there are *several* systems you need to solve that all have the same coefficient matrix.

Solution Without showing our steps, we compute

$$A^{-1}=\begin{bmatrix} 94 & 15 & -12\\ 7 & 1 & -1\\ 31 & 5 & -4\end{bmatrix}.$$

We then find the solution to $A\vec{x}=\vec{b}$ by computing $A^{-1}\vec{b}$:

$$\begin{aligned}\vec{x}&=A^{-1}\vec{b}\\ &=\begin{bmatrix} 94 & 15 & -12\\ 7 & 1 & -1\\ 31 & 5 & -4\end{bmatrix}\begin{bmatrix}-15\\ 57\\ -46\end{bmatrix}\\ &=\begin{bmatrix}-3\\ -2\\ 4\end{bmatrix}.\end{aligned}$$

We can easily check our answer:

$$\begin{bmatrix} 1 & 0 & -3\\ -3 & -4 & 10\\ 4 & -5 & -11\end{bmatrix}\begin{bmatrix}-3\\ -2\\ 4\end{bmatrix}=\begin{bmatrix}-15\\ 57\\ -46\end{bmatrix}.$$

As odd as it may sound, *knowing* a matrix is invertible is useful; actually computing the inverse isn't. This is discussed at the end of the next section.

Knowing a matrix is invertible is incredibly useful. Among many other reasons, if you know A is invertible, then you know for sure that $A\vec{x}=\vec{b}$ has a solution (as we just stated in Theorem 17). In the next section we'll demonstrate many different properties of invertible matrices, including stating several different ways in which we know that a matrix is invertible.

Exercises 5.3

Problems

In Exercises 1 – 8, A matrix A is given. Find A^{-1} using Theorem 16, if it exists.

1. $\begin{bmatrix} 1 & 5 \\ -5 & -24 \end{bmatrix}$

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

3. $\begin{bmatrix} 3 & 0 \\ 0 & 7 \end{bmatrix}$

4. $\begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix}$

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

6. $\begin{bmatrix} 3 & 7 \\ 2 & 4 \end{bmatrix}$

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

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

In Exercises 9 – 28, a matrix A is given. Find A^{-1} using Key Idea 21, if it exists.

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

10. $\begin{bmatrix} -5 & -2 \\ 9 & 2 \end{bmatrix}$

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

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

13. $\begin{bmatrix} 25 & -10 & -4 \\ -18 & 7 & 3 \\ -6 & 2 & 1 \end{bmatrix}$

14. $\begin{bmatrix} 2 & 3 & 4 \\ -3 & 6 & 9 \\ -1 & 9 & 13 \end{bmatrix}$

15. $\begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & -7 \\ 20 & 7 & -48 \end{bmatrix}$

16. $\begin{bmatrix} -4 & 1 & 5 \\ -5 & 1 & 9 \\ -10 & 2 & 19 \end{bmatrix}$

17. $\begin{bmatrix} 5 & -1 & 0 \\ 7 & 7 & 1 \\ -2 & -8 & -1 \end{bmatrix}$

18. $\begin{bmatrix} 1 & -5 & 0 \\ -2 & 15 & 4 \\ 4 & -19 & 1 \end{bmatrix}$

19. $\begin{bmatrix} 25 & -8 & 0 \\ -78 & 25 & 0 \\ 48 & -15 & 1 \end{bmatrix}$

20. $\begin{bmatrix} 1 & 0 & 0 \\ 7 & 5 & 8 \\ -2 & -2 & -3 \end{bmatrix}$

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

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

23. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ -19 & -9 & 0 & 4 \\ 33 & 4 & 1 & -7 \\ 4 & 2 & 0 & -1 \end{bmatrix}$

24. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 27 & 1 & 0 & 4 \\ 18 & 0 & 1 & 4 \\ 4 & 0 & 0 & 1 \end{bmatrix}$

25. $\begin{bmatrix} -15 & 45 & -3 & 4 \\ 55 & -164 & 15 & -15 \\ -215 & 640 & -62 & 59 \\ -4 & 12 & 0 & 1 \end{bmatrix}$

26. $\begin{bmatrix} 1 & 0 & 2 & 8 \\ 0 & 1 & 0 & 0 \\ 0 & -4 & -29 & -110 \\ 0 & -3 & -5 & -19 \end{bmatrix}$

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

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

In Exercises 29 – 36, a matrix A and a vector $\vec{b}$ are given. Solve the equation $A\vec{x} = \vec{b}$ using Theorem 17.

29. $A = \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 21 \\ 13 \end{bmatrix}$

30. $A = \begin{bmatrix} 1 & -4 \\ 4 & -15 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 21 \\ 77 \end{bmatrix}$

31. $A = \begin{bmatrix} 9 & 70 \\ -4 & -31 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

32. $A = \begin{bmatrix} 10 & -57 \\ 3 & -17 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -14 \\ -4 \end{bmatrix}$

33. $A = \begin{bmatrix} 1 & 2 & 12 \\ 0 & 1 & 6 \\ -3 & 0 & 1 \end{bmatrix},$

$\vec{b} = \begin{bmatrix} -17 \\ -5 \\ 20 \end{bmatrix}$

34. $A = \begin{bmatrix} 1 & 0 & -3 \\ 8 & -2 & -13 \\ 12 & -3 & -20 \end{bmatrix},$

$\vec{b} = \begin{bmatrix} -34 \\ -159 \\ -243 \end{bmatrix}$

35. $A = \begin{bmatrix} 5 & 0 & -2 \\ -8 & 1 & 5 \\ -2 & 0 & 1 \end{bmatrix},$

$\vec{b} = \begin{bmatrix} 33 \\ -70 \\ -15 \end{bmatrix}$

36. $A = \begin{bmatrix} 1 & -6 & 0 \\ 0 & 1 & 0 \\ 2 & -8 & 1 \end{bmatrix},$

$\vec{b} = \begin{bmatrix} -69 \\ 10 \\ -102 \end{bmatrix}$

5.4 Properties of the Matrix Inverse

AS YOU READ . . .

1. What does it mean to say that two statements are "equivalent?"
2. T/F: If A is not invertible, then $A\vec{x} = \vec{0}$ could have no solutions.
3. T/F: If A is not invertible, then $A\vec{x} = \vec{b}$ could have infinite solutions.
4. What is the inverse of the inverse of A?
5. T/F: Solving $A\vec{x} = \vec{b}$ using Gaussian elimination is faster than using the inverse of A.

We ended the previous section by stating that invertible matrices are important. Since they are, in this section we study invertible matrices in two ways. First, we look at ways to tell whether or not a matrix is invertible, and second, we study properties of invertible matrices (that is, how they interact with other matrix operations).

We start with collecting ways in which we know that a matrix is invertible. We actually already know the truth of this theorem from our work in the previous section, but it is good to list the following statements in one place. As we move through other sections, we'll add on to this theorem.

Theorem 18 Invertible Matrix Theorem

Let A be an $n \times n$ matrix. The following statements are equivalent.

(a) A is invertible.

(b) There exists a matrix B such that $BA = I$.

(c) There exists a matrix C such that $AC = I$.

(d) The reduced row echelon form of A is I.

(e) The equation $A\vec{x} = \vec{b}$ has exactly one solution for every $n \times 1$ vector $\vec{b}$.

(f) The equation $A\vec{x} = \vec{0}$ has exactly one solution (namely, $\vec{x} = \vec{0}$).

Note: Theorem 18 gives us several different ways of saying what is essentially the same thing (logically speaking). Theorems like this are very useful, since it's often the case that, in a given situation, one of the conditions can easily be checked, allowing us to immediately obtain information that might be difficult (or impossible) to verify directly.

Let's make note of a few things about the Invertible Matrix Theorem.

1. First, note that the theorem uses the phrase "the following statements are *equivalent*." When two or more statements are equivalent, it means that the truth of any one of them implies that the rest are also true; if any one of the statements is false, then they are all false. So, for example, if we determined that the equation $A\vec{x} = \vec{0}$ had exactly one solution (and A was an $n \times n$ matrix) then we would know that A was invertible, that $A\vec{x} = \vec{b}$ had only one solution, that the reduced row echelon form of A was I, etc.

2. Let's go through each of the statements and see why we already knew they all said essentially the same thing.

(a) This simply states that A is invertible – that is, that there exists a matrix A^{-1} such that $A^{-1}A = AA^{-1} = I$. We'll go on to show why all the other statements basically tell us "A is invertible."

(b) If we know that A is invertible, then we already know that there is a matrix B where $BA = I$. That is part of the definition of invertible. However, we can also "go the other way." Recall from Theorem 14 that even if all we know is that there is a matrix B where $BA = I$, then we also know that $AB = I$. That is, we know that B is the inverse of A (and hence A is invertible).

(c) We use the same logic as in the previous statement to show why this is the same as "A is invertible."

(d) If A is invertible, we can find the inverse by using Key Idea 21 (which in turn depends on Theorem 14). The crux of Key Idea 21 is that the reduced row echelon form of A is I; if it is something else, we can't find A^{-1} (it doesn't exist). Knowing that A is invertible means that the reduced row echelon form of A is I. We can go the other way; if we know that the reduced row echelon form of A is I, then we can employ Key Idea 21 to find A^{-1}, so A is invertible.

(e) We know from Theorem 17 that if A is invertible, then given any vector $\vec{b}$, $A\vec{x} = \vec{b}$ has always has exactly one solution, namely $\vec{x} = A^{-1}\vec{b}$. However, we can go the other way; let's say we know that $A\vec{x} = \vec{b}$ always has exactly solution. How can we conclude that A is invertible?

Think about how we, up to this point, determined the solution to $A\vec{x} = \vec{b}$. We set up the augmented matrix $\begin{bmatrix} A & \vec{b} \end{bmatrix}$ and put it into reduced row echelon form. We know that getting the identity matrix on the left means that we had a unique solution (and not getting the identity means we either have no solution or infinite solutions). So getting I on the left means having a unique solution; having I on the left means that the reduced row echelon form of A is I, which we know from above is the same as A being invertible.

(f) This is the same as the above; simply replace the vector $\vec{b}$ with the vector $\vec{0}$.

So we came up with a list of statements that are all *equivalent* to the statement "A is invertible." Again, if we know that if any one of them is true (or false), then they are all true (or all false).

Theorem 18 states formally that if A is invertible, then $A\vec{x} = \vec{b}$ has exactly one solution, namely $A^{-1}\vec{b}$. What if A is not invertible? What are the possibilities for solutions to $A\vec{x} = \vec{b}$?

We know that $A\vec{x} = \vec{b}$ *cannot* have exactly one solution; if it did, then by our theorem it would be invertible. Recalling that linear equations have either one solution, infinite solutions, or no solution, we are left with the latter options when A is not invertible. This idea is important and so we'll state it again as a Key Idea.

Key Idea 22 Solutions to $A\vec{x} = \vec{b}$ and the Invertibility of A

Consider the system of linear equations $A\vec{x} = \vec{b}$.

1. If A is invertible, then $A\vec{x} = \vec{b}$ has exactly one solution, namely $A^{-1}\vec{b}$.
2. If A is not invertible, then $A\vec{x} = \vec{b}$ has either infinite solutions or no solution.

In Theorem 18 we've come up with a list of ways in which we can tell whether or not a matrix is invertible. At the same time, we have come up with a list of properties of invertible matrices – things we know that are true about them. (For instance, if we know that A is invertible, then we know that $A\vec{x} = \vec{b}$ has only one solution.)

We now go on to discover other properties of invertible matrices. Specifically, we want to find out how invertibility interacts with other matrix operations. For instance, if we know that A and B are invertible, what is the inverse of $A + B$? What is the inverse of AB? What is "the inverse of the inverse?" We'll explore these questions through an example.

Example 116 Exploring properties of the inverse

Let

$$A = \begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} -2 & 0 \\ 1 & 1 \end{bmatrix}.$$

Find:

1. A^{-1}
2. B^{-1}
3. $(AB)^{-1}$
4. $(A^{-1})^{-1}$
5. $(A + B)^{-1}$
6. $(5A)^{-1}$

In addition, try to find connections between each of the above.

Solution

1. Computing A^{-1} is straightforward; we'll use Theorem 16.

$$A^{-1} = \frac{1}{3}\begin{bmatrix} 1 & -2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1/3 & -2/3 \\ 0 & 1 \end{bmatrix}$$

2. We compute B^{-1} in the same way as above.

$$B^{-1} = \frac{1}{-2}\begin{bmatrix} 1 & 0 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} -1/2 & 0 \\ 1/2 & 1 \end{bmatrix}$$

3. To compute $(AB)^{-1}$, we first compute AB:

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

We now apply Theorem 16 to find $(AB)^{-1}$.

$$(AB)^{-1} = \frac{1}{-6}\begin{bmatrix} 1 & -2 \\ -1 & -4 \end{bmatrix} = \begin{bmatrix} -1/6 & 1/3 \\ 1/6 & 2/3 \end{bmatrix}$$

4. To compute $(A^{-1})^{-1}$, we simply apply Theorem 16 to A^{-1}:

$$(A^{-1})^{-1} = \frac{1}{1/3}\begin{bmatrix} 1 & 2/3 \\ 0 & 1/3 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix}.$$

5. To compute $(A+B)^{-1}$, we first compute $A+B$ then apply Theorem 16:

$$A + B = \begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -2 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}.$$

Hence

$$(A+B)^{-1} = \frac{1}{0}\begin{bmatrix} 2 & -2 \\ -1 & 1 \end{bmatrix} = !$$

Our last expression is really nonsense; we know that if $ad - bc = 0$, then the given matrix is not invertible. That is the case with $A + B$, so we conclude that $A + B$ is not invertible.

6. To compute $(5A)^{-1}$, we compute $5A$ and then apply Theorem 16.

$$(5A)^{-1} = \left(\begin{bmatrix} 15 & 10 \\ 0 & 5 \end{bmatrix}\right)^{-1} = \frac{1}{75}\begin{bmatrix} 5 & -10 \\ 0 & 15 \end{bmatrix} = \begin{bmatrix} 1/15 & -2/15 \\ 0 & 1/5 \end{bmatrix}$$

We now look for connections between A^{-1}, B^{-1}, $(AB)^{-1}$, $(A^{-1})^{-1}$ and $(A+B)^{-1}$.

3. Is there some sort of relationship between $(AB)^{-1}$ and A^{-1} and B^{-1}? A first guess that seems plausible is $(AB)^{-1} = A^{-1}B^{-1}$. Is this true? Using our work from above, we have

$$A^{-1}B^{-1} = \begin{bmatrix} 1/3 & -2/3 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -1/2 & 0 \\ 1/2 & 1 \end{bmatrix} = \begin{bmatrix} -1/2 & -2/3 \\ 1/2 & 1 \end{bmatrix}.$$

Obviously, this is not equal to $(AB)^{-1}$. Before we do some further guessing, let's think about what the inverse of AB is supposed to do. The inverse – let's call it C – is supposed to be a matrix such that

$$(AB)C = C(AB) = I.$$

In examining the expression $(AB)C$, we see that we want B to somehow "cancel" with C. What "cancels" B? An obvious answer is B^{-1}. This gives us a thought: perhaps we got the order of A^{-1} and B^{-1} wrong before. After all, we were hoping to find that

$$ABA^{-1}B^{-1} \stackrel{?}{=} I,$$

but algebraically speaking, it is hard to cancel out these terms.(Recall that matrix multiplication is not commutative: $AB \neq BA$ in general.) However, switching the order of A^{-1} and B^{-1} gives us some hope. Is $(AB)^{-1} = B^{-1}A^{-1}$? Let's see.

$$\begin{aligned}(AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} && \text{(regrouping by the associative property)} \\ &= AIA^{-1} && (BB^{-1} = I) \\ &= AA^{-1} && (AI = A) \\ &= I && (AA^{-1} = I)\end{aligned}$$

Thus it seems that $(AB)^{-1} = B^{-1}A^{-1}$. Let's confirm this with our example matrices.

$$B^{-1}A^{-1} = \begin{bmatrix} -1/2 & 0 \\ 1/2 & 1 \end{bmatrix}\begin{bmatrix} 1/3 & -2/3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1/6 & 1/3 \\ 1/6 & 2/3 \end{bmatrix} = (AB)^{-1}.$$

It worked!

4. Is there some sort of connection between $(A^{-1})^{-1}$ and A? The answer is pretty obvious: they are equal. The "inverse of the inverse" returns one to the original matrix.

5. Is there some sort of relationship between $(A+B)^{-1}$, A^{-1} and B^{-1}? Certainly, if we were forced to make a guess without working any examples, we would guess that

$$(A+B)^{-1} \stackrel{?}{=} A^{-1} + B^{-1}.$$

However, we saw that in our example, the matrix $(A+B)$ isn't even invertible. This pretty much kills any hope of a connection.

6. Is there a connection between $(5A)^{-1}$ and A^{-1}? Consider:

$$\begin{aligned}(5A)^{-1} &= \begin{bmatrix} 1/15 & -2/15 \\ 0 & 1/5 \end{bmatrix} \\ &= \frac{1}{5}\begin{bmatrix} 1/3 & -2/3 \\ 0 & 1/5 \end{bmatrix} \\ &= \frac{1}{5}A^{-1}\end{aligned}$$

Yes, there is a connection!

The fact that invertibility works well with matrix multiplication should not come as a surprise. After all, saying that A is invertible makes a statement about the muliplicative properties of A. It says that I can multiply A with a special matrix to get I. Invertibility, in and of itself, says nothing about matrix addition, therefore we should not be too surprised that it doesn't work well with it.

Let's summarize the results of this example. If A and B are both invertible matrices, then so is their product, AB. We demonstrated this with our example, and there is more to be said. Let's suppose that A and B are $n \times n$ matrices, but we don't yet know if they are invertible. If AB is invertible, then each of A and B are; if AB is not invertible, then A or B is also not invertible.

In short, invertibility "works well" with matrix multiplication. However, we saw that it doesn't work well with matrix addition. Knowing that A and B are invertible does not help us find the inverse of $(A+B)$; in fact, the latter matrix may not even be invertible.

Let's do one more example, then we'll summarize the results of this section in a theorem.

Example 117 **Computing the inverse of a diagonal matrix**

Find the inverse of $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -7 \end{bmatrix}$.

Solution We'll find A^{-1} using Key Idea 21.

$$\begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 & 1 & 0 \\ 0 & 0 & -7 & 0 & 0 & 1 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1/7 \end{bmatrix}$$

Therefore

$$A^{-1} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & -1/7 \end{bmatrix}.$$

The matrix A in the previous example is a *diagonal* matrix: the only nonzero entries of A lie on the *diagonal*. The relationship between A and A^{-1} in the above example seems pretty strong, and it holds true in general. We'll state this and summarize the results of this section with the following theorem.

Theorem 19 Properties of Invertible Matrices

Let A and B be $n \times n$ invertible matrices. Then:

1. AB is invertible; $(AB)^{-1} = B^{-1}A^{-1}$.
2. A^{-1} is invertible; $(A^{-1})^{-1} = A$.
3. nA is invertible for any nonzero scalar n; $(nA)^{-1} = \frac{1}{n}A^{-1}$.
4. If A is a diagonal matrix, with diagonal entries $d_1, d_2, \cdots, d_n$, where none of the diagonal entries are 0, then A^{-1} exists and is a diagonal matrix. Furthermore, the diagonal entries of A^{-1} are $1/d_1, 1/d_2, \cdots, 1/d_n$.

Furthermore,

1. If a product AB is not invertible, then A or B is not invertible.
2. If A or B are not invertible, then AB is not invertible.

We still haven't formally defined *diagonal*, but the definition is rather visual so we risk it. See Definition 45 on page 218 for more details.

Yes, real people do solve linear equations in real life. Not just mathematicians, but economists, engineers, and scientists of all flavours regularly need to solve linear equations, and the matrices they use are often *huge*.

Most people see matrices at work without thinking about it. Digital pictures are simply "rectangular arrays" of numbers representing colours – they are matrices of colours. Many of the standard image processing operations involve matrix operations. The author's wife has a "7 megapixel" camera which creates pictures that are 3072×2304 in size, giving over 7 million pixels, and that isn't even considered a "large" picture these days.

We end this section with a comment about solving systems of equations "in real life." Solving a system $A\vec{x} = \vec{b}$ by computing $A^{-1}\vec{b}$ seems pretty slick, so it would make sense that this is the way it is normally done. However, in practice, this is rarely done. There are two main reasons why this is the case.

First, computing A^{-1} and $A^{-1}\vec{b}$ is "expensive" in the sense that it takes up a lot of computing time. Certainly, our calculators have no trouble dealing with the 3×3 cases we often consider in this textbook, but in real life the matrices being considered are very large (as in, hundreds of thousand rows and columns). Computing A^{-1} alone is rather impractical, and we waste a lot of time if we come to find out that A^{-1} does not exist. Even if we already know what A^{-1} is, computing $A^{-1}\vec{b}$ is computationally expensive – Gaussian elimination is faster.

Secondly, computing A^{-1} using the method we've described often gives rise to numerical roundoff errors. Even though computers often do computations with an accuracy to more than 8 decimal places, after thousands of computations, rounding off can cause big errors. (A "small" $1,000 \times 1,000$ matrix has $1,000,000$ entries! That's a lot of places to have roundoff errors accumulate!) It is not unheard of to have a computer compute A^{-1} for a large matrix, and then immediately have it compute AA^{-1} and *not* get the identity matrix. (The result is usually very close, with the numbers on the diagonal close to 1 and the other entries near 0. But it isn't exactly the identity matrix.)

Therefore, in real life, solutions to $A\vec{x} = \vec{b}$ are usually found using the methods we learned in Section 5.1. It turns out that even with all of our advances in mathematics, it is hard to beat the basic method that Gauss introduced a long time ago.

Exercises 5.4

Problems

In Exercises 1 – 4, matrices A and B are given. Compute $(AB)^{-1}$ and $B^{-1}A^{-1}$.

1. $A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 5 \\ 2 & 5 \end{bmatrix}$

2. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 7 & 1 \\ 2 & 1 \end{bmatrix}$

3. $A = \begin{bmatrix} 2 & 5 \\ 3 & 8 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \\ 1 & 4 \end{bmatrix}$

4. $A = \begin{bmatrix} 2 & 4 \\ 2 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 2 \\ 6 & 5 \end{bmatrix}$

In Exercises 5 – 8, a 2×2 matrix A is given. Compute A^{-1} and $(A^{-1})^{-1}$ using Theorem 16.

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

6. $A = \begin{bmatrix} 3 & 5 \\ 2 & 4 \end{bmatrix}$

7. $A = \begin{bmatrix} 2 & 7 \\ 1 & 3 \end{bmatrix}$

8. $A = \begin{bmatrix} 9 & 0 \\ 7 & 9 \end{bmatrix}$

9. Find 2×2 matrices A and B that are each invertible, but $A + B$ is not.

10. Create a random 6×6 matrix A, then have a calculator or computer compute AA^{-1}. Was the identity matrix returned exactly? Comment on your results.

11. Use a calculator or computer to compute AA^{-1}, where

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \\ 1 & 16 & 81 & 256 \end{bmatrix}.$$

Was the identity matrix returned exactly? Comment on your results.

6: Operations on Matrices

In the previous chapter we learned about matrix arithmetic: adding, subtracting, and multiplying matrices, finding inverses, and multiplying by scalars. In this chapter we learn about some operations that we perform *on* matrices. We can think of them as functions: you input a matrix, and you get something back. One of these operations, the transpose, will return another matrix. With the other operations, the trace and the determinant, we input matrices and get numbers in return, an idea that is different than what we have seen before.

6.1 The Matrix Transpose

AS YOU READ . . .

1. T/F: If A is a 3×5 matrix, then A^T will be a 5×3 matrix.
2. Where are there zeros in an upper triangular matrix?
3. T/F: A matrix is symmetric if it doesn't change when you take its transpose.
4. What is the transpose of the transpose of A?
5. Give 2 other terms to describe symmetric matrices besides "interesting."

We jump right in with a definition.

Definition 44 Transpose

Let A be an $m \times n$ matrix. The *tranpsose of A*, denoted A^T, is the $n \times m$ matrix whose columns are the respective rows of A.

Examples will make this definition clear.

Example 118 Taking the transpose of a matrix

Find the transpose of $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.

Solution Note that A is a 2×3 matrix, so A^T will be a 3×2 matrix. By the definition, the first column of A^T is the first row of A; the second column of A^T is the second row of A. Therefore,

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

Example 119 Computing transposes

Find the transpose of the following matrices.

$$A = \begin{bmatrix} 7 & 2 & 9 & 1 \\ 2 & -1 & 3 & 0 \\ -5 & 3 & 0 & 11 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 10 & -2 \\ 3 & -5 & 7 \\ 4 & 2 & -3 \end{bmatrix} \quad C = \begin{bmatrix} 1 & -1 & 7 & 8 & 3 \end{bmatrix}$$

Solution We find each transpose using the definition without explanation. Make note of the dimensions of the original matrix and the dimensions of its transpose.

$$A^T = \begin{bmatrix} 7 & 2 & -5 \\ 2 & -1 & 3 \\ 9 & 3 & 0 \\ 1 & 0 & 11 \end{bmatrix} \quad B^T = \begin{bmatrix} 1 & 3 & 4 \\ 10 & -5 & 2 \\ -2 & 7 & -3 \end{bmatrix} \quad C^T = \begin{bmatrix} 1 \\ -1 \\ 7 \\ 8 \\ 3 \end{bmatrix}$$

Notice that with matrix B, when we took the transpose, the *diagonal* did not change. We can see what the diagonal is below where we rewrite B and B^T with the diagonal in bold. We'll follow this by a definition of what we mean by "the diagonal of a matrix," along with a few other related definitions.

$$B = \begin{bmatrix} \mathbf{1} & 10 & -2 \\ 3 & \mathbf{-5} & 7 \\ 4 & 2 & \mathbf{-3} \end{bmatrix} \quad B^T = \begin{bmatrix} \mathbf{1} & 3 & 4 \\ 10 & \mathbf{-5} & 2 \\ -2 & 7 & \mathbf{-3} \end{bmatrix}$$

It is probably pretty clear why we call those entries "the diagonal." Here is the formal definition.

Definition 45 The Diagonal, a Diagonal Matrix, Triangular Matrices

Let A be an $m \times n$ matrix. The *diagonal of A* consists of the entries a_{11}, a_{22}, ... of A.

A *diagonal matrix* is an $n \times n$ matrix in which the only nonzero entries lie on the diagonal.

An *upper (lower) triangular* matrix is a matrix in which any nonzero entries lie on or above (below) the diagonal.

Example 120 Classifying matrices

Consider the matrices A, B, C and I_4, as well as their transposes, where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{bmatrix}.$$

Identify the diagonal of each matrix, and state whether each matrix is diagonal, upper triangular, lower triangular, or none of the above.

Solution We first compute the transpose of each matrix.

$$A^T = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{bmatrix} \quad B^T = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad C^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 3 & 5 & 6 & 0 \end{bmatrix}$$

Note that $I_4^T = I_4$.

The diagonals of A and A^T are the same, consisting of the entries 1, 4 and 6. The diagonals of B and B^T are also the same, consisting of the entries 3, 7 and -1. Finally, the diagonals of C and C^T are the same, consisting of the entries 1, 4 and 6.

The matrix A is upper triangular; the only nonzero entries lie on or above the diagonal. Likewise, A^T is lower triangular.

The matrix B is diagonal. By their definitions, we can also see that B is both upper and lower triangular. Likewise, I_4 is diagonal, as well as upper and lower triangular.

Finally, C is upper triangular, with C^T being lower triangular.

Make note of the definitions of diagonal and triangular matrices. We specify that a diagonal matrix must be square, but triangular matrices don't have to be. ("Most" of the time, however, the ones we study are.) Also, as we mentioned before in the example, by definition a diagonal matrix is also both upper and lower triangular. Finally, notice that by definition, the transpose of an upper triangular matrix is a lower triangular matrix, and vice-versa.

There are many questions to probe concerning the transpose operations. The first set of questions we'll investigate involve the matrix arithmetic we learned from last chapter. We do this investigation by way of examples, and then summarize what we have learned at the end.

Remember, this is what mathematicians do. We learn something new, and then we ask lots of questions about it. Often the first questions we ask are along the lines of "How does this new thing relate to the old things I already know about?"

Example 121 **Adding transposed matrices**

Let

$$A=\begin{bmatrix}1&2&3\\4&5&6\end{bmatrix} \text{ and } B=\begin{bmatrix}1&2&1\\3&-1&0\end{bmatrix}.$$

Find A^T+B^T and $(A+B)^T$.

Solution We note that

$$A^T=\begin{bmatrix}1&4\\2&5\\3&6\end{bmatrix} \text{ and } B^T=\begin{bmatrix}1&3\\2&-1\\1&0\end{bmatrix}.$$

Therefore

$$\begin{aligned}A^T+B^T&=\begin{bmatrix}1&4\\2&5\\3&6\end{bmatrix}+\begin{bmatrix}1&3\\2&-1\\1&0\end{bmatrix}\\&=\begin{bmatrix}2&7\\4&4\\4&6\end{bmatrix}.\end{aligned}$$

Also,

$$\begin{aligned}(A+B)^T&=\left(\begin{bmatrix}1&2&3\\4&5&6\end{bmatrix}+\begin{bmatrix}1&2&1\\3&-1&0\end{bmatrix}\right)^T\\&=\left(\begin{bmatrix}2&4&4\\7&4&6\end{bmatrix}\right)^T\\&=\begin{bmatrix}2&7\\4&4\\4&6\end{bmatrix}.\end{aligned}$$

It looks like "the sum of the transposes is the transpose of the sum." (This is kind of fun to say, especially when said fast. Regardless of how fast we say it, we should think about this statement. The "is" represents "equals." The stuff before "is" equals the stuff afterwards.) This should lead us to wonder how the transpose works with multiplication.

Example 122 **Multiplying transposed matrices**
Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Find $(AB)^T$, A^TB^T and B^TA^T.

Solution We first note that

$$A^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \text{ and } B^T = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 1 \end{bmatrix}.$$

Find $(AB)^T$:

$$\begin{aligned} (AB)^T &= \left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \end{bmatrix} \right)^T \\ &= \left(\begin{bmatrix} 3 & 2 & 1 \\ 7 & 6 & 1 \end{bmatrix} \right)^T \\ &= \begin{bmatrix} 3 & 7 \\ 2 & 6 \\ 1 & 1 \end{bmatrix} \end{aligned}$$

Now find A^TB^T:

$$\begin{aligned} A^TB^T &= \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 1 \end{bmatrix} \\ &= \text{Not defined!} \end{aligned}$$

So we can't compute A^TB^T. Let's finish by computing B^TA^T:

$$\begin{aligned} B^TA^T &= \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 7 \\ 2 & 6 \\ 1 & 1 \end{bmatrix} \end{aligned}$$

We may have suspected that $(AB)^T = A^TB^T$. We saw that this wasn't the case, though – and not only was it not equal, the second product wasn't even defined! Oddly enough, though, we saw that $(AB)^T = B^TA^T$. (Then again, maybe this isn't all that "odd." It is reminiscent of the fact that, when invertible, $(AB)^{-1} = B^{-1}A^{-1}$.) To help understand why this is true, look back at the work above and confirm the steps of each multiplication.

We have one more arithmetic operation to look at: the inverse.

Example 123 **Inverting a transposed matrix**
Let

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

Find $(A^{-1})^T$ and $(A^T)^{-1}$.

Solution We first find A^{-1} and A^T:

$$A^{-1} = \begin{bmatrix} 4 & -7 \\ -1 & 2 \end{bmatrix} \text{ and } A^T = \begin{bmatrix} 2 & 1 \\ 7 & 4 \end{bmatrix}.$$

Finding $(A^{-1})^T$:

$$\begin{aligned} (A^{-1})^T &= \begin{bmatrix} 4 & -7 \\ -1 & 2 \end{bmatrix}^T \\ &= \begin{bmatrix} 4 & -1 \\ -7 & 2 \end{bmatrix} \end{aligned}$$

Finding $(A^T)^{-1}$:

$$\begin{aligned} (A^T)^{-1} &= \begin{bmatrix} 2 & 1 \\ 7 & 4 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 4 & -1 \\ -7 & 2 \end{bmatrix} \end{aligned}$$

It seems that "the inverse of the transpose is the transpose of the inverse." (Again, we should think about this statement. The part before "is" states that we take the transpose of a matrix, then find the inverse. The part after "is" states that we find the inverse of the matrix, then take the transpose. Since these two statements are linked by an "is," they are equal.)

We have just looked at some examples of how the transpose operation interacts with matrix arithmetic operations. (These examples don't *prove* anything, other than it worked in specific examples.) We now give a theorem that tells us that what we saw wasn't a coincidence, but rather is always true.

Theorem 20 Properties of the Matrix Transpose

Let A and B be matrices where the following operations are defined. Then:

1. $(A+B)^T = A^T + B^T$ and $(A-B)^T = A^T - B^T$
2. $(kA)^T = kA^T$
3. $(AB)^T = B^TA^T$
4. $(A^{-1})^T = (A^T)^{-1}$
5. $(A^T)^T = A$

We included in the theorem two ideas we didn't discuss already. First, that $(kA)^T = kA^T$. This is probably obvious. It doesn't matter when you multiply a matrix by a scalar when dealing with transposes.

The second "new" item is that $(A^T)^T = A$. That is, if we take the transpose of a matrix, then take its transpose again, what do we have? The original matrix.

Now that we know some properties of the transpose operation, we are tempted to play around with it and see what happens. For instance, if A is an $m\times n$ matrix, we know that A^T is an $n \times m$ matrix. So no matter what matrix A we start with, we can always perform the multiplication AA^T (and also A^TA) and the result is a square matrix!

Another thing to ask ourselves as we "play around" with the transpose: suppose A is a square matrix. Is there anything special about $A + A^T$? The following example has us try out these ideas.

Example 124 **The matrices AA^T, $A + A^T$, and $A - A^T$**

Let

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

Find AA^T, $A + A^T$ and $A - A^T$.

Solution Finding AA^T:

$$\begin{aligned} AA^T &= \begin{bmatrix} 2 & 1 & 3 \\ 2 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \\ 1 & -1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 14 & 6 & 5 \\ 6 & 4 & 3 \\ 5 & 3 & 2 \end{bmatrix} \end{aligned}$$

Finding $A + A^T$:

$$\begin{aligned} A + A^T &= \begin{bmatrix} 2 & 1 & 3 \\ 2 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 2 & 1 \\ 1 & -1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 3 & 4 \\ 3 & -2 & 1 \\ 4 & 1 & 2 \end{bmatrix} \end{aligned}$$

Finding $A - A^T$:

$$\begin{aligned} A - A^T &= \begin{bmatrix} 2 & 1 & 3 \\ 2 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 2 & 1 \\ 1 & -1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -1 & 2 \\ 1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} \end{aligned}$$

Let's look at the matrices we've formed in this example. First, consider AA^T. Something seems to be nice about this matrix – look at the location of the 6's, the 5's and the 3's. More precisely, let's look at the transpose of AA^T. We should notice that if we take the transpose of this matrix, we have the very same matrix. That is,

$$\left(\begin{bmatrix} 14 & 6 & 5 \\ 6 & 4 & 3 \\ 5 & 3 & 2 \end{bmatrix} \right)^T = \begin{bmatrix} 14 & 6 & 5 \\ 6 & 4 & 3 \\ 5 & 3 & 2 \end{bmatrix}!$$

We'll formally define this in a moment, but a matrix that is equal to its transpose is called *symmetric*.

Look at the next part of the example; what do we notice about $A + A^T$? We should see that it, too, is symmetric. Finally, consider the last part of the example: do we notice anything about $A - A^T$?

We should immediately notice that it is not symmetric, although it does seem "close." Instead of it being equal to its transpose, we notice that this matrix is the *opposite* of its transpose. We call this type of matrix *skew symmetric.* (Some mathematicians use the term *antisymmetric*) We formally define these matrices here.

Definition 46 Symmetric and Skew Symmetric Matrices

A matrix A is *symmetric* if $A^T = A$.

A matrix A is *skew symmetric* if $A^T = -A$.

Note that in order for a matrix to be either symmetric or skew symmetric, it must be square.

So why was AA^T symmetric in our previous example? Did we just luck out? (Of course not.) Let's take the transpose of AA^T and see what happens.

$$\begin{aligned} (AA^T)^T &= (A^T)^T(A)^T && \text{transpose multiplication rule} \\ &= AA^T && (A^T)^T = A \end{aligned}$$

We have just *proved* that no matter what matrix A we start with, the matrix AA^T will be symmetric. Nothing in our string of equalities even demanded that A be a square matrix; it is always true.

We can do a similar proof to show that as long as A is square, $A + A^T$ is a symmetric matrix. (Why do we say that A has to be square?) We'll instead show here that if A is a square matrix, then $A - A^T$ is skew symmetric.

$$\begin{aligned} (A - A^T)^T &= A^T - (A^T)^T && \text{transpose subtraction rule} \\ &= A^T - A \\ &= -(A - A^T) \end{aligned}$$

So we took the transpose of $A - A^T$ and we got $-(A - A^T)$; this is the definition of being skew symmetric.

We'll take what we learned from Example 123 and put it in a box. (We've already proved most of this is true; the rest we leave to solve in the Exercises.)

Theorem 21 Symmetric and Skew Symmetric Matrices

1. Given any matrix A, the matrices AA^T and A^TA are symmetric.
2. Let A be a square matrix. The matrix $A + A^T$ is symmetric.
3. Let A be a square matrix. The matrix $A - A^T$ is skew symmetric.

Why do we care about the transpose of a matrix? Why do we care about symmetric matrices?

There are two answers that each answer both of these questions. First, we are interested in the transpose of a matrix and symmetric matrices because they are interesting. One particularly interesting thing about symmetric and skew symmetric matrices is this: consider the sum of $(A + A^T)$ and $(A - A^T)$:

$$(A + A^T) + (A - A^T) = 2A.$$

This gives us an idea: if we were to multiply both sides of this equation by $\frac{1}{2}$, then the right hand side would just be A. This means that

$$A = \underbrace{\frac{1}{2}(A + A^T)}_{\text{symmetric}} + \underbrace{\frac{1}{2}(A - A^T)}_{\text{skew symmetric}}.$$

That is, any matrix A can be written as the sum of a symmetric and skew symmetric matrix. That's interesting.

The second reason we care about them is that they are very useful and important in various areas of mathematics. The transpose of a matrix turns out to be an important operation; symmetric matrices have many nice properties that make solving certain types of problems possible.

Most of this text focuses on the preliminaries of matrix algebra, and the actual uses are beyond our current scope. One easy to describe example is curve fitting. Suppose we are given a large set of data points that, when plotted, look roughly quadratic. How do we find the quadratic that "best fits" this data? The solution can be found using matrix algebra, and specifically a matrix called the *pseudoinverse*. If A is a matrix, the pseudoinverse of A is the matrix $A^\dagger = (A^TA)^{-1}A^T$ (assuming that the inverse exists). We aren't going to worry about what all the above means; just notice that it has a cool sounding name and the transpose appears twice.

In the next section we'll learn about the trace, another operation that can be performed on a matrix that is relatively simple to compute but can lead to some deep results.

Exercises 6.1

Problems

In Exercises 1 – 24, a matrix A is given. Find A^T; make note if A is upper/lower triangular, diagonal, symmetric and/or skew symmetric.

1. $\begin{bmatrix} -7 & 4 \\ 4 & -6 \end{bmatrix}$

2. $\begin{bmatrix} 3 & 1 \\ -7 & 8 \end{bmatrix}$

3. $\begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix}$

4. $\begin{bmatrix} 13 & -3 \\ -3 & 1 \end{bmatrix}$

5. $\begin{bmatrix} -5 & -9 \\ 3 & 1 \\ -10 & -8 \end{bmatrix}$

6. $\begin{bmatrix} -2 & 10 \\ 1 & -7 \\ 9 & -2 \end{bmatrix}$

7. $\begin{bmatrix} 4 & -7 & -4 & -9 \\ -9 & 6 & 3 & -9 \end{bmatrix}$

8. $\begin{bmatrix} 3 & -10 & 0 & 6 \\ -10 & -2 & -3 & 1 \end{bmatrix}$

9. $\begin{bmatrix} -7 & -8 & 2 & -3 \end{bmatrix}$

10. $\begin{bmatrix} -9 & 8 & 2 & -7 \end{bmatrix}$

11. $\begin{bmatrix} -9 & 4 & 10 \\ 6 & -3 & -7 \\ -8 & 1 & -1 \end{bmatrix}$

12. $\begin{bmatrix} 4 & -5 & 2 \\ 1 & 5 & 9 \\ 9 & 2 & 3 \end{bmatrix}$

13. $\begin{bmatrix} 4 & 0 & -2 \\ 0 & 2 & 3 \\ -2 & 3 & 6 \end{bmatrix}$

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

15. $\begin{bmatrix} 2 & -5 & -3 \\ 5 & 5 & -6 \\ 7 & -4 & -10 \end{bmatrix}$

16. $\begin{bmatrix} 0 & -6 & 1 \\ 6 & 0 & 4 \\ -1 & -4 & 0 \end{bmatrix}$

17. $\begin{bmatrix} 4 & 2 & -9 \\ 5 & -4 & -10 \\ -6 & 6 & 9 \end{bmatrix}$

18. $\begin{bmatrix} 4 & 0 & 0 \\ -2 & -7 & 0 \\ 4 & -2 & 5 \end{bmatrix}$

19. $\begin{bmatrix} -3 & -4 & -5 \\ 0 & -3 & 5 \\ 0 & 0 & -3 \end{bmatrix}$

20. $\begin{bmatrix} 6 & -7 & 2 & 6 \\ 0 & -8 & -1 & 0 \\ 0 & 0 & 1 & -7 \end{bmatrix}$

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

22. $\begin{bmatrix} 6 & -4 & -5 \\ -4 & 0 & 2 \\ -5 & 2 & -2 \end{bmatrix}$

23. $\begin{bmatrix} 0 & 1 & -2 \\ -1 & 0 & 4 \\ 2 & -4 & 0 \end{bmatrix}$

24. $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

6.2 The Matrix Trace

AS YOU READ . . .

1. T/F: We only compute the trace of square matrices.
2. T/F: One can tell if a matrix is invertible by computing the trace.

In the previous section, we learned about an operation we can peform on matrices, namely the transpose. Given a matrix A, we can "find the transpose of A," which is another matrix. In this section we learn about a new operation called the *trace*. It is a different type of operation than the transpose. Given a matrix A, we can "find the trace of A," which is not a matrix but rather a number. We formally define it here.

Definition 47 The Trace

Let A be an $n \times n$ matrix. The *trace of* A, denoted $\text{tr}(A)$, is the sum of the diagonal elements of A. That is,

$$\text{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}.$$

This seems like a simple definition, and it really is. Just to make sure it is clear, let's practice.

Example 125 Computing the trace of a matrix
Find the trace of A, B, C and I_4, where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \ B = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 8 & 1 \\ -2 & 7 & -5 \end{bmatrix} \text{ and } C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}.$$

Solution To find the trace of A, note that the diagonal elements of A are 1 and 4. Therefore, $\text{tr}(A) = 1 + 4 = 5$.

We see that the diagonal elements of B are 1, 8 and -5, so $\text{tr}(B) = 1+8-5 = 4$.

The matrix C is not a square matrix, and our definition states that we must start with a square matrix. Therefore $\text{tr}(C)$ is not defined.

Finally, the diagonal of I_4 consists of four 1s. Therefore $\text{tr}(I_4) = 4$.

Now that we have defined the trace of a matrix, we should think like mathematicians and ask some questions. The first questions that should pop into our minds should be along the lines of "How does the trace work with other matrix operations?" (Recall that we asked a similar question once we learned about the transpose.) We should think about how the trace works with matrix addition, scalar multiplication, matrix multiplication, matrix inverses, and the transpose.

We'll give a theorem that will formally tell us what is true in a moment, but first let's play with two sample matrices and see if we can see what will happen. Let

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 3 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & 2 & -1 \end{bmatrix}.$$

It should be clear that $\text{tr}(A) = 5$ and $\text{tr}(B) = 3$. What is $\text{tr}(A + B)$?

$$\begin{aligned}\text{tr}(A+B) &= \text{tr}\left(\begin{bmatrix} 2 & 1 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 3 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & 2 & -1 \end{bmatrix}\right) \\ &= \text{tr}\left(\begin{bmatrix} 4 & 1 & 4 \\ 1 & 2 & -1 \\ 3 & 1 & 2 \end{bmatrix}\right) \\ &= 8\end{aligned}$$

Something to think about: we know that not all square matrices are invertible. Would we be able to tell just by the trace? That seems unlikely.

So we notice that $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$. This probably isn't a coincidence.

How does the trace work with scalar multiplication? If we multiply A by 4, then the diagonal elements will be 8, 0 and 12, so $\text{tr}(4A) = 20$. Is it a coincidence that this is 4 times the trace of A?

Let's move on to matrix multiplication. How will the trace of AB relate to the traces of A and B? Let's see:

$$\begin{aligned}\text{tr}(AB) &= \text{tr}\left(\begin{bmatrix} 2 & 1 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 3 \end{bmatrix}\begin{bmatrix} 2 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & 2 & -1 \end{bmatrix}\right) \\ &= \text{tr}\left(\begin{bmatrix} 3 & 8 & -1 \\ 4 & -2 & 3 \\ 7 & 4 & 0 \end{bmatrix}\right) \\ &= 1\end{aligned}$$

It isn't exactly clear what the relationship is among $\text{tr}(A)$, $\text{tr}(B)$ and $\text{tr}(AB)$. Before moving on, let's find $\text{tr}(BA)$:

$$\begin{aligned}\text{tr}(BA) &= \text{tr}\left(\begin{bmatrix} 2 & 0 & 1 \\ -1 & 2 & 0 \\ 0 & 2 & -1 \end{bmatrix}\begin{bmatrix} 2 & 1 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 3 \end{bmatrix}\right) \\ &= \text{tr}\left(\begin{bmatrix} 7 & 1 & 9 \\ 2 & -1 & -5 \\ 1 & 1 & -5 \end{bmatrix}\right) \\ &= 1\end{aligned}$$

We notice that $\text{tr}(AB) = \text{tr}(BA)$. Is this coincidental?

How are the traces of A and A^{-1} related? We compute A^{-1} and find that

$$A^{-1} = \begin{bmatrix} 1/17 & 6/17 & 1/17 \\ 9/17 & 3/17 & -8/17 \\ 2/17 & -5/17 & 2/17 \end{bmatrix}.$$

Therefore $\text{tr}(A^{-1}) = 6/17$. Again, the relationship isn't clear.

Finally, let's see how the trace is related to the transpose. We actually don't have to formally compute anything. Recall from the previous section that the diagonals of A and A^T are identical; therefore, $\text{tr}(A) = \text{tr}(A^T)$. That, we know for sure, isn't a coincidence.

We now formally state what equalities are true when considering the interaction of the trace with other matrix operations.

> **Theorem 22 Properties of the Matrix Trace**
>
> Let A and B be $n \times n$ matrices. Then:
>
> 1. $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$
> 2. $\text{tr}(A - B) = \text{tr}(A) - \text{tr}(B)$
> 3. $\text{tr}(kA) = k \cdot \text{tr}(A)$
> 4. $\text{tr}(AB) = \text{tr}(BA)$
> 5. $\text{tr}(A^T) = \text{tr}(A)$

One of the key things to note here is what this theorem does *not* say. It says nothing about how the trace relates to inverses. The reason for the silence in these areas is that there simply is not a relationship.

We end this section by again wondering why anyone would care about the trace of matrix. One reason mathematicians are interested in it is that it can give a measurement of the "size" of a matrix.

Consider the following 2×2 matrices:

$$A = \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 6 & 7 \\ 11 & -4 \end{bmatrix}.$$

These matrices have the same trace, yet B clearly has bigger elements in it. So how can we use the trace to determine a "size" of these matrices? We can consider $\text{tr}(A^TA)$ and $\text{tr}(B^TB)$.

$$\begin{aligned} \text{tr}(A^TA) &= \text{tr}\left(\begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix}\right) \\ &= \text{tr}\left(\begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix}\right) \\ &= 7 \end{aligned}$$

$$\begin{aligned} \text{tr}(B^TB) &= \text{tr}\left(\begin{bmatrix} 6 & 11 \\ 7 & -4 \end{bmatrix}\begin{bmatrix} 6 & 7 \\ 11 & -4 \end{bmatrix}\right) \\ &= \text{tr}\left(\begin{bmatrix} 157 & -2 \\ -2 & 65 \end{bmatrix}\right) \\ &= 222 \end{aligned}$$

Our concern is not how to interpret what this "size" measurement means, but rather to demonstrate that the trace (along with the transpose) can be used to give (perhaps useful) information about a matrix.

This example brings to light many interesting ideas that we'll flesh out just a little bit here.

1. Notice that the elements of A are $1, -2, 1$ and 1. Add the squares of these numbers: $1^2 + (-2)^2 + 1^2 + 1^2 = 7 = \text{tr}(A^TA)$.

   Notice that the elements of B are $6, 7, 11$ and -4. Add the squares of these numbers: $6^2 + 7^2 + 11^2 + (-4)^2 = 222 = \text{tr}(B^TB)$.

   Can you see why this is true? When looking at multiplying A^TA, focus only on where the elements on the diagonal come from since they are the only ones that matter when taking the trace.

2. You can confirm on your own that regardless of the dimensions of A, $\text{tr}(A^TA) = \text{tr}(AA^T)$. To see why this is true, consider the previous point. (Recall also that A^TA and AA^T are always square, regardless of the dimensions of A.)

3. Mathematicians are actually more interested in $\sqrt{\text{tr}(A^TA)}$ than just $\text{tr}(A^TA)$. The reason for this is a bit complicated; the short answer is that "it works better." The reason "it works better" is related to the Pythagorean Theorem, all of all things. If we know that the legs of a right triangle have length a and b, we are more interested in $\sqrt{a^2 + b^2}$ than just $a^2 + b^2$. Of course, this explanation raises more questions than it answers; our goal here is just to whet your appetite and get you to do some more reading. A Numerical Linear Algebra book would be a good place to start.

There are many different measurements of a matrix size. In this text, we just refer to its dimensions. Some measurements of size refer the magnitude of the elements in the matrix. The next section describes yet another measurement of matrix size.

Exercises 6.2

Problems

In Exercises 1 – 15, find the trace of the given matrix.

1. $\begin{bmatrix} 1 & -5 \\ 9 & 5 \end{bmatrix}$

2. $\begin{bmatrix} -3 & -10 \\ -6 & 4 \end{bmatrix}$

3. $\begin{bmatrix} 7 & 5 \\ -5 & -4 \end{bmatrix}$

4. $\begin{bmatrix} -6 & 0 \\ -10 & 9 \end{bmatrix}$

5. $\begin{bmatrix} -4 & 1 & 1 \\ -2 & 0 & 0 \\ -1 & -2 & -5 \end{bmatrix}$

6. $\begin{bmatrix} 0 & -3 & 1 \\ 5 & -5 & 5 \\ -4 & 1 & 0 \end{bmatrix}$

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

8. $\begin{bmatrix} 4 & 2 & -1 \\ -4 & 1 & 4 \\ 0 & -5 & 5 \end{bmatrix}$

9. $\begin{bmatrix} 2 & 6 & 4 \\ -1 & 8 & -10 \end{bmatrix}$

10. $\begin{bmatrix} 6 & 5 \\ 2 & 10 \\ 3 & 3 \end{bmatrix}$

11. $\begin{bmatrix} -10 & 6 & -7 & -9 \\ -2 & 1 & 6 & -9 \\ 0 & 4 & -4 & 0 \\ -3 & -9 & 3 & -10 \end{bmatrix}$

12. $\begin{bmatrix} 5 & 2 & 2 & 2 \\ -7 & 4 & -7 & -3 \\ 9 & -9 & -7 & 2 \\ -4 & 8 & -8 & -2 \end{bmatrix}$

13. I_4

14. I_n

15. A matrix A that is skew symmetric.

In Exercises 16 – 19, verify Theorem 22 by:

1. **Showing that tr(A)+tr(B) = tr($A + B$) and**
2. **Showing that tr(AB) = tr(BA).**

16. $A = \begin{bmatrix} 1 & -1 \\ 9 & -6 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 0 \\ -6 & 3 \end{bmatrix}$

17. $A = \begin{bmatrix} 0 & -8 \\ 1 & 8 \end{bmatrix}, \quad B = \begin{bmatrix} -4 & 5 \\ -4 & 2 \end{bmatrix}$

18. $A = \begin{bmatrix} -8 & -10 & 10 \\ 10 & 5 & -6 \\ -10 & 1 & 3 \end{bmatrix}$

$B = \begin{bmatrix} -10 & -4 & -3 \\ -4 & -5 & 4 \\ 3 & 7 & 3 \end{bmatrix}$

19. $A = \begin{bmatrix} -10 & 7 & 5 \\ 7 & 7 & -5 \\ 8 & -9 & 2 \end{bmatrix}$

$B = \begin{bmatrix} -3 & -4 & 9 \\ 4 & -1 & -9 \\ -7 & -8 & 10 \end{bmatrix}$

6.3 The Determinant

AS YOU READ . . .

1. T/F: The determinant of a matrix is always positive.
2. T/F: To compute the determinant of a 3×3 matrix, one needs to compute the determinants of 3 2×2 matrices.
3. Give an example of a 2×2 matrix with a determinant of 3.

In this chapter so far we've learned about the transpose (an operation on a matrix that returns another matrix) and the trace (an operation on a square matrix that returns a number). In this section we'll learn another operation on square matrices that returns a number, called the *determinant*. We give a pseudo-definition of the determinant here.

> The *determinant* of an $n \times n$ matrix A is a number, denoted $\det(A)$, that is determined by A.

That definition isn't meant to explain everything; it just gets us started by making us realize that the determinant is a number. The determinant is kind of a tricky thing to define. Once you know and understand it, it isn't that hard, but getting started is a bit complicated. (It's similar to learning to ride a bike. The riding itself isn't hard, it is getting started that's difficult.) We start simply; we define the determinant for 2×2 matrices.

Definition 48 Determinant of 2×2 Matrices

Let

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

The *determinant of A*, denoted by

$$\det(A) \text{ or } \begin{vmatrix} a & b \\ c & d \end{vmatrix},$$

is $ad - bc$.

We've seen the expression $ad - bc$ before. In Section 5.3, we saw that a 2×2 matrix A has inverse

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

as long as $ad - bc \neq 0$; otherwise, the inverse does not exist. We can rephrase the above statement now: If $\det(A) \neq 0$, then

$$A^{-1} = \frac{1}{\det(A)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

A brief word about the notation: notice that we can refer to the determinant by using what *looks like* absolute value bars around the entries of a matrix. We discussed at the end of the last section the idea of measuring the "size" of a matrix, and mentioned that there are many different ways to measure size. The determinant is one such way. Just as the absolute value of a number measures its size (and ignores its sign), the determinant of a matrix is a measurement of the size of the matrix. (Be careful, though: $\det(A)$ can be negative!)

Let's practice.

Example 126 **Computing 2×2 determinants**
Find the determinant of A, B and C where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \; B = \begin{bmatrix} 3 & -1 \\ 2 & 7 \end{bmatrix} \text{ and } C = \begin{bmatrix} 1 & -3 \\ -2 & 6 \end{bmatrix}.$$

SOLUTION Finding the determinant of A:

$$\begin{aligned} \det(A) &= \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} \\ &= 1(4) - 2(3) \\ &= -2. \end{aligned}$$

Similar computations show that $\det(B) = 3(7) - (-1)(2) = 23$ and $\det(C) = 1(6) - (-3)(-2) = 0$.

Finding the determinant of a 2×2 matrix is pretty straightforward. It is natural to ask next "How do we compute the determinant of matrices that are not 2×2?" We first need to define some terms.

Definition 49 **Matrix Minor, Cofactor**

Let A be an $n \times n$ matrix. The *i,j minor of A*, denoted $A_{i,j}$, is the determinant of the $(n-1) \times (n-1)$ matrix formed by deleting the i^{th} row and j^{th} column of A.

The *i,j-cofactor of A* is the number

$$C_{ij} = (-1)^{i+j} A_{i,j}.$$

Notice that this definition makes reference to taking the determinant of a matrix, while we haven't yet defined what the determinant is beyond 2×2 matrices. We recognize this problem, and we'll see how far we can go before it becomes an issue.

Example 127 **Computing minors and cofactors**
Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 2 & 0 & 8 \\ -3 & 5 & 7 & 2 \\ -1 & 9 & -4 & 6 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Find $A_{1,3}$, $A_{3,2}$, $B_{2,1}$, $B_{4,3}$ and their respective cofactors.

SOLUTION To compute the minor $A_{1,3}$, we remove the first row and third column of A then take the determinant.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \Rightarrow \begin{bmatrix} \cancel{1} & \cancel{2} & \cancel{3} \\ 4 & 5 & \cancel{6} \\ 7 & 8 & \cancel{9} \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix}$$

$$A_{1,3} = \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = 32 - 35 = -3.$$

The corresponding cofactor, $C_{1,3}$, is

$$C_{1,3} = (-1)^{1+3}A_{1,3} = (-1)^4(-3) = -3.$$

The minor $A_{3,2}$ is found by removing the third row and second column of A then taking the determinant.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & \cancel{2} & 3 \\ 4 & \cancel{5} & 6 \\ \cancel{7} & \cancel{8} & \cancel{9} \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix}$$

$$A_{3,2} = \begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} = 6 - 12 = -6.$$

The corresponding cofactor, $C_{3,2}$, is

$$C_{3,2} = (-1)^{3+2}A_{3,2} = (-1)^5(-6) = 6.$$

The minor $B_{2,1}$ is found by removing the second row and first column of B then taking the determinant.

$$B = \begin{bmatrix} 1 & 2 & 0 & 8 \\ -3 & 5 & 7 & 2 \\ -1 & 9 & -4 & 6 \\ 1 & 1 & 1 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} \cancel{1} & 2 & 0 & 8 \\ \cancel{-3} & \cancel{5} & \cancel{7} & \cancel{2} \\ \cancel{-1} & 9 & -4 & 6 \\ \cancel{1} & 1 & 1 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 0 & 8 \\ 9 & -4 & 6 \\ 1 & 1 & 1 \end{bmatrix}$$

$$B_{2,1} = \begin{vmatrix} 2 & 0 & 8 \\ 9 & -4 & 6 \\ 1 & 1 & 1 \end{vmatrix} \overset{!}{=} ?$$

We're a bit stuck. We don't know how to find the determinant of this 3×3 matrix. We'll come back to this later. The corresponding cofactor is

$$C_{2,1} = (-1)^{2+1}B_{2,1} = -B_{2,1},$$

whatever this number happens to be.

The minor $B_{4,3}$ is found by removing the fourth row and third column of B then taking the determinant.

$$B = \begin{bmatrix} 1 & 2 & 0 & 8 \\ -3 & 5 & 7 & 2 \\ -1 & 9 & -4 & 6 \\ 1 & 1 & 1 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 2 & \cancel{0} & 8 \\ -3 & 5 & \cancel{7} & 2 \\ -1 & 9 & \cancel{-4} & 6 \\ \cancel{1} & \cancel{1} & \cancel{1} & \cancel{1} \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 2 & 8 \\ -3 & 5 & 2 \\ -1 & 9 & 6 \end{bmatrix}$$

$$B_{4,3} = \begin{vmatrix} 1 & 2 & 8 \\ -3 & 5 & 2 \\ -1 & 9 & 6 \end{vmatrix} \overset{!}{=} ?$$

Again, we're stuck. We won't be able to fully compute $C_{4,3}$; all we know so far is that

$$C_{4,3} = (-1)^{4+3}B_{4,3} = (-1)B_{4,3}.$$

Once we learn how to compute determinates for matrices larger than 2×2 we can come back and finish this exercise.

In our previous example we ran into a bit of trouble. By our definition, in order to compute a minor of an $n \times n$ matrix we needed to compute the determinant of a $(n-1) \times (n-1)$ matrix. This was fine when we started with a 3×3 matrix, but when we got up to a 4×4 matrix (and larger) we run into trouble.

We are almost ready to define the determinant for any square matrix; we need one last definition.

Definition 50 Cofactor Expansion

Let A be an $n \times n$ matrix.

The *cofactor expansion of A along the i^{th} row* is the sum

$$a_{i,1}C_{i,1} + a_{i,2}C_{i,2} + \cdots + a_{i,n}C_{i,n}.$$

The *cofactor expansion of A down the j^{th} column* is the sum

$$a_{1,j}C_{1,j} + a_{2,j}C_{2,j} + \cdots + a_{n,j}C_{n,j}.$$

The notation of this definition might be a little intimidating, so let's look at an example.

Example 128 Computing cofactor exapansions

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Find the cofactor expansions along the second row and down the first column.

Solution By the definition, the cofactor expansion along the second row is the sum

$$a_{2,1}C_{2,1} + a_{2,2}C_{2,2} + a_{2,3}C_{2,3}.$$

(Be sure to compare the above line to the definition of cofactor expansion, and see how the "i" in the definition is replaced by "2" here.)

We'll find each cofactor and then compute the sum.

$$C_{2,1} = (-1)^{2+1}\begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} = (-1)(-6) = 6 \qquad \left(\begin{array}{c}\text{we removed the second row and} \\ \text{first column of } A \text{ to compute the} \\ \text{minor}\end{array}\right)$$

$$C_{2,2} = (-1)^{2+2}\begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix} = (1)(-12) = -12 \qquad \left(\begin{array}{c}\text{we removed the second row and} \\ \text{second column of } A \text{ to compute} \\ \text{the minor}\end{array}\right)$$

$$C_{2,3} = (-1)^{2+3}\begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} = (-1)(-6) = 6 \qquad \left(\begin{array}{c}\text{we removed the second row and} \\ \text{third column of } A \text{ to compute the} \\ \text{minor}\end{array}\right)$$

Thus the cofactor expansion along the second row is

$$\begin{aligned} a_{2,1}C_{2,1} + a_{2,2}C_{2,2} + a_{2,3}C_{2,3} &= 4(6) + 5(-12) + 6(6) \\ &= 24 - 60 + 36 \\ &= 0 \end{aligned}$$

At the moment, we don't know what to do with this cofactor expansion; we've just successfully found it.

We move on to find the cofactor expansion down the first column. By the definition, this sum is

$$a_{1,1}C_{1,1} + a_{2,1}C_{2,1} + a_{3,1}C_{3,1}.$$

(Again, compare this to the above definition and see how we replaced the "j" with "1.")

We find each cofactor:

$$C_{1,1} = (-1)^{1+1}\begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} = (1)(-3) = -3 \qquad \left(\begin{array}{c}\text{we removed the first row and first} \\ \text{column of } A \text{ to compute the minor}\end{array}\right)$$

$$C_{2,1} = (-1)^{2+1}\begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} = (-1)(-6) = 6 \qquad \left(\text{we computed this cofactor above}\right)$$

$$C_{3,1} = (-1)^{3+1}\begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} = (1)(-3) = -3 \qquad \left(\begin{array}{c}\text{we removed the third row and first} \\ \text{column of } A \text{ to compute the minor}\end{array}\right)$$

The cofactor expansion down the first column is

$$\begin{aligned} a_{1,1}C_{1,1} + a_{2,1}C_{2,1} + a_{3,1}C_{3,1} &= 1(-3) + 4(6) + 7(-3) \\ &= -3 + 24 - 21 \\ &= 0 \end{aligned}$$

Is it a coincidence that both cofactor expansions were 0? We'll answer that in a while.

This section is entitled "The Determinant," yet we don't know how to compute it yet except for 2×2 matrices. We finally define it now.

Definition 51 The Determinant

The *determinant* of an $n \times n$ matrix A, denoted $\det(A)$ or $|A|$, is a number given by the following:

- if A is a 1×1 matrix $A = [a]$, then $\det(A) = a$.
- if A is a 2×2 matrix
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$
then $\det(A) = ad - bc$.
- if A is an $n \times n$ matrix, where $n \geq 2$, then $\det(A)$ is the number found by taking the cofactor expansion along the first row of A. That is,
$$\det(A) = a_{1,1}C_{1,1} + a_{1,2}C_{1,2} + \cdots + a_{1,n}C_{1,n}.$$

Notice that in order to compute the determinant of an $n \times n$ matrix, we need to compute the determinants of n $(n-1) \times (n-1)$ matrices. This can be a lot of work. We'll later learn how to shorten some of this. First, let's practice.

Example 129 Computing a 3×3 determinant

Find the determinant of

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Solution Notice that this is the matrix from Example 127. The cofactor expansion along the first row is

$$\det(A) = a_{1,1}C_{1,1} + a_{1,2}C_{1,2} + a_{1,3}C_{1,3}.$$

We'll compute each cofactor first then take the appropriate sum.

$$C_{1,1} = (-1)^{1+1}A_{1,1} = 1 \cdot \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} = 45 - 48 = -3$$

$$C_{1,2} = (-1)^{1+2}A_{1,2} = (-1) \cdot \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} = (-1)(36 - 42) = 6$$

$$C_{1,3} = (-1)^{1+3}A_{1,3} = 1 \cdot \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = 32 - 35 = -3$$

Therefore the determinant of A is

$$\det(A) = 1(-3) + 2(6) + 3(-3) = 0.$$

Example 130 **Another 3×3 determinant**

Find the determinant of

$$A = \begin{bmatrix} 3 & 6 & 7 \\ 0 & 2 & -1 \\ 3 & -1 & 1 \end{bmatrix}.$$

Solution We'll compute each cofactor first then find the determinant.

$$C_{1,1} = (-1)^{1+1}A_{1,1} = 1 \cdot \begin{vmatrix} 2 & -1 \\ -1 & 1 \end{vmatrix} = 2 - 1 = 1$$

$$C_{1,2} = (-1)^{1+2}A_{1,2} = (-1) \cdot \begin{vmatrix} 0 & -1 \\ 3 & 1 \end{vmatrix} = (-1)(0 + 3) = -3$$

$$C_{1,3} = (-1)^{1+3}A_{1,3} = 1 \cdot \begin{vmatrix} 0 & 2 \\ 3 & -1 \end{vmatrix} = 0 - 6 = -6$$

Thus the determinant is

$$\det(A) = 3(1) + 6(-3) + 7(-6) = -57.$$

Example 131 **Computing a 4×4 determinant**

Find the determinant of

$$A = \begin{bmatrix} 1 & 2 & 1 & 2 \\ -1 & 2 & 3 & 4 \\ 8 & 5 & -3 & 1 \\ 5 & 9 & -6 & 3 \end{bmatrix}.$$

Solution This, quite frankly, will take quite a bit of work. In order to compute this determinant, we need to compute 4 minors, each of which requires finding the determinant of a 3×3 matrix! Complaining won't get us any closer to the solution, (But it might make us feel a little better. Glance ahead: do you see how much work we have to do?!?) so let's get started. We first compute the cofactors:

$$\begin{aligned} C_{1,1} &= (-1)^{1+1}A_{1,1} \\ &= 1 \cdot \begin{vmatrix} 2 & 3 & 4 \\ 5 & -3 & 1 \\ 9 & -6 & 3 \end{vmatrix} \qquad \left(\begin{array}{l}\text{we must compute the determinant} \\ \text{of this } 3 \times 3 \text{ matrix}\end{array}\right) \\ &= 2 \cdot (-1)^{1+1} \begin{vmatrix} -3 & 1 \\ -6 & 3 \end{vmatrix} + 3 \cdot (-1)^{1+2} \begin{vmatrix} 5 & 1 \\ 9 & 3 \end{vmatrix} + 4 \cdot (-1)^{1+3} \begin{vmatrix} 5 & -3 \\ 9 & -6 \end{vmatrix} \\ &= 2(-3) + 3(-6) + 4(-3) \\ &= -36 \end{aligned}$$

$$\begin{aligned}
C_{1,2} &= (-1)^{1+2}A_{1,2} \\
&= (-1)\cdot\begin{vmatrix} -1 & 3 & 4 \\ 8 & -3 & 1 \\ 5 & -6 & 3 \end{vmatrix} \qquad \begin{pmatrix}\text{we must compute the determinant}\\ \text{of this } 3\times 3 \text{ matrix}\end{pmatrix} \\
&= (-1)\underbrace{\left[(-1)\cdot(-1)^{1+1}\begin{vmatrix} -3 & 1 \\ -6 & 3 \end{vmatrix} + 3\cdot(-1)^{1+2}\begin{vmatrix} 8 & 1 \\ 5 & 3 \end{vmatrix} + 4\cdot(-1)^{1+3}\begin{vmatrix} 8 & -3 \\ 5 & -6 \end{vmatrix}\right]}_{\text{the determinate of the } 3\times 3 \text{ matrix}} \\
&= (-1)\left[(-1)(-3) + 3(-19) + 4(-33)\right] \\
&= 186
\end{aligned}$$

$$\begin{aligned}
C_{1,3} &= (-1)^{1+3}A_{1,3} \\
&= 1\cdot\begin{vmatrix} -1 & 2 & 4 \\ 8 & 5 & 1 \\ 5 & 9 & 3 \end{vmatrix} \qquad \begin{pmatrix}\text{we must compute the determinant}\\ \text{of this } 3\times 3 \text{ matrix}\end{pmatrix} \\
&= (-1)\cdot(-1)^{1+1}\begin{vmatrix} 5 & 1 \\ 9 & 3 \end{vmatrix} + 2\cdot(-1)^{1+2}\begin{vmatrix} 8 & 1 \\ 5 & 3 \end{vmatrix} + 4\cdot(-1)^{1+3}\begin{vmatrix} 8 & 5 \\ 5 & 9 \end{vmatrix} \\
&= (-1)(6) + 2(-19) + 4(47) \\
&= 144
\end{aligned}$$

$$\begin{aligned}
C_{1,4} &= (-1)^{1+4}A_{1,4} \\
&= (-1)\cdot\begin{vmatrix} -1 & 2 & 3 \\ 8 & 5 & -3 \\ 5 & 9 & -6 \end{vmatrix} \qquad \begin{pmatrix}\text{we must compute the determinant}\\ \text{of this } 3\times 3 \text{ matrix}\end{pmatrix} \\
&= (-1)\underbrace{\left[(-1)\cdot(-1)^{1+1}\begin{vmatrix} 5 & -3 \\ 9 & -6 \end{vmatrix} + 2\cdot(-1)^{1+2}\begin{vmatrix} 8 & -3 \\ 5 & -6 \end{vmatrix} + 3\cdot(-1)^{1+3}\begin{vmatrix} 8 & 5 \\ 5 & 9 \end{vmatrix}\right]}_{\text{the determinate of the } 3\times 3 \text{ matrix}} \\
&= (-1)\left[(-1)(-3) + 2(33) + 3(47)\right] \\
&= -210
\end{aligned}$$

We've computed our four cofactors. All that is left is to compute the cofactor expansion.

$$\det(A) = 1(-36) + 2(186) + 1(144) + 2(-210) = 60.$$

As a way of "visualizing" this, let's write out the cofactor expansion again but including the matrices in their place.

$$\begin{aligned}\det(A) &= a_{1,1}C_{1,1} + a_{1,2}C_{1,2} + a_{1,3}C_{1,3} + a_{1,4}C_{1,4} \\ &= 1(-1)^2 \underbrace{\begin{vmatrix} 2 & 3 & 4 \\ 5 & -3 & 1 \\ 9 & -6 & 3 \end{vmatrix}}_{=-36} + 2(-1)^3 \underbrace{\begin{vmatrix} -1 & 3 & 4 \\ 8 & -3 & 1 \\ 5 & -6 & 3 \end{vmatrix}}_{=-186} \\ &+ \\ &\quad 1(-1)^4 \underbrace{\begin{vmatrix} -1 & 2 & 4 \\ 8 & 5 & 1 \\ 5 & 9 & 3 \end{vmatrix}}_{=144} + 2(-1)^5 \underbrace{\begin{vmatrix} -1 & 2 & 3 \\ 8 & 5 & -3 \\ 5 & 9 & -6 \end{vmatrix}}_{=210} \\ &= 60\end{aligned}$$

That certainly took a while; it required more than 50 multiplications (we didn't count the additions). To compute the determinant of a 5×5 matrix, we'll need to compute the determinants of five 4×4 matrices, meaning that we'll need over 250 multiplications! Not only is this a lot of work, but there are just too many ways to make silly mistakes. (The author made three when the above example was originally typed.) There are some tricks to make this job easier, but regardless we see the need to employ technology. Even then, technology quickly bogs down. A 25×25 matrix is considered "small" by today's standards, but it is essentially impossible for a computer to compute its determinant by only using cofactor expansion; it too needs to employ "tricks."

It is common for mathematicians, scientists and engineers to consider linear systems with thousands of equations and variables.

In the next section we will learn some of these tricks as we learn some of the properties of the determinant. Right now, let's review the essentials of what we have learned.

1. The determinant of a square matrix is a number that is determined by the matrix.
2. We find the determinant by computing the cofactor expansion along the first row.
3. To compute the determinant of an $n \times n$ matrix, we need to compute n determinants of $(n-1) \times (n-1)$ matrices.

Exercises 6.3

Problems

In Exercises 1 – 8, find the determinant of the 2×2 matrix.

1. $\begin{bmatrix} 10 & 7 \\ 8 & 9 \end{bmatrix}$

2. $\begin{bmatrix} 6 & -1 \\ -7 & 8 \end{bmatrix}$

3. $\begin{bmatrix} -1 & -7 \\ -5 & 9 \end{bmatrix}$

4. $\begin{bmatrix} -10 & -1 \\ -4 & 7 \end{bmatrix}$

5. $\begin{bmatrix} 8 & 10 \\ 2 & -3 \end{bmatrix}$

6. $\begin{bmatrix} 10 & -10 \\ -10 & 0 \end{bmatrix}$

7. $\begin{bmatrix} 1 & -3 \\ 7 & 7 \end{bmatrix}$

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

In Exercises 9 – 12, a matrix A is given.

(a) Construct the submatrices used to compute the minors $A_{1,1}$, $A_{1,2}$ and $A_{1,3}$.

(b) Find the cofactors $C_{1,1}$, $C_{1,2}$, and $C_{1,3}$.

9. $\begin{bmatrix} -7 & -3 & 10 \\ 3 & 7 & 6 \\ 1 & 6 & 10 \end{bmatrix}$

10. $\begin{bmatrix} -2 & -9 & 6 \\ -10 & -6 & 8 \\ 0 & -3 & -2 \end{bmatrix}$

11. $\begin{bmatrix} -5 & -3 & 3 \\ -3 & 3 & 10 \\ -9 & 3 & 9 \end{bmatrix}$

12. $\begin{bmatrix} -6 & -4 & 6 \\ -8 & 0 & 0 \\ -10 & 8 & -1 \end{bmatrix}$

In Exercises 13 – 24, find the determinant of the given matrix using cofactor expansion along the first row.

13. $\begin{bmatrix} 3 & 2 & 3 \\ -6 & 1 & -10 \\ -8 & -9 & -9 \end{bmatrix}$

14. $\begin{bmatrix} 8 & -9 & -2 \\ -9 & 9 & -7 \\ 5 & -1 & 9 \end{bmatrix}$

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

16. $\begin{bmatrix} 1 & -2 & 1 \\ 5 & 5 & 4 \\ 4 & 0 & 0 \end{bmatrix}$

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

18. $\begin{bmatrix} 3 & -1 & 0 \\ -3 & 0 & -4 \\ 0 & -1 & -4 \end{bmatrix}$

19. $\begin{bmatrix} -5 & 0 & -4 \\ 2 & 4 & -1 \\ -5 & 0 & -4 \end{bmatrix}$

20. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}$

21. $\begin{bmatrix} 0 & 0 & -1 & -1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix}$

22. $\begin{bmatrix} -1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & -1 & -1 \end{bmatrix}$

23. $\begin{bmatrix} -5 & 1 & 0 & 0 \\ -3 & -5 & 2 & 5 \\ -2 & 4 & -3 & 4 \\ 5 & 4 & -3 & 3 \end{bmatrix}$

24. $\begin{bmatrix} 2 & -1 & 4 & 4 \\ 3 & -3 & 3 & 2 \\ 0 & 4 & -5 & 1 \\ -2 & -5 & -2 & -5 \end{bmatrix}$

25. Let A be a 2×2 matrix;

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

Show why $\det(A) = ad - bc$ by computing the cofactor expansion of A along the first row.

6.4 Properties of the Determinant

AS YOU READ . . .

1. Having the choice to compute the determinant of a matrix using cofactor expansion along any row or column is most useful when there are lots of what in a row or column?
2. Which elementary row operation does not change the determinant of a matrix?
3. Why do mathematicians rarely smile?
4. T/F: When computers are used to compute the determinant of a matrix, cofactor expansion is rarely used.

In the previous section we learned how to compute the determinant. In this section we learn some of the properties of the determinant, and this will allow us to compute determinants more easily. In the next section we will see one application of determinants.

We start with a theorem that gives us more freedom when computing determinants.

Theorem 23 Cofactor Expansion Along Any Row or Column

Let A be an $n \times n$ matrix. The determinant of A can be computed using cofactor expansion along any row or column of A.

We alluded to this fact way back after Example 127. We had just learned what cofactor expansion was and we practiced along the second row and down the third column. Later, we found the determinant of this matrix by computing the cofactor expansion along the first row. In all three cases, we got the number 0. This wasn't a coincidence. The above theorem states that all three expansions were actually computing the determinant.

How does this help us? By giving us freedom to choose any row or column to use for the expansion, we can choose a row or column that looks "most appealing." This usually means "it has lots of zeros." We demonstrate this principle below.

Example 132 Computing a 4×4 determinant

Find the determinant of

$$A = \begin{bmatrix} 1 & 2 & 0 & 9 \\ 2 & -3 & 0 & 5 \\ 7 & 2 & 3 & 8 \\ -4 & 1 & 0 & 2 \end{bmatrix}.$$

Solution Our first reaction may well be "Oh no! Not another 4×4 determinant!" However, we can use cofactor expansion along any row or column that we choose. The third column looks great; it has lots of zeros in it. The cofactor expansion along this column is

$$\begin{aligned} \det(A) &= a_{1,3}C_{1,3} + a_{2,3}C_{2,3} + a_{3,3}C_{3,3} + a_{4,3}C_{4,3} \\ &= 0 \cdot C_{1,3} + 0 \cdot C_{2,3} + 3 \cdot C_{3,3} + 0 \cdot C_{4,3} \end{aligned}$$

The wonderful thing here is that three of our cofactors are multiplied by 0. We won't bother computing them since they will not contribute to the determinant. Thus

$$\begin{aligned}\det(A) &= 3 \cdot C_{3,3} \\ &= 3 \cdot (-1)^{3+3} \cdot \begin{vmatrix} 1 & 2 & 9 \\ 2 & -3 & 5 \\ -4 & 1 & 2 \end{vmatrix} \\ &= 3 \cdot (-147) \qquad \left(\begin{array}{c}\text{we computed the determinant of the } 3 \times 3 \text{ matrix} \\ \text{without showing our work; it is } -147\end{array}\right) \\ &= -447\end{aligned}$$

Wow. That was a lot simpler than computing all that we did in Example 130. Of course, in that example, we didn't really have any shortcuts that we could have employed. Our next example involves a 5×5 determinant. At first, this looks like trouble, until we realize that the matrix is *triangular*. As we'll see, this makes our job much easier.

Example 133 Computing the determinant of a 5×5 (triangular) matrix
Find the determinant of

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 6 & 7 & 8 & 9 \\ 0 & 0 & 10 & 11 & 12 \\ 0 & 0 & 0 & 13 & 14 \\ 0 & 0 & 0 & 0 & 15 \end{bmatrix}.$$

Solution Since we can expand along any row or column, things are not as bad as they might at first seem. In fact, this problem is very easy. What row or column should we choose to find the determinant along? There are two obvious choices: the first column or the last row. Both have 4 zeros in them. We choose the first column. We omit most of the cofactor expansion, since most of it is just 0:

$$\det(A) = 1 \cdot (-1)^{1+1} \cdot \begin{vmatrix} 6 & 7 & 8 & 9 \\ 0 & 10 & 11 & 12 \\ 0 & 0 & 13 & 14 \\ 0 & 0 & 0 & 15 \end{vmatrix}.$$

Similarly, this determinant is not bad to compute; we again choose to use cofactor expansion along the first column. Note: technically, this cofactor expansion is $6 \cdot (-1)^{1+1} A_{1,1}$; we are going to drop the $(-1)^{1+1}$ terms from here on out in this example (it will show up a lot...).

$$\det(A) = 1 \cdot 6 \cdot \begin{vmatrix} 10 & 11 & 12 \\ 0 & 13 & 14 \\ 0 & 0 & 15 \end{vmatrix}.$$

You can probably can see a trend. We'll finish out the steps without explaining each one.

$$\begin{aligned}\det(A) &= 1 \cdot 6 \cdot 10 \cdot \begin{vmatrix} 13 & 14 \\ 0 & 15 \end{vmatrix} \\ &= 1 \cdot 6 \cdot 10 \cdot 13 \cdot 15 \\ &= 11700\end{aligned}$$

We see that the final determinant is the product of the diagonal entries. This works for any triangular matrix (and since diagonal matrices are triangular, it works for diagonal matrices as well). This is an important enough idea that we'll put it into a box.

Key Idea 23 The Determinant of Triangular Matrices

The determinant of a triangular matrix is the product of its diagonal elements.

It is now again time to start thinking like a mathematician. Remember, mathematicians see something new and often ask "How does this relate to things I already know?" So now we ask, "If we change a matrix in some way, how is its determinant changed?"

The standard way that we change matrices is through elementary row operations. If we perform an elementary row operation on a matrix, how will the determinant of the new matrix compare to the determinant of the original matrix?

Let's experiment first and then we'll officially state what happens.

Example 134 Row operations and determinants

Let

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

Let B be formed from A by doing one of the following elementary row operations:

1. $2R_1 + R_2 \rightarrow R_2$

2. $5R_1 \rightarrow R_1$

3. $R_1 \leftrightarrow R_2$

Find $\det(A)$ as well as $\det(B)$ for each of the row operations above.

Solution It is straightforward to compute $\det(A) = -2$.

Let B be formed by performing the row operation in 1) on A; thus

$$B = \begin{bmatrix} 1 & 2 \\ 5 & 8 \end{bmatrix}.$$

It is clear that $\det(B) = -2$, the same as $\det(A)$.

Now let B be formed by performing the elementary row operation in 2) on A; that is,

$$B = \begin{bmatrix} 5 & 10 \\ 3 & 4 \end{bmatrix}.$$

We can see that $\det(B) = -10$, which is $5 \cdot \det(A)$.

Finally, let B be formed by the third row operation given; swap the two rows of A. We see that

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

and that $\det(B) = 2$, which is $(-1) \cdot \det(A)$.

We've seen in the above example that there seems to be a relationship between the determinants of matrices "before and after" being changed by elementary row operations. Certainly, one example isn't enough to base a theory on, and we have not proved anything yet. Regardless, the following theorem is true.

Theorem 24 The Determinant and Elementary Row Operations

Let A be an $n \times n$ matrix and let B be formed by performing one elementary row operation on A.

1. If B is formed from A by adding a scalar multiple of one row to another, then $\det(B) = \det(A)$.
2. If B is formed from A by multiplying one row of A by a scalar k, then $\det(B) = k \cdot \det(A)$.
3. If B is formed from A by interchanging two rows of A, then $\det(B) = -\det(A)$.

Let's put this theorem to use in an example.

Example 135 Using row operations to compute a determinant

Let

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

Compute $\det(A)$, then find the determinants of the following matrices by inspection using Theorem 24.

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 7 & 7 & 7 \end{bmatrix} \quad D = \begin{bmatrix} 1 & -1 & -2 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Solution Computing $\det(A)$ by cofactor expansion down the first column or along the second row seems like the best choice, utilizing the one zero in the matrix. We can quickly confirm that $\det(A) = 1$.

To compute $\det(B)$, notice that the rows of A were rearranged to form B. There are different ways to describe what happened; saying $R_1 \leftrightarrow R_2$ was followed by $R_1 \leftrightarrow R_3$ produces B from A. Since there were *two* row swaps, $\det(B) = (-1)(-1)\det(A) = \det(A) = 1$.

Notice that C is formed from A by multiplying the third row by 7. Thus $\det(C) = 7 \cdot \det(A) = 7$.

It takes a little thought, but we can form D from A by the operation $-3R_2 + R_1 \rightarrow R_1$. This type of elementary row operation does not change determinants, so $\det(D) = \det(A)$.

Let's continue to think like mathematicians; mathematicians tend to remember "problems" they've encountered in the past, and when they learn something new, in the backs of their minds they try to apply their new knowledge to solve their old problem. (This is why mathematicians rarely smile: they are remembering their problems)

What "problem" did we recently uncover? We stated in the last chapter that even computers could not compute the determinant of large matrices with cofactor expansion. How then can we compute the determinant of large matrices?

We just learned two interesting and useful facts about matrix determinants. First, the determinant of a triangular matrix is easy to compute: just multiply the diagonal elements. Secondly, we know how elementary row operations affect the determinant. Put these two ideas together: given any square matrix, we can use elementary row operations to put the matrix in triangular form, find the determinant of the new matrix (which is easy), and then adjust that number by recalling what elementary operations we performed.

Example 136 **Using row operations to reduce a determinant to triangular form**

Find the determinant of A by first putting A into a triangular form, where

$$A = \begin{bmatrix} 2 & 4 & -2 \\ -1 & -2 & 5 \\ 3 & 2 & 1 \end{bmatrix}.$$

Solution In putting A into a triangular form, we need not worry about getting leading 1s, but it does tend to make our life easier as we work out a problem by hand. So let's scale the first row by $1/2$:

$$\frac{1}{2}R_1 \to R_1 \qquad \begin{bmatrix} 1 & 2 & -1 \\ -1 & -2 & 5 \\ 3 & 2 & 1 \end{bmatrix}.$$

Now let's get 0s below this leading 1:

$$\begin{array}{c} R_1 + R_2 \to R_2 \\ -3R_1 + R_3 \to R_3 \end{array} \qquad \begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 4 \\ 0 & -4 & 4 \end{bmatrix}.$$

We can finish in one step; by interchanging rows 2 and 3 we'll have our matrix in triangular form.

$$R_2 \leftrightarrow R_3 \qquad \begin{bmatrix} 1 & 2 & -1 \\ 0 & -4 & 4 \\ 0 & 0 & 4 \end{bmatrix}.$$

Let's name this last matrix B. The determinant of B is easy to compute as it is triangular; $\det(B) = -16$. We can use this to find $\det(A)$.

Recall the steps we used to transform A into B. They are:

$$\begin{array}{c} \frac{1}{2}R_1 \to R_1 \\ R_1 + R_2 \to R_2 \\ -3R_1 + R_3 \to R_3 \\ R_2 \leftrightarrow R_3 \end{array}$$

The first operation multiplied a row of A by $\frac{1}{2}$. This means that the resulting matrix had a determinant that was $\frac{1}{2}$ the determinant of A.

The next two operations did not affect the determinant at all. The last operation, the row swap, changed the sign. Combining these effects, we know that

$$-16 = \det(B) = (-1)\frac{1}{2}\det(A).$$

Solving for $\det(A)$ we have that $\det(A) = 32$.

In practice, we don't need to keep track of operations where we add multiples of one row to another; they simply do not affect the determinant. Also, in practice, these steps are carried out by a computer, and computers don't care about leading 1s. Therefore, row scaling operations are rarely used. The only things to keep track of are row swaps, and even then all we care about are the number of row swaps. An odd number of row swaps means that the original determinant has the opposite sign of the triangular form matrix; an even number of row swaps means they have the same determinant.

Let's practice this again.

Example 137 **Effect of elementary row operations ion the determinant**
The matrix B was formed from A using the following elementary row operations, though not necessarily in this order. Find $\det(A)$.

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \qquad \begin{array}{l} 2R_1 \to R_1 \\ \frac{1}{3}R_3 \to R_3 \\ R_1 \leftrightarrow R_2 \\ 6R_1 + R_2 \to R_2 \end{array}$$

Solution It is easy to compute $\det(B) = 24$. In looking at our list of elementary row operations, we see that only the first three have an effect on the determinant. Therefore

$$24 = \det(B) = 2 \cdot \frac{1}{3} \cdot (-1) \cdot \det(A)$$

and hence

$$\det(A) = -36.$$

In the previous example, we may have been tempted to "rebuild" A using the elementary row operations and then computing the determinant. This can be done, but in general it is a bad idea; it takes too much work and it is too easy to make a mistake.

Let's think some more like a mathematician. How does the determinant work with other matrix operations that we know? Specifically, how does the determinant interact with matrix addition, scalar multiplication, matrix multiplication, the transpose and the trace? We'll again do an example to get an idea of what is going on, then give a theorem to state what is true.

Example 138 **Determinants and matrix operations**
Let

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}.$$

Find the determinants of the matrices A, B, $A + B$, $3A$, AB, A^T, A^{-1}. Can you find any connections between these values?

Solution We can quickly compute that $\det(A) = -2$ and that $\det(B) = 7$.

$$\begin{aligned} \det(A - B) &= \det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}\right) \\ &= \begin{vmatrix} -1 & 1 \\ 0 & -1 \end{vmatrix} \\ &= 1 \end{aligned}$$

It's tough to find a connection between $\det(A - B)$, $\det(A)$ and $\det(B)$.

$$\begin{aligned} \det(3A) &= \begin{vmatrix} 3 & 6 \\ 9 & 12 \end{vmatrix} \\ &= -18 \end{aligned}$$

We can figure this one out; multiplying one row of A by 3 increases the determinant by a factor of 3; doing it again (and hence multiplying both rows by 3) increases the determinant again by a factor of 3. Therefore $\det(3A) = 3 \cdot 3 \cdot \det(A)$, or $3^2 \cdot \det(A)$.

$$\begin{aligned} \det(AB) &= \det\left(\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}\right) \\ &= \begin{vmatrix} 8 & 11 \\ 18 & 23 \end{vmatrix} \\ &= -14 \end{aligned}$$

This one seems clear; $\det(AB) = \det(A)\det(B)$.

$$\begin{aligned} \det\left(A^T\right) &= \begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} \\ &= -2 \end{aligned}$$

Obviously $\det\left(A^T\right) = \det(A)$; is this always going to be the case? If we think about it, we can see that the cofactor expansion along the first *row* of A will give us the same result as the cofactor expansion along the first *column* of A^T.

Seeing that expansion along the first row agrees with expansion along the first column can be a bit tricky to think out in your head. Try it with a 3×3 matrix A and see how it works. All the 2×2 submatrices that are created in A^T are the transpose of those found in A; this doesn't matter since it is easy to see that the determinant isn't affected by the transpose in a 2×2 matrix.

$$\begin{aligned} \det\left(A^{-1}\right) &= \begin{vmatrix} -2 & 1 \\ 3/2 & -1/2 \end{vmatrix} \\ &= 1 - 3/2 \\ &= -1/2 \end{aligned}$$

It seems as though

$$\det\left(A^{-1}\right) = \frac{1}{\det(A)}.$$

We now state a theorem which will confirm our conjectures from the previous example.

Theorem 25 Determinant Properties

Let A and B be $n \times n$ matrices and let k be a scalar. The following are true:

1. $\det(kA) = k^n \cdot \det(A)$
2. $\det(A^T) = \det(A)$
3. $\det(AB) = \det(A)\det(B)$
4. If A is invertible, then
$$\det\left(A^{-1}\right) = \frac{1}{\det(A)}.$$
5. A matrix A is invertible if and only if $\det(A) \neq 0$.

This last statement of the above theorem is significant: what happens if $\det(A) = 0$? It seems that $\det\left(A^{-1}\right) =$"1/0", which is undefined. There actually isn't a problem here; it turns out that if $\det(A) = 0$, then A is not invertible (hence part 5 of Theorem 25). This allows us to add on to our Invertible Matrix Theorem.

Theorem 26 Invertible Matrix Theorem

Let A be an $n \times n$ matrix. The following statements are equivalent.

(a) A is invertible.

(g) $\det(A) \neq 0$.

This new addition to the Invertible Matrix Theorem is very useful; we'll refer back to it in Chapter 8 when we discuss eigenvalues.

In the next section we'll see how the determinant can be used to solve systems of linear equations.

Exercises 6.4

Problems

In Exercises 1 – 14, find the determinant of the given matrix using cofactor expansion along any row or column you choose.

1. $\begin{bmatrix} 1 & 2 & 3 \\ -5 & 0 & 3 \\ 4 & 0 & 6 \end{bmatrix}$

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

3. $\begin{bmatrix} -4 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -5 \end{bmatrix}$

4. $\begin{bmatrix} 0 & -3 & 1 \\ 0 & 0 & 5 \\ -4 & 1 & 0 \end{bmatrix}$

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

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

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

8. $\begin{bmatrix} 0 & 4 & -4 \\ 3 & 1 & -3 \\ -3 & -4 & 0 \end{bmatrix}$

9. $\begin{bmatrix} 5 & -5 & 0 & 1 \\ 2 & 4 & -1 & -1 \\ 5 & 0 & 0 & 4 \\ -1 & -2 & 0 & 5 \end{bmatrix}$

10. $\begin{bmatrix} -1 & 3 & 3 & 4 \\ 0 & 0 & 0 & 0 \\ 4 & -5 & -2 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix}$

11. $\begin{bmatrix} -5 & -5 & 0 & -2 \\ 0 & 0 & 5 & 0 \\ 1 & 3 & 3 & 1 \\ -4 & -2 & -1 & -5 \end{bmatrix}$

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

13. $\begin{bmatrix} 4 & 0 & 5 & 1 & 0 \\ 1 & 0 & 3 & 1 & 5 \\ 2 & 2 & 0 & 2 & 2 \\ 1 & 0 & 0 & 0 & 0 \\ 4 & 4 & 2 & 5 & 3 \end{bmatrix}$

14. $\begin{bmatrix} 2 & 1 & 1 & 1 & 1 \\ 4 & 1 & 2 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 3 & 2 & 0 & 3 \\ 5 & 0 & 5 & 0 & 4 \end{bmatrix}$

In Exercises 15 – 18, a matrix M and det(M) are given. Matrices A, B and C are formed by performing operations on M. Determine the determinants of A, B and C using Theorems 24 and 25, and indicate the operations used to form A, B and C.

15. $M = \begin{bmatrix} 0 & 3 & 5 \\ 3 & 1 & 0 \\ -2 & -4 & -1 \end{bmatrix}$,
$\det(M) = -41$.

(a) $A = \begin{bmatrix} 0 & 3 & 5 \\ -2 & -4 & -1 \\ 3 & 1 & 0 \end{bmatrix}$

(b) $B = \begin{bmatrix} 0 & 3 & 5 \\ 3 & 1 & 0 \\ 8 & 16 & 4 \end{bmatrix}$

(c) $C = \begin{bmatrix} 3 & 4 & 5 \\ 3 & 1 & 0 \\ -2 & -4 & -1 \end{bmatrix}$

16. $M = \begin{bmatrix} 9 & 7 & 8 \\ 1 & 3 & 7 \\ 6 & 3 & 3 \end{bmatrix}$,
$\det(M) = 45$.

(a) $A = \begin{bmatrix} 18 & 14 & 16 \\ 1 & 3 & 7 \\ 6 & 3 & 3 \end{bmatrix}$

(b) $B = \begin{bmatrix} 9 & 7 & 8 \\ 1 & 3 & 7 \\ 96 & 73 & 83 \end{bmatrix}$

(c) $C = \begin{bmatrix} 9 & 1 & 6 \\ 7 & 3 & 3 \\ 8 & 7 & 3 \end{bmatrix}$

17. $M = \begin{bmatrix} 5 & 1 & 5 \\ 4 & 0 & 2 \\ 0 & 0 & 4 \end{bmatrix}$,
$\det(M) = -16$.

(a) $A = \begin{bmatrix} 0 & 0 & 4 \\ 5 & 1 & 5 \\ 4 & 0 & 2 \end{bmatrix}$

(b) $B = \begin{bmatrix} -5 & -1 & -5 \\ -4 & 0 & -2 \\ 0 & 0 & 4 \end{bmatrix}$

(c) $C = \begin{bmatrix} 15 & 3 & 15 \\ 12 & 0 & 6 \\ 0 & 0 & 12 \end{bmatrix}$

18. $M = \begin{bmatrix} 5 & 4 & 0 \\ 7 & 9 & 3 \\ 1 & 3 & 9 \end{bmatrix}$,
$\det(M) = 120$.

(a) $A = \begin{bmatrix} 1 & 3 & 9 \\ 7 & 9 & 3 \\ 5 & 4 & 0 \end{bmatrix}$

(b) $B = \begin{bmatrix} 5 & 4 & 0 \\ 14 & 18 & 6 \\ 3 & 9 & 27 \end{bmatrix}$

(c) $C = \begin{bmatrix} -5 & -4 & 0 \\ -7 & -9 & -3 \\ -1 & -3 & -9 \end{bmatrix}$

In Exercises 19 – 22, matrices A and B are given. Verify part 3 of Theorem 25 by computing det(A), det(B) and det(AB).

19. $A = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$,
$B = \begin{bmatrix} 0 & -4 \\ 1 & 3 \end{bmatrix}$

20. $A = \begin{bmatrix} 3 & -1 \\ 4 & 1 \end{bmatrix}$,
$B = \begin{bmatrix} -4 & -1 \\ -5 & 3 \end{bmatrix}$

21. $A = \begin{bmatrix} -4 & 4 \\ 5 & -2 \end{bmatrix}$,
$B = \begin{bmatrix} -3 & -4 \\ 5 & -3 \end{bmatrix}$

22. $A = \begin{bmatrix} -3 & -1 \\ 2 & -3 \end{bmatrix}$,
$B = \begin{bmatrix} 0 & 0 \\ 4 & -4 \end{bmatrix}$

In Exercises 23 – 30, find the determinant of the given matrix.

23. $\begin{bmatrix} 3 & 2 & 3 \\ -6 & 1 & -10 \\ -8 & -9 & -9 \end{bmatrix}$

24. $\begin{bmatrix} 8 & -9 & -2 \\ -9 & 9 & -7 \\ 5 & -1 & 9 \end{bmatrix}$

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

26. $\begin{bmatrix} 1 & -2 & 1 \\ 5 & 5 & 4 \\ 4 & 0 & 0 \end{bmatrix}$

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

28. $\begin{bmatrix} 3 & -1 & 0 \\ -3 & 0 & -4 \\ 0 & -1 & -4 \end{bmatrix}$

29. $\begin{bmatrix} -5 & 0 & -4 \\ 2 & 4 & -1 \\ -5 & 0 & -4 \end{bmatrix}$

30. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}$

6.5 Cramer's Rule

AS YOU READ . . .

1. T/F: Cramer's Rule is another method to compute the determinant of a matrix.

2. T/F: Cramer's Rule is often used because it is more efficient than Gaussian elimination.

3. Mathematicians use what word to describe the connections between seemingly unrelated ideas?

In the previous sections we have learned about the determinant, but we haven't given a really good reason *why* we would want to compute it. This section shows one application of the determinant: solving systems of linear equations. We introduce this idea in terms of a theorem, then we will practice.

The closest we came to motivating the determinant is that if $\det(A) = 0$, then we know that A is not invertible. But it seems that there may be easier ways to check.
It is interesting to note that despite the presentation given here, determinants actually pre-date the modern usage of matrices by more than a century. Cramer's rule was published by Cramer in 1750, and the term matrix was introduced by James Joseph Sylvester in 1850. (Even then, Sylvester's description of matrices was in terms of minors – that's right, determinants.) The interested reader is encouraged to read up on the history of the subject. (Wikipedia is not a bad place to start.)

Theorem 27 Cramer's Rule

Let A be an $n \times n$ matrix with $\det(A) \neq 0$ and let $\vec{b}$ be an $n \times 1$ column vector. Then the linear system

$$A\vec{x} = \vec{b}$$

has solution

$$x_i = \frac{\det\big(A_i(\vec{b})\big)}{\det(A)},$$

where $A_i(\vec{b})$ is the matrix formed by replacing the i^{th} column of A with $\vec{b}$.

Example 139 Using Cramer's Rule

Use Cramer's Rule to solve the linear system $A\vec{x} = \vec{b}$ where

$$A = \begin{bmatrix} 1 & 5 & -3 \\ 1 & 4 & 2 \\ 2 & -1 & 0 \end{bmatrix} \quad \text{and} \quad \vec{b} = \begin{bmatrix} -36 \\ -11 \\ 7 \end{bmatrix}.$$

Solution We first compute the determinant of A to see if we can apply Cramer's Rule.

$$\det(A) = \begin{vmatrix} 1 & 5 & -3 \\ 1 & 4 & 2 \\ 2 & -1 & 0 \end{vmatrix} = 49.$$

Since $\det(A) \neq 0$, we can apply Cramer's Rule. Following Theorem 27, we compute $\det\big(A_1(\vec{b})\big)$, $\det\big(A_2(\vec{b})\big)$ and $\det\big(A_3(\vec{b})\big)$.

$$\det\big(A_1(\vec{b})\big) = \begin{vmatrix} -36 & 5 & -3 \\ -11 & 4 & 2 \\ 7 & -1 & 0 \end{vmatrix} = 49.$$

(We used a bold font to show where $\vec{b}$ replaced the first column of A.)

$$\det\left(A_2(\vec{b})\right) = \begin{vmatrix} 1 & -36 & -3 \\ 1 & -11 & 2 \\ 2 & 7 & 0 \end{vmatrix} = -245.$$

$$\det\left(A_3(\vec{b})\right) = \begin{vmatrix} 1 & 5 & -36 \\ 1 & 4 & -11 \\ 2 & -1 & 7 \end{vmatrix} = 196.$$

Therefore we can compute $\vec{x}$:

$$x_1 = \frac{\det\left(A_1(\vec{b})\right)}{\det(A)} = \frac{49}{49} = 1$$

$$x_2 = \frac{\det\left(A_2(\vec{b})\right)}{\det(A)} = \frac{-245}{49} = -5$$

$$x_3 = \frac{\det\left(A_3(\vec{b})\right)}{\det(A)} = \frac{196}{49} = 4$$

Therefore

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -5 \\ 4 \end{bmatrix}.$$

Example 140 Using Cramer's Rule

Use Cramer's Rule to solve the linear system $A\vec{x} = \vec{b}$ where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ and } \vec{b} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Solution The determinant of A is -2, so we can apply Cramer's Rule.

$$\det\left(A_1(\vec{b})\right) = \begin{vmatrix} -1 & 2 \\ 1 & 4 \end{vmatrix} = -6.$$

$$\det\left(A_2(\vec{b})\right) = \begin{vmatrix} 1 & -1 \\ 3 & 1 \end{vmatrix} = 4.$$

Therefore

$$x_1 = \frac{\det\left(A_1(\vec{b})\right)}{\det(A)} = \frac{-6}{-2} = 3$$

$$x_2 = \frac{\det\left(A_2(\vec{b})\right)}{\det(A)} = \frac{4}{-2} = -2$$

and

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}.$$

We learned in Section 6.4 that when considering a linear system $A\vec{x} = \vec{b}$ where A is square, if $\det(A) \neq 0$ then A is invertible and $A\vec{x} = \vec{b}$ has exactly one solution. We also stated in Key Idea 22 that if $\det(A) = 0$, then A is not invertible and so therefore either $A\vec{x} = \vec{b}$ has no solution or infinite solutions. Our method of figuring out which of these cases applied was to form the augmented matrix $\begin{bmatrix} A & \vec{b} \end{bmatrix}$, put it into reduced row echelon form, and then interpret the results.

Cramer's Rule specifies that $\det(A) \neq 0$ (so we are guaranteed a solution). When $\det(A) = 0$ we are not able to discern whether infinite solutions or no solution exists for a given vector $\vec{b}$. Cramer's Rule is only applicable to the case when exactly one solution exists.

We end this section with a practical consideration. We have mentioned before that finding determinants is a computationally intensive operation. To solve a linear system with 3 equations and 3 unknowns, we need to compute 4 determinants. Just think: with 10 equations and 10 unknowns, we'd need to compute 11 really hard determinants of 10×10 matrices! That is a lot of work!

The upshot of this is that Cramer's Rule makes for a poor choice in solving numerical linear systems. It simply is not done in practice; it is hard to beat Gaussian elimination.

So why include it? *Because its truth is amazing.* The determinant is a very strange operation; it produces a number in a very odd way. It should seem incredible to the reader that by manipulating determinants in a particular way, we can solve linear systems.

In the next chapter we'll see another use for the determinant. Meanwhile, try to develop a deeper appreciation of math: odd, complicated things that seem completely unrelated often are intricately tied together. Mathematicians see these connections and describe them as "beautiful."

A version of Cramer's Rule is often taught in introductory differential equations courses as it can be used to find solutions to certain linear differential equations. In this situation, the entries of the matrices are functions, not numbers, and hence computing determinants is easier than using Gaussian elimination. Again, though, as the matrices get large, other solution methods are resorted to.

Exercises 6.5

Problems

In Exercises 1 – 12, matrices A and $\vec{b}$ are given.

(a) Give det(A) and det(A_i) for all i.

(b) Use Cramer's Rule to solve $A\vec{x} = \vec{b}$. If Cramer's Rule cannot be used to find the solution, then state whether or not a solution exists.

1. $A = \begin{bmatrix} 7 & -7 \\ -7 & 9 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 28 \\ -26 \end{bmatrix}$

2. $A = \begin{bmatrix} 9 & 5 \\ -4 & -7 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -45 \\ 20 \end{bmatrix}$

3. $A = \begin{bmatrix} -8 & 16 \\ 10 & -20 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -48 \\ 60 \end{bmatrix}$

4. $A = \begin{bmatrix} 0 & -6 \\ 9 & -10 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 6 \\ -17 \end{bmatrix}$

5. $A = \begin{bmatrix} 2 & 10 \\ -1 & 3 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 42 \\ 19 \end{bmatrix}$

6. $A = \begin{bmatrix} 7 & 14 \\ -2 & -4 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$

7. $A = \begin{bmatrix} 3 & 0 & -3 \\ 5 & 4 & 4 \\ 5 & 5 & -4 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 24 \\ 0 \\ 31 \end{bmatrix}$

8. $A = \begin{bmatrix} 4 & 9 & 3 \\ -5 & -2 & -13 \\ -1 & 10 & -13 \end{bmatrix},$
$\vec{b} = \begin{bmatrix} -28 \\ 35 \\ 7 \end{bmatrix}$

9. $A = \begin{bmatrix} 4 & -4 & 0 \\ 5 & 1 & -1 \\ 3 & -1 & 2 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 16 \\ 22 \\ 8 \end{bmatrix}$

10. $A = \begin{bmatrix} 1 & 0 & -10 \\ 4 & -3 & -10 \\ -9 & 6 & -2 \end{bmatrix},$
$\vec{b} = \begin{bmatrix} -40 \\ -94 \\ 132 \end{bmatrix}$

11. $A = \begin{bmatrix} 7 & -4 & 25 \\ -2 & 1 & -7 \\ 9 & -7 & 34 \end{bmatrix},$
$\vec{b} = \begin{bmatrix} -1 \\ -3 \\ 5 \end{bmatrix}$

12. $A = \begin{bmatrix} -6 & -7 & -7 \\ 5 & 4 & 1 \\ 5 & 4 & 8 \end{bmatrix},$
$\vec{b} = \begin{bmatrix} 58 \\ -35 \\ -49 \end{bmatrix}$

7: The Complex Numbers

7.1 Complex Numbers

We now move on to the study of the set of **complex numbers**. As you may recall, the complex numbers fill an algebraic gap left by the real numbers. There is no real number x with $x^2 = -1$, since for any real number $x^2 \geq 0$. However, we could formally extract square roots and write $x = \pm\sqrt{-1}$. We build the complex numbers by relabelling the quantity $\sqrt{-1}$ as i, the unfortunately misnamed **imaginary unit**. The number i, while not a real number, is defined so that it plays along well with real numbers and acts very much like any other radical expression. For instance, $3(2i) = 6i$, $7i - 3i = 4i$, $(2 - 7i) + (3 + 4i) = 5 - 3i$, and so forth. The key properties which distinguish i from the real numbers are listed below.

Historically, the lack of solutions to the equation $x^2 = -1$ had nothing to do with the development of the complex numbers. Until the 19th century, equations such as $x^2 = -1$ would have been considered in the context of the analytic geometry of Descartes. The lack of solutions simply indicated that the graph $y = x^2$ did not intersect the line $y = -1$. The more remarkable case was that of *cubic* equations, of the form $x^3 = ax + b$. In this case a **real** solution is *guaranteed*, but there are cases where one needs **complex** numbers to find it! For details, see the excellent book *Visual Complex Analysis*, by Tristan Needham.

Definition 52 The imaginary unit

The imaginary unit i satisfies the two following properties:

1. $i^2 = -1$
2. If c is a real number with $c \geq 0$ then $\sqrt{-c} = i\sqrt{c}$

Note the use of the indefinite article 'a'. Whatever beast is chosen to be i, $-i$ is the other square root of -1.

Property 1 in Definition 52 establishes that i does act as a square root of -1, and property 2 establishes what we mean by the 'principal square root' of a negative real number. In property 2, it is important to remember the restriction on c. For example, it is perfectly acceptable to say $\sqrt{-4} = i\sqrt{4} = i(2) = 2i$. However, $\sqrt{-(-4)} \neq i\sqrt{-4}$, otherwise, we'd get

$$2 = \sqrt{4} = \sqrt{-(-4)} = i\sqrt{-4} = i(2i) = 2i^2 = 2(-1) = -2,$$

which is unacceptable. The moral of this story is that the general properties of radicals do not apply for even roots of negative quantities. With Definition 52 in place, we are now in position to define the **complex numbers**.

Some Technical Mathematics textbooks label the imaginary unit 'j', usually to avoid confusion with the use of the letter i to denote electric current. While it carries the adjective 'imaginary', these numbers have essential real-world implications. For example, every electronic device owes its existence to the study of 'imaginary' numbers.

Definition 53 Complex number

A **complex number** is a number of the form $a + bi$, where a and b are real numbers and i is the imaginary unit. The set of complex numbers is denoted $\mathbb{C}$.

Complex numbers include things you'd normally expect, like $3 + 2i$ and $\frac{2}{5} - i\sqrt{3}$. However, don't forget that a or b could be zero, which means numbers like $3i$ and 6 are also complex numbers. In other words, don't forget that the complex numbers *include* the real numbers, so 0 and $\pi - \sqrt{21}$ are both considered complex numbers. We want to study the arithmetic of complex numbers, but before we can do so, we first need to make sure we understand what it means for two complex numbers to be equal.

Definition 54 Equality of complex numbers

Let $z = a + ib$ and $w = c + id$ be two complex numbers. We say that z and w are **equal**, and write $z = w$, if and only if $a = c$ and $b = d$.

The arithmetic of complex numbers is as you would expect. The only things you need to remember are the two properties in Definition 52. The next example should help recall how these animals behave.

Example 141 Arithmetic with complex numbers
Perform the indicated operations.

1. $(1 - 2i) - (3 + 4i)$
2. $(1 - 2i)(3 + 4i)$
3. $\dfrac{1 - 2i}{3 - 4i}$
4. $\sqrt{-3}\sqrt{-12}$
5. $\sqrt{(-3)(-12)}$
6. $(x - [1 + 2i])(x - [1 - 2i])$

Solution

1. As mentioned earlier, we treat expressions involving i as we would any other radical. We distribute and combine like terms:

$$\begin{aligned}(1 - 2i) - (3 + 4i) &= 1 - 2i - 3 - 4i && \text{Distribute}\\ &= -2 - 6i && \text{Gather like terms}\end{aligned}$$

   Technically, we'd have to rewrite our answer $-2-6i$ as $(-2)+(-6)i$ to be (in the strictest sense) 'in the form $a + bi$'. That being said, even pedants have their limits, and we'll consider $-2 - 6i$ good enough.

2. Using the Distributive Property (a.k.a. F.O.I.L.), we get

$$\begin{aligned}(1 - 2i)(3 + 4i) &= (1)(3) + (1)(4i) - (2i)(3) - (2i)(4i) && \text{F.O.I.L.}\\ &= 3 + 4i - 6i - 8i^2 \\ &= 3 - 2i - 8(-1) && i^2 = -1\\ &= 3 - 2i + 8 \\ &= 11 - 2i\end{aligned}$$

3. How in the world are we supposed to simplify $\frac{1-2i}{3-4i}$? Well, we deal with the denominator $3 - 4i$ as we would any other denominator containing two terms, one of which is a square root: we and multiply both numerator and denominator by $3 + 4i$, the (complex) conjugate of $3 - 4i$. Doing so

produces

$$\begin{aligned}
\frac{1-2i}{3-4i} &= \frac{(1-2i)(3+4i)}{(3-4i)(3+4i)} && \text{Equivalent Fractions}\\
&= \frac{3+4i-6i-8i^2}{9-16i^2} && \text{F.O.I.L.}\\
&= \frac{3-2i-8(-1)}{9-16(-1)} && i^2=-1\\
&= \frac{11-2i}{25}\\
&= \frac{11}{25}-\frac{2}{25}i
\end{aligned}$$

4. We use property 2 of Definition 52 first, then apply the rules of radicals applicable to real numbers to get $\sqrt{-3}\sqrt{-12} = \left(i\sqrt{3}\right)\left(i\sqrt{12}\right) = i^2\sqrt{3\cdot 12} = -\sqrt{36} = -6$.

5. We adhere to the order of operations here and perform the multiplication before the radical to get $\sqrt{(-3)(-12)} = \sqrt{36} = 6$.

6. We can brute force multiply using the distributive property and see that

$$\begin{aligned}
(x-[1+2i])(x-[1-2i]) &= x^2 - x[1-2i] - x[1+2i] + [1-2i][1+2i]\\
&= x^2 - x + 2ix - x - 2ix + 1 - 2i + 2i - 4i^2\\
&= x^2 - 2x + 1 - 4(-1)\\
&= x^2 - 2x + 5
\end{aligned}$$

In the previous example, we used the idea of a 'conjugate' to divide two complex numbers. (You may recall using conjugates to rationalize expressions involving square roots.) More generally, the **complex conjugate** of a complex number $a+bi$ is the number $a-bi$. The notation commonly used for complex conjugation is a 'bar': $\overline{a+bi} = a-bi$. For example, $\overline{3+2i} = 3-2i$ and $\overline{3-2i} = 3+2i$. To find $\overline{6}$, we note that $\overline{6} = \overline{6+0i} = 6-0i = 6$, so $\overline{6} = 6$. Similarly, $\overline{4i} = -4i$, since $\overline{4i} = \overline{0+4i} = 0-4i = -4i$. Note that $\overline{3+\sqrt{5}} = 3+\sqrt{5}$, not $3-\sqrt{5}$, since $\overline{3+\sqrt{5}} = \overline{3+\sqrt{5}+0i} = 3+\sqrt{5}-0i = 3+\sqrt{5}$. Here, the conjugation specified by the 'bar' notation involves reversing the sign before $i=\sqrt{-1}$, not before $\sqrt{5}$. The properties of the conjugate are summarized in the following theorem.

Theorem 28 Properties of the Complex Conjugate

Let z and w be complex numbers.

- $\overline{\overline{z}} = z$
- $\overline{z+w} = \overline{z} + \overline{w}$
- $\overline{zw} = \overline{z}\,\overline{w}$
- $\overline{z^n} = \left(\overline{z}\right)^n$, for any natural number n
- z is a real number if and only if $\overline{z} = z$.

Essentially, Theorem 28 says that complex conjugation works well with addition, multiplication and powers. The proofs of these properties can best be achieved by writing out $z = a + bi$ and $w = c + di$ for real numbers a, b, c and d. Next, we compute the left and right sides of each equation and verify that they are the same.

The proof of the first property is a very quick exercise. To prove the second property, we compare $\overline{z+w}$ with $\overline{z} + \overline{w}$. We have $\overline{z} + \overline{w} = \overline{a+bi} + \overline{c+di} = a - bi + c - di$. To find $\overline{z+w}$, we first compute

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

so

$$\overline{z+w} = \overline{(a+c)+(b+d)i} = (a + c) - (b + d)i = a - bi + c - di = \overline{z} + \overline{w}$$

As such, we have established $\overline{z+w} = \overline{z} + \overline{w}$. The proof for multiplication works similarly. The proof that the conjugate works well with powers can be viewed as a repeated application of the product rule. A rigorous proof requires a technique called Mathematical Induction, which is beyond the scope of this course, so we omit the proof. The last property is a characterization of real numbers. If z is real, then $z = a + 0i$, so $\overline{z} = a - 0i = a = z$. On the other hand, if $z = \overline{z}$, then $a + bi = a - bi$ which means $b = -b$ so $b = 0$. Hence, $z = a + 0i = a$ and is real.

Proof by Mathematical Induction usually isn't encountered until Math 2000. It provides a way of formally proving statements that are claimed to hold true for all natural numbers.

The formal definition of a **field** is usually not encountered until a second course in abstract algebra, such as Math 4500. For now, you should think of a field as any number system where the rules of arithmetic behave exactly as you expect them to.

It is worth noting that although the arithmetic of complex numbers seems, at first impression, to be very different and strange compared to the arithmetic of real numbers, it actually satisfies all the same properties. (Some of which we implicitly assumed in the previous example.) In fact, both $\mathbb{R}$ and $\mathbb{C}$ are examples of what are called **fields**.

Theorem 29 Properties of Complex Arithmetic

The addition and multiplication of complex numbers satisfy the following properties:

- **Closure under addition:** For any complex numbers z and w, $z + w$ is a complex number.
- **Commutativity of addition:** For any complex numbers z and w, $z + w = w + z$.
- **Associativity of addition:** For any complex numbers z_1, z_2, z_3, $z_1 + (z_2 + z_3) = (z_1 + z_2) + z_3$.
- **Additive identity:** There exists a complex number 0 such that $z + 0 = 0 + z = z$ for every complex number z.
- **Additive inverses:** For every complex number z there exists a complex number $-z$ such that $z + (-z) = -z + z = 0$.
- **Closure under multiplication:** For any complex numbers z and w, zw is a complex number.
- **Commutativity of multiplication:** For any complex numbers z and w, $zw = wz$.
- **Assiciativity of multiplication:** For any complex numbers z_1, z_2, z_3, $z_1(z_2z_3) = (z_1z_2)z_3$
- **Multiplicative identity:** There exists a complex number 1 such that $1 \cdot z = z \cdot 1 = z$ for every complex number z.
- **Multiplicative inverses:** For every complex number $z \neq 0$, there exists a complex number z^{-1} such that $zz^{-1} = z^{-1}z = 1$.
- **Distributive property:** For all complex numbers z_1, z_2, z_3, we have $z_1(z_2 + z_3) = z_1z_2 + z_1z_3$.

We leave the proof of Theorem 29 as a long (but straightforward) exercises. Working through the proof is a good way to confirm for yourself that you understand the corresponding rules for real number arithmetic from Section 1.2, and how the properties for complex arithmetic are inherited from their real counterparts.

We now consider the problem of solving quadratic equations. Consider $x^2 - 2x + 5 = 0$. The discriminant $b^2 - 4ac = -16$ is negative, so we know from the quadratic formula that there are no *real* solutions, since the Quadratic Formula would involve the term $\sqrt{-16}$. Complex numbers, however, are built just for such situations, so we can go ahead and apply the Quadratic Formula to get:

$$x = \frac{-(-2) \pm \sqrt{(-2)^2 - 4(1)(5)}}{2(1)} = \frac{2 \pm \sqrt{-16}}{2} = \frac{2 \pm 4i}{2} = 1 \pm 2i.$$

We're assuming some prior familiarity on the part of the reader where quadratic equations are concerned. If you're a bit rusty when it comes to finding *real* solutions to quadratic equations (and in particular, the quadratic formula), you may want to check out the review materials available on the "Math Basics" Moodle page.

Example 142 Finding complex solutions
Find the complex solutions to the following equations.

Remember, all real numbers are complex numbers, so 'complex solutions' means both real and non-real answers.

1. $\dfrac{2x}{x+1} = x+3$ 2. $2t^4 = 9t^2 + 5$ 3. $z^3 + 1 = 0$

Solution

1. Clearing fractions yields a quadratic equation so we collect all terms on one side and apply the Quadratic Formula.

$$\begin{aligned} \frac{2x}{x+1} &= x+3 && \\ 2x &= (x+3)(x+1) && \text{Clear denominators} \\ 2x &= x^2 + x + 3x + 3 && \text{F.O.I.L.} \\ 2x &= x^2 + 4x + 3 && \text{Gather like terms} \\ 0 &= x^2 + 2x + 3 && \text{Subtract } 2x \end{aligned}$$

From here, we apply the Quadratic Formula

$$\begin{aligned} x &= \frac{-2 \pm \sqrt{2^2 - 4(1)(3)}}{2(1)} && \text{Quadratic Formula} \\ &= \frac{-2 \pm \sqrt{-8}}{2} && \text{Simplify} \\ &= \frac{-2 \pm i\sqrt{8}}{2} && \text{Definition of } i \\ &= \frac{-2 \pm i2\sqrt{2}}{2} && \text{Product Rule for Radicals} \\ &= \frac{\cancel{2}(-1 \pm i\sqrt{2})}{\cancel{2}} && \text{Factor and reduce} \\ &= -1 \pm i\sqrt{2} \end{aligned}$$

We get two answers: $x = -1 + i\sqrt{2}$ and its conjugate $x = -1 - i\sqrt{2}$. Checking both of these answers reviews all of the salient points about complex number arithmetic and is therefore strongly encouraged.

2. Since we have three terms, and the exponent on one term ('4' on t^4) is exactly twice the exponent on the other ('2' on t^2), we have a Quadratic in Disguise. We proceed accordingly.

$$\begin{aligned} 2t^4 &= 9t^2 + 5 && \\ 2t^4 - 9t^2 - 5 &= 0 && \text{Subtract } 9t^2 \text{ and } 5 \\ (2t^2 + 1)(t^2 - 5) &= 0 && \text{Factor} \\ 2t^2 + 1 = 0 \quad &\text{or} \quad t^2 = 5 && \text{Zero Product Property} \end{aligned}$$

From $2t^2 + 1 = 0$ we get $2t^2 = -1$, or $t^2 = -\frac{1}{2}$. We extract square roots as follows:

$$t = \pm\sqrt{-\frac{1}{2}} = \pm i\sqrt{\frac{1}{2}} = \pm i\frac{\sqrt{1}}{\sqrt{2}} = \pm i\frac{1}{\sqrt{2}} = \pm\frac{i\sqrt{2}}{2},$$

where we have rationalized the denominator per convention. From $t^2 = 5$, we get $t = \pm\sqrt{5}$. In total, we have four complex solutions - two real: $t = \pm\sqrt{5}$ and two non-real: $t = \pm\frac{i\sqrt{2}}{2}$.

3. To find the *real* solutions to $z^3 + 1 = 0$, we can subtract the 1 from both sides and extract cube roots: $z^3 = -1$, so $z = \sqrt[3]{-1} = -1$. It turns out there are two more non-real complex number solutions to this equation. To get at these, we factor:

$$\begin{aligned} z^3 + 1 &= 0 \\ (z+1)(z^2 - z + 1) &= 0 && \text{Factor (Sum of Two Cubes)} \\ z + 1 = 0 \quad &\text{or} \quad z^2 - z + 1 = 0 \end{aligned}$$

From $z + 1 = 0$, we get our real solution $z = -1$. From $z^2 - z + 1 = 0$, we apply the Quadratic Formula to get:

$$z = \frac{-(-1) \pm \sqrt{(-1)^2 - 4(1)(1)}}{2(1)} = \frac{1 \pm \sqrt{-3}}{2} = \frac{1 \pm i\sqrt{3}}{2}$$

Thus we get *three* solutions to $z^3 + 1 = 0$ - one real: $z = -1$ and two non-real: $z = \frac{1 \pm i\sqrt{3}}{2}$. As always, the reader is encouraged to test their algebraic mettle and check these solutions.

It is no coincidence that the non-real solutions to the equations in Example 141 appear in complex conjugate pairs. Any time we use the Quadratic Formula to solve an equation with real coefficients, the answers will form a complex conjugate pair owing to the $\pm$ in the Quadratic Formula. This is stated formally in the following theorem.

Theorem 30 Discriminant Theorem

Given a Quadratic Equation $AX^2 + BX + C = 0$, where A, B and C are real numbers, let $D = B^2 - 4AC$ be the discriminant.

- If $D > 0$, there are two distinct real number solutions to the equation.
- If $D = 0$, there is one (repeated) real number solution.

  Note: 'Repeated' here comes from the fact that 'both' solutions $\frac{-B \pm 0}{2A}$ reduce to $-\frac{B}{2A}$.
- If $D < 0$, there are two non-real solutions which form a complex conjugate pair.

Theorem 30 tells us that if ever we obtain non-real zeros to a quadratic function with real coefficients, the zeros will be a complex conjugate pair. (Do you see why?) Next, we note that in Example 140, part 6, we found $(x - [1 + 2i])(x - [1 - 2i]) = x^2 - 2x + 5$. This demonstrates that the factor theorem holds even for non-real zeros, i.e, $x = 1 + 2i$ is a zero of f, and, sure enough, $(x - [1 + 2i])$ is a factor of $f(x)$. It turns out that polynomial division works the same way for all complex numbers, real and non-real alike, so the Factor and Remainder Theorems hold as well. But how do we know if a general polynomial has any complex zeros at all? We have many examples of polynomials with no real zeros. Can there be polynomials with no zeros whatsoever? The answer to that last question is "No." and the theorem which provides that answer is The Fundamental Theorem of Algebra.

Theorem 31 The Fundamental Theorem of Algebra

Suppose f is a polynomial function with complex number coefficients of degree $n \geq 1$, then f has at least one complex zero.

The Fundamental Theorem of Algebra has since been proved many times, using many different methods, by many mathematicians. There are probably very few, if any, results in mathematics with the variety of proofs this result has. Unfortunately, none of the proofs can be understood within the realm of this text, but if the reader is sufficiently interested, a collection of proofs can be found at this website.

The Fundamental Theorem of Algebra is an example of an 'existence' theorem in Mathematics. It guarantees the existence of at least one zero, but gives us no algorithm to use in finding it. The authors are fully aware that the full impact and profound nature of the Fundamental Theorem of Algebra is lost on most students, and that's fine. It took mathematicians literally hundreds of years to prove the theorem in its full generality, and some of that history is recorded in this Wikipedia article. Note that the Fundamental Theorem of Algebra applies to not only polynomial functions with real coefficients, but to those with complex number coefficients as well.

Suppose f is a polynomial of degree $n \geq 1$. The Fundamental Theorem of Algebra guarantees us at least one complex zero, z_1, and as such, the Factor Theorem guarantees that $f(x)$ factors as $f(x) = (x - z_1)\, q_1(x)$ for a polynomial function q_1, of degree exactly $n - 1$. If $n - 1 \geq 1$, then the Fundamental Theorem of Algebra guarantees a complex zero of q_1 as well, say z_2, so then the Factor Theorem gives us $q_1(x) = (x - z_2)\, q_2(x)$, and hence $f(x) = (x - z_1)(x - z_2)\, q_2(x)$. We can continue this process exactly n times, at which point our quotient polynomial q_n has degree 0 so it's a constant. This argument gives us the following factorization theorem.

Theorem 32 Complex Factorization Theorem

Suppose f is a polynomial function with complex number coefficients. If the degree of f is n and $n \geq 1$, then f has exactly n complex zeros, counting multiplicity. If z_1, z_2, ..., z_k are the distinct zeros of f, with multiplicities m_1, m_2, ..., m_k, respectively, then $f(x) = a\,(x - z_1)^{m_1}(x - z_2)^{m_2} \cdots (x - z_k)^{m_k}$.

To complete our study of the arithmetic of complex numbers, we should discuss powers and roots. Computing powers can be done using the form $z = x + iy$, but it quickly becomes unpleasant (try computing $(4 + 3i)^7$, for example). Roots, on the other hand, are nearly impossible. Luckily for us, there is a better way: using the *polar form* of complex numbers. Before we get to this discussion, however, we need to pause to introduce the *polar coordinate system* for the Cartesian coordinate system.

Exercises 7.1

Problems

In Exercises 1 – 10, use the given complex numbers z and w to find and simplify the following:

- $z + w$
- zw
- z^2
- $\frac{1}{z}$
- $\frac{z}{w}$
- $\frac{w}{z}$
- $\overline{z}$
- $z\overline{z}$
- $(\overline{z})^2$

1. $z = 2 + 3i$, $w = 4i$
2. $z = 1 + i$, $w = -i$
3. $z = i$, $w = -1 + 2i$
4. $z = 4i$, $w = 2 - 2i$
5. $z = 3 - 5i$, $w = 2 + 7i$
6. $z = -5 + i$, $w = 4 + 2i$
7. $z = \sqrt{2} - i\sqrt{2}$, $w = \sqrt{2} + i\sqrt{2}$
8. $z = 1 - i\sqrt{3}$, $w = -1 - i\sqrt{3}$
9. $z = \frac{1}{2} + \frac{\sqrt{3}}{2}i$, $w = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$
10. $z = -\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i$, $w = -\frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i$

In Exercises 11 – 18, simplify the quantity.

11. $\sqrt{-49}$
12. $\sqrt{-9}$
13. $\sqrt{-25}\sqrt{-4}$
14. $\sqrt{(-25)(-4)}$
15. $\sqrt{-9}\sqrt{-16}$
16. $\sqrt{(-9)(-16)}$
17. $\sqrt{-(-9)}$
18. $-\sqrt{(-9)}$

We know that $i^2 = -1$ which means $i^3 = i^2 \cdot i = (-1) \cdot i = -i$ and $i^4 = i^2 \cdot i^2 = (-1)(-1) = 1$. In Exercises 19 – 26, use this information to simplify the given power of i.

19. i^5
20. i^6
21. i^7
22. i^8
23. i^{15}
24. i^{26}
25. i^{117}
26. i^{304}

In Exercises 27 – 35, find all complex solutions.

27. $3x^2 + 6 = 4x$
28. $15t^2 + 2t + 5 = 3t(t^2 + 1)$
29. $3y^2 + 4 = y^4$
30. $\frac{2}{1 - w} = w$
31. $\frac{y}{3} - \frac{3}{y} = y$
32. $\frac{x^3}{2x - 1} = \frac{x}{3}$
33. $x = \frac{2}{\sqrt{5 - x}}$
34. $\frac{5y^4 + 1}{y^2 - 1} = 3y^2$
35. $z^4 = 16$
36. Multiply and simplify: $\left(x - [3 - i\sqrt{23}]\right)\left(x - [3 + i\sqrt{23}]\right)$

7.2 Polar Coordinates

In Section 1.3, we introduced the Cartesian coordinates of a point in the plane as a means of assigning ordered pairs of numbers to points in the plane. We defined the Cartesian coordinate plane using two number lines – one horizontal and one vertical – which intersect at right angles at a point we called the 'origin'. To plot a point, say $P(-3,4)$, we start at the origin, travel horizontally to the left 3 units, then up 4 units. Alternatively, we could start at the origin, travel up 4 units, then to the left 3 units and arrive at the same location. For the most part, the 'motions' of the Cartesian system (over and up) describe a rectangle, and most points can be thought of as the corner diagonally across the rectangle from the origin.(Excluding, of course, the points in which one or both coordinates are 0.) For this reason, the Cartesian coordinates of a point are often called 'rectangular' coordinates. In this section, we introduce a new system for assigning coordinates to points in the plane – **polar coordinates**. We start with an origin point, called the **pole**, and a ray called the **polar axis**. We then locate a point *P* using two coordinates, (r,θ), where *r* represents a *directed* distance from the pole (we will explain more about this momentarily) and θ is a measure of rotation from the polar axis. Roughly speaking, the polar coordinates (r,θ) of a point measure 'how far out' the point is from the pole (that's *r*), and 'how far to rotate' from the polar axis, (that's θ).

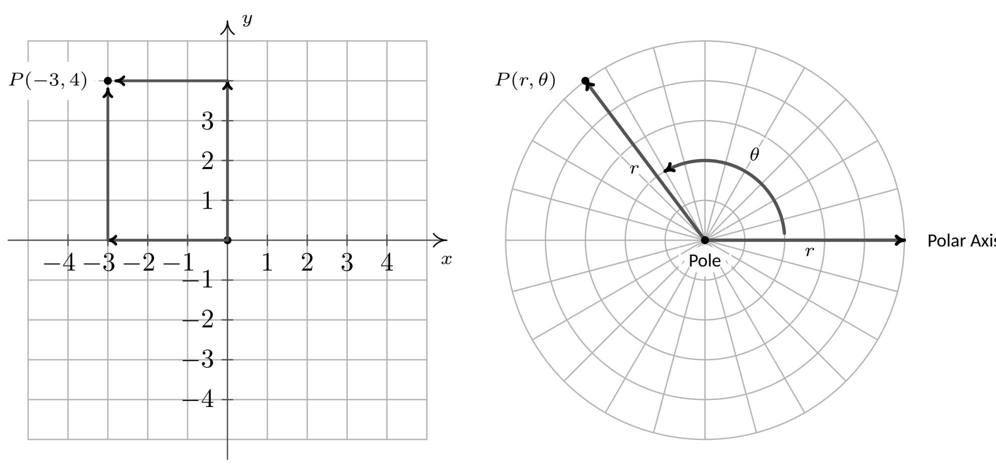


Figure 7.1: Rectangular vs. Polar Coordinates

For example, if we wished to plot the point *P* with polar coordinates $\left(4, \frac{5\pi}{6}\right)$, we'd start at the pole, move out along the polar axis 4 units, then rotate $\frac{5\pi}{6}$ radians counter-clockwise, as shown in Figure 7.2.

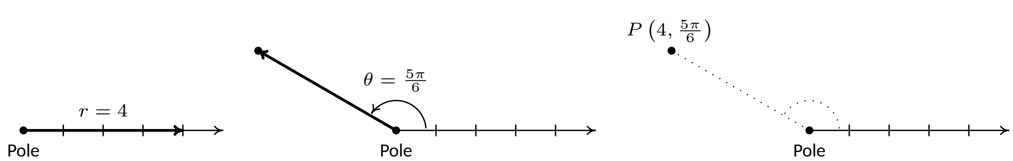


Figure 7.2: Locating a point using polar coordinates

We may also visualize this process by thinking of the rotation first.(As with anything in Mathematics, the more ways you have to look at something, the better. The authors encourage the reader to take time to think about both approaches to plotting points given in polar coordinates.) To plot $P\left(4, \frac{5\pi}{6}\right)$ this way, we rotate $\frac{5\pi}{6}$ counter-clockwise from the polar axis, then move outwards from the pole 4 units, as shown in Figure 7.3. Essentially we are locating a point on the terminal side of $\frac{5\pi}{6}$ which is 4 units away from the pole.

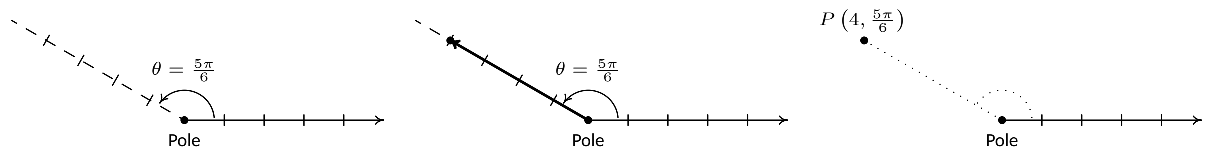


Figure 7.3: Performing the rotation first

If $r < 0$, we begin by moving in the opposite direction on the polar axis from the pole. For example, to plot $Q\left(-3.5, \frac{\pi}{4}\right)$ we have the steps shown in Figure 7.4.

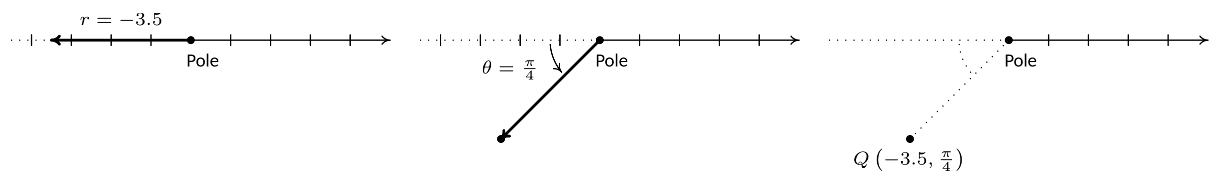


Figure 7.4: Using polar coordinates when $r < 0$

If we interpret the angle first, we rotate $\frac{\pi}{4}$ radians, then move back through the pole 3.5 units. Here we are locating a point 3.5 units away from the pole on the terminal side of $\frac{5\pi}{4}$, not $\frac{\pi}{4}$.

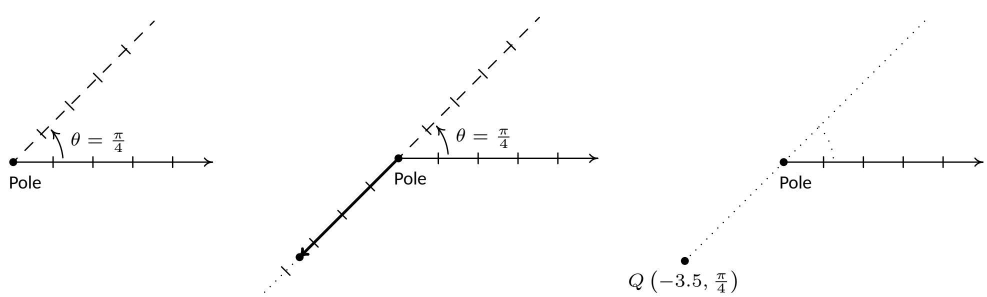


Figure 7.5: Performing the rotation first to plot the point in Figure 7.4

As you may have guessed, $\theta < 0$ means the rotation away from the polar axis is clockwise instead of counter-clockwise. Hence, to plot $R\left(3.5, -\frac{3\pi}{4}\right)$ we have the following.

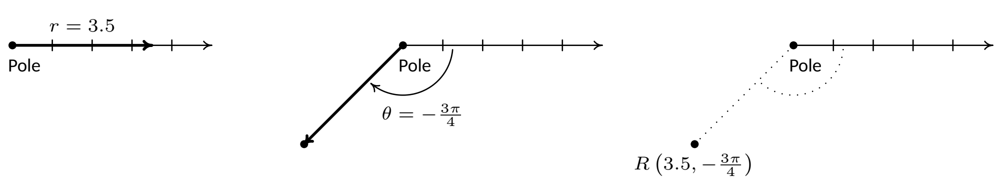


Figure 7.6: $\theta = -\frac{3\pi}{4} < 0$ produces a clockwise rotation

From an 'angles first' approach, we rotate $-\frac{3\pi}{4}$ then move out 3.5 units from the pole. We see that R is the point on the terminal side of $\theta = -\frac{3\pi}{4}$ which is 3.5 units from the pole.

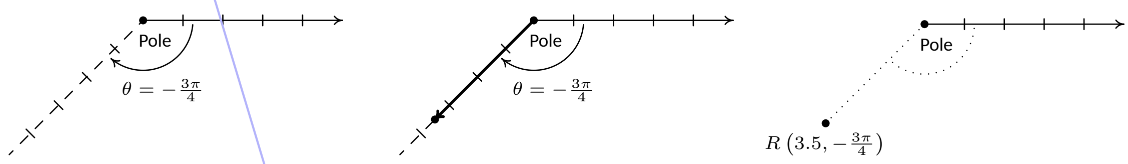


Figure 7.7: Rotating first with $\theta < 0$

The points Q and R above are, in fact, the same point despite the fact that their polar coordinate representations are different. Unlike Cartesian coordinates where (a, b) and (c, d) represent the same point if and only if $a = c$ and $b = d$, a point can be represented by infinitely many polar coordinate pairs. We explore this notion more in the following example.

Example 143 **Plotting points in polar coordinates**
For each point in polar coordinates given below plot the point and then give two additional expressions for the point, one of which has $r > 0$ and the other with $r < 0$.

1. $P\left(2, 240^\circ\right)$
2. $P\left(-4, \frac{7\pi}{6}\right)$
3. $P\left(117, -\frac{5\pi}{2}\right)$
4. $P\left(-3, -\frac{\pi}{4}\right)$

SOLUTION

1. Whether we move 2 units along the polar axis and then rotate 240° or rotate 240° then move out 2 units from the pole, we plot $P\left(2, 240^\circ\right)$ in Figure 7.8 below.

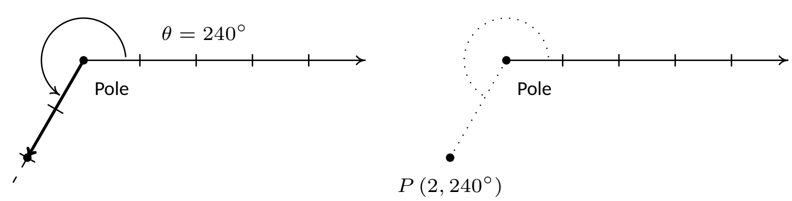


Figure 7.8: Plotting $P(2, 240^\circ)$

We now set about finding alternate descriptions (r, θ) for the point P. Since P is 2 units from the pole, $r = \pm 2$. Next, we choose angles θ for each of the r values. The given representation for P is $(2, 240^\circ)$ so the angle θ we choose for the $r = 2$ case must be coterminal with 240°. (Can you see why?) One such angle is $\theta = -120^\circ$ so one answer for this case is $(2, -120^\circ)$. For the case $r = -2$, we visualize our rotation starting 2 units to the left of the pole. From this position, we need only to rotate $\theta = 60^\circ$ to arrive at location coterminal with 240°. Hence, our answer here is $(-2, 60^\circ)$. We check our answers by plotting them in Figure 7.9.

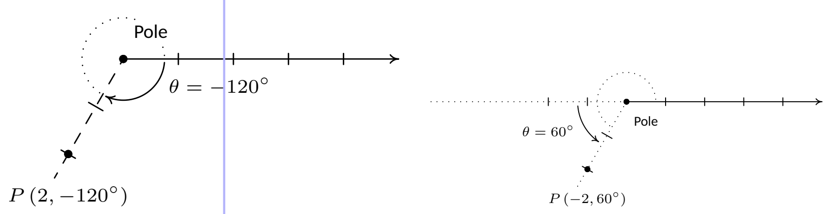


Figure 7.9: Alternate polar reprentations of $P(2, 240^\circ)$

2. We plot $\left(-4, \frac{7\pi}{6}\right)$ by first moving 4 units to the left of the pole and then rotating $\frac{7\pi}{6}$ radians. Since $r = -4 < 0$, we find our point lies 4 units from the pole on the terminal side of $\frac{\pi}{6}$.

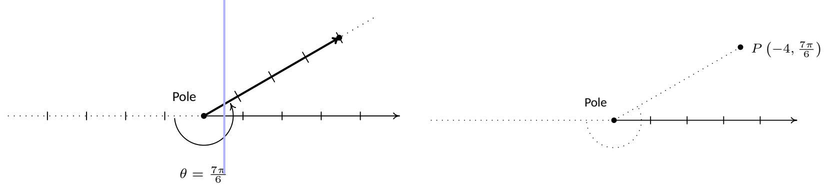


Figure 7.10: Plotting $P(-4, \frac{7\pi}{6})$

To find alternate descriptions for P, we note that the distance from P to the pole is 4 units, so any representation (r, θ) for P must have $r = \pm 4$. As we noted above, P lies on the terminal side of $\frac{\pi}{6}$, so this, coupled with $r = 4$, gives us $\left(4, \frac{\pi}{6}\right)$ as one of our answers. To find a different representation for P with $r = -4$, we may choose any angle coterminal with the angle in the original representation of $P\left(-4, \frac{7\pi}{6}\right)$. We pick $-\frac{5\pi}{6}$ and get $\left(-4, -\frac{5\pi}{6}\right)$ as our second answer.

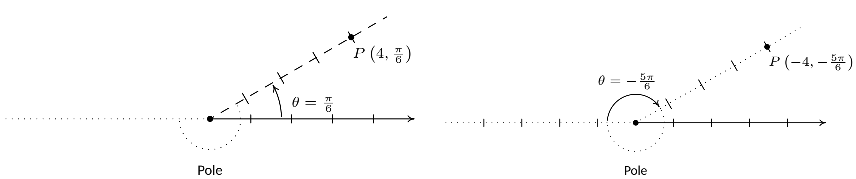


Figure 7.11: Alternate polar representations of $P(-4, \frac{7\pi}{6})$

3. To plot $P\left(117, -\frac{5\pi}{2}\right)$, we move along the polar axis 117 units from the pole and rotate *clockwise* $\frac{5\pi}{2}$ radians as illustrated in Figure 7.12 below.

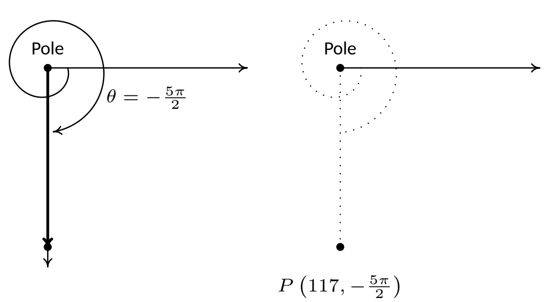


Figure 7.12: Plotting $P(117, -\frac{5\pi}{2})$

Since P is 117 units from the pole, any representation (r, θ) for P satisfies $r = \pm 117$. For the $r = 117$ case, we can take θ to be any angle coterminal with $-\frac{5\pi}{2}$. In this case, we choose $\theta = \frac{3\pi}{2}$, and get $\left(117, \frac{3\pi}{2}\right)$ as one answer. For the $r = -117$ case, we visualize moving left 117 units from the pole and then rotating through an angle θ to reach P. We find that $\theta = \frac{\pi}{2}$ satisfies this requirement, so our second answer is $\left(-117, \frac{\pi}{2}\right)$.

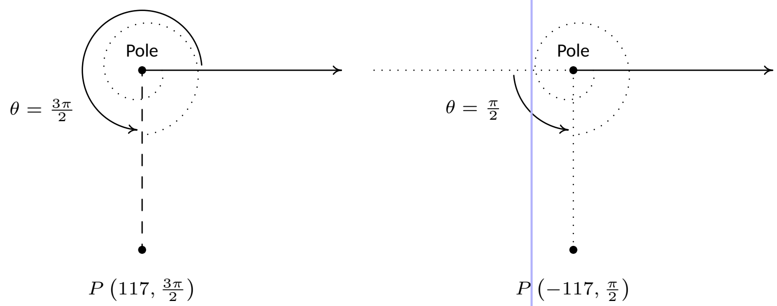


Figure 7.13: Alternate polar representations of $P(117, -\frac{5\pi}{2})$

4. We move three units to the left of the pole and follow up with a clockwise rotation of $\frac{\pi}{4}$ radians to plot $P\left(-3, -\frac{\pi}{4}\right)$. We see that P lies on the terminal side of $\frac{3\pi}{4}$.

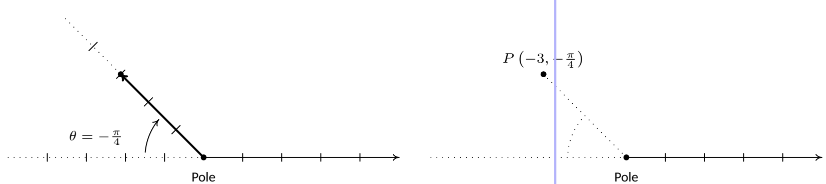


Figure 7.14: Plotting $P(-3, -\frac{\pi}{4})$

Since P lies on the terminal side of $\frac{3\pi}{4}$, one alternative representation for P is $\left(3, \frac{3\pi}{4}\right)$. To find a different representation for P with $r = -3$, we may choose any angle coterminal with $-\frac{\pi}{4}$. We choose $\theta = \frac{7\pi}{4}$ for our final answer $\left(-3, \frac{7\pi}{4}\right)$.

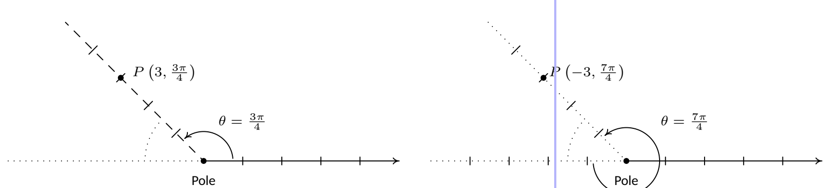


Figure 7.15: Alternate polar representations of $P(-3, -\frac{\pi}{4})$

Now that we have had some practice with plotting points in polar coordinates, it should come as no surprise that any given point expressed in polar coordinates has infinitely many other representations in polar coordinates. The following result characterizes when two sets of polar coordinates determine the same point in the plane. It could be considered as a definition or a theorem, depending on your point of view. We state it as a property of the polar coordinate system.

Key Idea 24 Equivalent Representations of Points in Polar Coordinates

Suppose (r, θ) and (r', θ') are polar coordinates where $r \neq 0$, $r' \neq 0$ and the angles are measured in radians. Then (r, θ) and (r', θ') determine the same point P if and only if one of the following is true:

- $r' = r$ and $\theta' = \theta + 2\pi k$ for some integer k
- $r' = -r$ and $\theta' = \theta + (2k+1)\pi$ for some integer k

All polar coordinates of the form $(0, \theta)$ represent the pole regardless of the value of θ.

The key to understanding this result, and indeed the whole polar coordinate system, is to keep in mind that

$$(r, \theta) \text{ means } (\text{directed distance from pole}, \text{angle of rotation}).$$

If $r = 0$, then no matter how much rotation is performed, the point never leaves the pole. Thus $(0, \theta)$ is the pole for all values of θ. Now let's assume that neither r nor r' is zero. If (r, θ) and (r', θ') determine the same point P then the (non-zero) distance from P to the pole in each case must be the same. Since this distance is controlled by the first coordinate, we have that either $r' = r$ or $r' = -r$. If $r' = r$, then when plotting (r, θ) and (r', θ'), the angles θ and θ' have the same initial side. Hence, if (r, θ) and (r', θ') determine the same point, we must have that θ' is coterminal with θ. We know that this means $\theta' = \theta + 2\pi k$ for some integer k, as required. If, on the other hand, $r' = -r$, then when plotting (r, θ) and (r', θ'), the initial side of θ' is rotated π radians away from the initial side of θ. In this case, θ' must be coterminal with $\pi + \theta$. Hence, $\theta' = \pi + \theta + 2\pi k$ which we rewrite as $\theta' = \theta + (2k+1)\pi$ for some integer k. Conversely, if $r' = r$ and $\theta' = \theta + 2\pi k$ for some integer k, then the points $P(r, \theta)$ and $P'(r', \theta')$ lie the same (directed) distance from the pole on the terminal sides of coterminal angles, and hence are the same point. Now suppose $r' = -r$ and $\theta' = \theta + (2k+1)\pi$ for some integer k. To plot P, we first move a directed distance r from the pole; to plot P', our first step is to move the same distance from the pole as P, but in the opposite direction. At this intermediate stage, we have two points equidistant from the pole rotated exactly π radians apart. Since $\theta' = \theta + (2k+1)\pi = (\theta + \pi) + 2\pi k$ for some integer k, we see that θ' is coterminal to $(\theta + \pi)$ and it is this extra π radians of rotation which aligns the points P and P'.

Next, we marry the polar coordinate system with the Cartesian (rectangular) coordinate system. To do so, we identify the pole and polar axis in the polar system to the origin and positive x-axis, respectively, in the rectangular system. We get the following result.

Theorem 33 Conversion Between Rectangular and Polar Coordinates

Suppose P is represented in rectangular coordinates as (x, y) and in polar coordinates as (r, θ). Then

- $x = r\cos(\theta)$ and $y = r\sin(\theta)$
- $x^2 + y^2 = r^2$ and $\tan(\theta) = \dfrac{y}{x}$ (provided $x \neq 0$)

In the case $r > 0$, Theorem 33 is an immediate consequence of the trigonometric definitions of sine and cosine along with the quotient identity $\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}$. If $r < 0$, then we know an alternate representation for (r, θ) is $(-r, \theta + \pi)$. Since $\cos(\theta + \pi) = -\cos(\theta)$ and $\sin(\theta + \pi) = -\sin(\theta)$, applying the theorem to $(-r, \theta + \pi)$ gives $x = (-r)\cos(\theta + \pi) = (-r)(-\cos(\theta)) = r\cos(\theta)$ and $y = (-r)\sin(\theta + \pi) = (-r)(-\sin(\theta)) = r\sin(\theta)$. Moreover, $x^2 + y^2 = (-r)^2 = r^2$, and $\frac{y}{x} = \tan(\theta + \pi) = \tan(\theta)$, so the theorem is true in this case, too. The remaining case is $r = 0$, in which case $(r, \theta) = (0, \theta)$ is the pole. Since the pole is identified with the origin $(0, 0)$ in rectangular coordinates, the theorem in this case amounts to checking '$0 = 0$.' The following example puts Theorem 33 to good use.

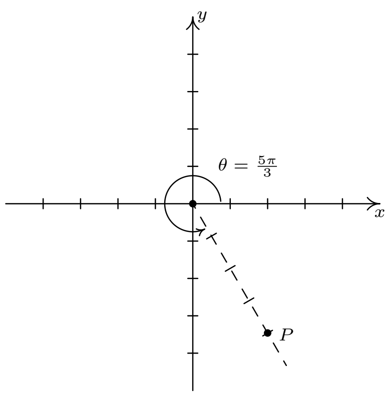


Figure 7.16: P has rectangular coordinates $(2, -2\sqrt{3})$ and polar coordinates $(4, \frac{5\pi}{3})$

Example 144 Converting from rectangular to polar coordinates

Convert each point in rectangular coordinates given below into polar coordinates with $r \geq 0$ and $0 \leq \theta < 2\pi$. Use exact values if possible and round any approximate values to two decimal places. Check your answer by converting them back to rectangular coordinates.

1. $P\left(2, -2\sqrt{3}\right)$
2. $Q(-3, -3)$
3. $R(0, -3)$
4. $S(-3, 4)$

Solution

1. Even though we are not explicitly told to do so, we can avoid many common mistakes by taking the time to plot the points before we do any calculations. Plotting $P\left(2, -2\sqrt{3}\right)$ shows that it lies in Quadrant IV. With $x = 2$ and $y = -2\sqrt{3}$, we get $r^2 = x^2 + y^2 = (2)^2 + \left(-2\sqrt{3}\right)^2 = 4 + 12 = 16$ so $r = \pm 4$. Since we are asked for $r \geq 0$, we choose $r = 4$. To find θ, we have that $\tan(\theta) = \frac{y}{x} = \frac{-2\sqrt{3}}{2} = -\sqrt{3}$. This tells us θ has a reference angle of $\frac{\pi}{3}$, and since P lies in Quadrant IV, we know θ is a Quadrant IV angle. We are asked to have $0 \leq \theta < 2\pi$, so we choose $\theta = \frac{5\pi}{3}$. Hence, our answer is $\left(4, \frac{5\pi}{3}\right)$. To check, we convert $(r, \theta) = \left(4, \frac{5\pi}{3}\right)$ back to rectangular coordinates and we find $x = r\cos(\theta) = 4\cos\left(\frac{5\pi}{3}\right) = 4\left(\frac{1}{2}\right) = 2$ and $y = r\sin(\theta) = 4\sin\left(\frac{5\pi}{3}\right) = 4\left(-\frac{\sqrt{3}}{2}\right) = -2\sqrt{3}$, as required.
2. The point $Q(-3, -3)$ lies in Quadrant III. Using $x = y = -3$, we get $r^2 = (-3)^2 + (-3)^2 = 18$ so $r = \pm\sqrt{18} = \pm 3\sqrt{2}$. Since we are asked for $r \geq 0$, we choose $r = 3\sqrt{2}$. We find $\tan(\theta) = \frac{-3}{-3} = 1$, which means θ has a reference angle of $\frac{\pi}{4}$. Since Q lies in Quadrant III, we choose $\theta = \frac{5\pi}{4}$, which satisfies the requirement that $0 \leq \theta < 2\pi$. Our final answer is $(r, \theta) = \left(3\sqrt{2}, \frac{5\pi}{4}\right)$. To check, we find $x = r\cos(\theta) = (3\sqrt{2})\cos\left(\frac{5\pi}{4}\right) =$

$(3\sqrt{2})\left(-\frac{\sqrt{2}}{2}\right) = -3$ and $y = r\sin(\theta) = (3\sqrt{2})\sin\left(\frac{5\pi}{4}\right) = (3\sqrt{2})\left(-\frac{\sqrt{2}}{2}\right) = -3$, so we are done.

3. The point $R(0,-3)$ lies along the negative y-axis. While we could go through the usual computations to find the polar form of R (since $x = 0$, we would have to determine θ geometrically), in this case we can find the polar coordinates of R using the definition. Since the pole is identified with the origin, we can easily tell the point R is 3 units from the pole, which means in the polar representation (r,θ) of R we know $r = \pm 3$. Since we require $r \geq 0$, we choose $r = 3$. Concerning θ, the angle $\theta = \frac{3\pi}{2}$ satisfies $0 \leq \theta < 2\pi$ with its terminal side along the negative y-axis, so our answer is $\left(3, \frac{3\pi}{2}\right)$. To check, we note $x = r\cos(\theta) = 3\cos\left(\frac{3\pi}{2}\right) = (3)(0) = 0$ and $y = r\sin(\theta) = 3\sin\left(\frac{3\pi}{2}\right) = 3(-1) = -3$.

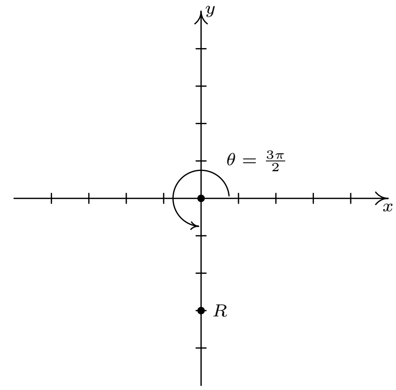


Figure 7.18: R has rectangular coordinates $(0,-3)$ and polar coordinates $(-3, \frac{3\pi}{2})$

4. The point $S(-3,4)$ lies in Quadrant II. With $x = -3$ and $y = 4$, we get $r^2 = (-3)^2 + (4)^2 = 25$ so $r = \pm 5$. As usual, we choose $r = 5 \geq 0$ and proceed to determine θ. We have $\tan(\theta) = \frac{y}{x} = \frac{4}{-3} = -\frac{4}{3}$, and since this isn't the tangent of one the common angles, we resort to using the arctangent function. Since θ lies in Quadrant II and must satisfy $0 \leq \theta < 2\pi$, we choose $\theta = \pi - \arctan\left(\frac{4}{3}\right)$ radians. Hence, our answer is $(r,\theta) = \left(5, \pi - \arctan\left(\frac{4}{3}\right)\right) \approx (5, 2.21)$. To check our answers requires a bit of tenacity since we need to simplify expressions of the form: $\cos\left(\pi - \arctan\left(\frac{4}{3}\right)\right)$ and $\sin\left(\pi - \arctan\left(\frac{4}{3}\right)\right)$. These are good review exercises and are hence left to the reader. We find $\cos\left(\pi - \arctan\left(\frac{4}{3}\right)\right) = -\frac{3}{5}$ and $\sin\left(\pi - \arctan\left(\frac{4}{3}\right)\right) = \frac{4}{5}$, so that $x = r\cos(\theta) = (5)\left(-\frac{3}{5}\right) = -3$ and $y = r\sin(\theta) = (5)\left(\frac{4}{5}\right) = 4$ which confirms our answer.

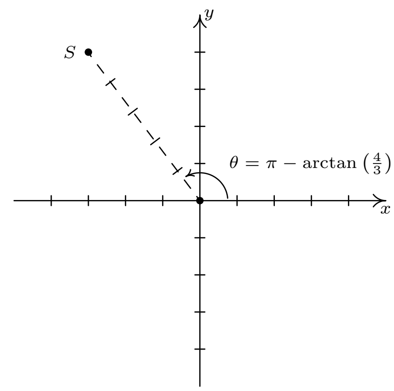


Figure 7.19: S has rectangular coordinates $(-3,4)$ and polar coordinates $(5, \pi - \arctan(\frac{4}{3}))$

Now that we've had practice converting representations of *points* between the rectangular and polar coordinate systems, we now set about converting *equations* from one system to another. Just as we've used equations in x and y to represent relations in rectangular coordinates, equations in the variables r and θ represent relations in polar coordinates. We convert equations between the two systems using Theorem 33 as the next examples illustrate.

Example 145 Converting equations from rectangular to polar

Convert each equation in rectangular coordinates into an equation in polar coordinates.

1. $(x-3)^2 + y^2 = 9$
2. $y = -x$
3. $y = x^2$

SOLUTION One strategy to convert an equation from rectangular to polar coordinates is to replace every occurrence of x with $r\cos(\theta)$ and every occurrence of y with $r\sin(\theta)$ and use identities to simplify. This is the technique we employ below.

1. We start by substituting $x = r\cos(\theta)$ and $y = \sin(\theta)$ into $(x-3)^2 + y^2 = 9$ and simplifying. With no real direction in which to proceed, we follow our mathematical instincts and see where they take us. (Experience is the mother of all instinct, and necessity is the mother of invention. Study this example and see what techniques are employed, then try your best to work through as many of the exercises as you can.)

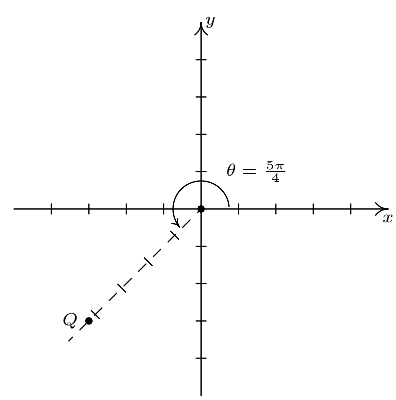


Figure 7.17: Q has rectangular coordinates $(-3,-3)$ and polar coordinates $(3\sqrt{2}, \frac{5\pi}{4})$

$$\begin{aligned}
(r\cos(\theta)-3)^2+(r\sin(\theta))^2 &= 9 \\
r^2\cos^2(\theta)-6r\cos(\theta)+9+r^2\sin^2(\theta) &= 9 \\
r^2\left(\cos^2(\theta)+\sin^2(\theta)\right)-6r\cos(\theta) &= 0 \\
r^2-6r\cos(\theta) &= 0 \quad (\cos^2(\theta)+\sin^2(\theta)=1) \\
r(r-6\cos(\theta)) &= 0 \qquad \text{Factor}
\end{aligned}$$

In Example 144.1, note that when we substitute $\theta = \frac{\pi}{2}$ into $r = 6\cos(\theta)$, we recover the point $r = 0$, so we aren't losing anything by disregarding $r = 0$.

Thus, we get $r = 0$ or $r = 6\cos(\theta)$. We know that the equation $(x-3)^2+y^2 = 9$ describes a circle, and since $r = 0$ describes just a point (namely the pole/origin), we choose $r = 6\cos(\theta)$ for our final answer.

2. Substituting $x = r\cos(\theta)$ and $y = r\sin(\theta)$ into $y = -x$ gives $r\sin(\theta) = -r\cos(\theta)$. Rearranging, we get $r\cos(\theta)+r\sin(\theta) = 0$ or $r(\cos(\theta)+\sin(\theta)) = 0$. This gives $r = 0$ or $\cos(\theta)+\sin(\theta) = 0$. Solving the latter equation for θ, we get $\theta = -\frac{\pi}{4}+\pi k$ for integers k. As we did in the previous example, we take a step back and think geometrically. We know $y = -x$ describes a line through the origin. As before, $r = 0$ describes the origin, but nothing else. Consider the equation $\theta = -\frac{\pi}{4}$. In this equation, the variable r is free, meaning it can assume any and all values including $r = 0$. If we imagine plotting points $(r, -\frac{\pi}{4})$ for all conceivable values of r (positive, negative and zero), we are essentially drawing the line containing the terminal side of $\theta = -\frac{\pi}{4}$ which is none other than $y = -x$. Hence, we can take as our final answer $\theta = -\frac{\pi}{4}$ here. (We could take it to be *any* of $\theta = -\frac{\pi}{4}+\pi k$ for integers k, but it's nice to keep things simple.)

3. We substitute $x = r\cos(\theta)$ and $y = r\sin(\theta)$ into $y = x^2$ and get $r\sin(\theta) = (r\cos(\theta))^2$, or $r^2\cos^2(\theta)-r\sin(\theta) = 0$. Factoring, we get $r(r\cos^2(\theta)-\sin(\theta)) = 0$ so that either $r = 0$ or $r\cos^2(\theta) = \sin(\theta)$. We can solve the latter equation for r by dividing both sides of the equation by $\cos^2(\theta)$, but as a general rule, we never divide through by a quantity that may be 0. In this particular case, we are safe since if $\cos^2(\theta) = 0$, then $\cos(\theta) = 0$, and for the equation $r\cos^2(\theta) = \sin(\theta)$ to hold, then $\sin(\theta)$ would also have to be 0. Since there are no angles with both $\cos(\theta) = 0$ and $\sin(\theta) = 0$, we are not losing any information by dividing both sides of $r\cos^2(\theta) = \sin(\theta)$ by $\cos^2(\theta)$. Doing so, we get $r = \frac{\sin(\theta)}{\cos^2(\theta)}$, or $r = \sec(\theta)\tan(\theta)$. As before, the $r = 0$ case is recovered in the solution $r = \sec(\theta)\tan(\theta)$ (let $\theta = 0$), so we state the latter as our final answer.

Example 146 Converting equations from polar to rectangular

Convert each equation in polar coordinates into an equation in rectangular coordinates.

1. $r = -3$
2. $\theta = \frac{4\pi}{3}$
3. $r = 1-\cos(\theta)$

Solution As a general rule, converting equations from polar to rectangular coordinates isn't as straight forward as the reverse process. We could solve $r^2 = x^2+y^2$ for r to get $r = \pm\sqrt{x^2+y^2}$ and solving $\tan(\theta) = \frac{y}{x}$ requires the arctangent function to get $\theta = \arctan\left(\frac{y}{x}\right)+\pi k$ for integers k. Neither of these expressions for r and θ are especially user-friendly, so we opt for a second strategy – rearrange the given polar equation so that the expressions $r^2 = x^2+y^2$, $r\cos(\theta) = x$, $r\sin(\theta) = y$ and/or $\tan(\theta) = \frac{y}{x}$ present themselves.

1. Starting with $r = -3$, we can square both sides to get $r^2 = (-3)^2$ or $r^2 = 9$. We may now substitute $r^2 = x^2 + y^2$ to get the equation $x^2 + y^2 = 9$. Recall that squaring an equation does not, in general, produce an equivalent equation. The concern here is that the equation $r^2 = 9$ might be satisfied by more points than $r = -3$. On the surface, this appears to be the case since $r^2 = 9$ is equivalent to $r = \pm 3$, not just $r = -3$. However, any point with polar coordinates $(3, \theta)$ can be represented as $(-3, \theta + \pi)$, which means any point (r, θ) whose polar coordinates satisfy the relation $r = \pm 3$ has an equivalent representation which satisfies $r = -3$.

2. We take the tangent of both sides the equation $\theta = \frac{4\pi}{3}$ to get $\tan(\theta) = \tan\left(\frac{4\pi}{3}\right) = \sqrt{3}$. Since $\tan(\theta) = \frac{y}{x}$, we get $\frac{y}{x} = \sqrt{3}$ or $y = x\sqrt{3}$. Of course, we pause a moment to wonder if, geometrically, the equations $\theta = \frac{4\pi}{3}$ and $y = x\sqrt{3}$ generate the same set of points. (In addition to taking the tangent of both sides of an equation (There are infinitely many solutions to $\tan(\theta) = \sqrt{3}$, and $\theta = \frac{4\pi}{3}$ is only one of them!), we also went from $\frac{y}{x} = \sqrt{3}$, in which x cannot be 0, to $y = x\sqrt{3}$ in which we assume x can be 0.) The same argument presented in number 2 applies equally well here so we are done.

3. Once again, we need to manipulate $r = 1 - \cos(\theta)$ a bit before using the conversion formulas given in Theorem 33. We could square both sides of this equation like we did in part 1 above to obtain an r^2 on the left hand side, but that does nothing helpful for the right hand side. Instead, we multiply both sides by r to obtain $r^2 = r - r\cos(\theta)$. We now have an r^2 and an $r\cos(\theta)$ in the equation, which we can easily handle, but we also have another r to deal with. Rewriting the equation as $r = r^2 + r\cos(\theta)$ and squaring both sides yields $r^2 = \left(r^2 + r\cos(\theta)\right)^2$. Substituting $r^2 = x^2 + y^2$ and $r\cos(\theta) = x$ gives $x^2 + y^2 = \left(x^2 + y^2 + x\right)^2$. Once again, we have performed some algebraic manoeuvres which may have altered the set of points described by the original equation. First, we multiplied both sides by r. This means that now $r = 0$ is a viable solution to the equation. In the original equation, $r = 1 - \cos(\theta)$, we see that $\theta = 0$ gives $r = 0$, so the multiplication by r doesn't introduce any new points. The squaring of both sides of this equation is also a reason to pause. Are there points with coordinates (r, θ) which satisfy $r^2 = \left(r^2 + r\cos(\theta)\right)^2$ but do not satisfy $r = r^2 + r\cos(\theta)$? Suppose (r', θ') satisfies $r^2 = \left(r^2 + r\cos(\theta)\right)^2$. Then $r' = \pm\left((r')^2 + r'\cos(\theta')\right)$. If we have that $r' = (r')^2 + r'\cos(\theta')$, we are done. What if $r' = -\left((r')^2 + r'\cos(\theta')\right) = -(r')^2 - r'\cos(\theta')$? We claim that the coordinates $(-r', \theta' + \pi)$, which determine the same point as (r', θ'), satisfy $r = r^2 + r\cos(\theta)$. If $r = -r'$ and $\theta = \theta' + \pi$, then we have

$$\begin{aligned}
r^2 + r\cos(\theta) &= (-r')^2 + (-r')\cos(\theta' + \pi) \\
&= (r')^2 - r'(-\cos(\theta')) && \text{Since } \cos(\theta' + \pi) = -\cos(\theta') \\
&= (r')^2 + r'\cos(\theta') \\
&= -r' && \text{Since } r' = -(r')^2 - r'\cos(\theta') \\
&= r.
\end{aligned}$$

Thus, the point $(-r', \theta' + \pi)$ satisfies $r = r^2 + r\cos(\theta)$, which means that any point (r, θ) which satisfies $r^2 = \left(r^2 + r\cos(\theta)\right)^2$ has a representation which satisfies $r = r^2 + r\cos(\theta)$, and we are done.

When we say that two representations of a point are 'equivalent', we mean that they represent the same point in the plane. As ordered pairs, $(3, 0)$ and $(-3, \pi)$ are different, but when interpreted as polar coordinates, they correspond to the same point in the plane. The same applies to equations defining sets of ordered pairs in the plane. For example, the equations $r^2 = 9$ and $r = -3$ represent different relations, since they correspond to different sets of ordered pairs of polar coordinates. However, since polar coordinates were defined geometrically to describe the location of points in the plane, we concern ourselves only with ensuring that the sets of *points* in the plane generated by two equations are the same.

In practice, much of the pedantic verification of the equivalence of equations in Examples 144 and 145 is left unsaid. Indeed, in most textbooks, squaring equations like $r = -3$ to arrive at $r^2 = 9$ happens without a second thought. Your instructor will ultimately decide how much, if any, justification is warranted. If you take anything away from these examples, it should be that relatively nice things in rectangular coordinates, such as $y = x^2$, can turn ugly in polar coordinates, and vice-versa. If nothing else, number 3 above shows the price we pay if we insist on always converting to back to the more familiar rectangular coordinate system.

Exercises 7.2

Problems

In Exercises 1 – 16, plot the point given in polar coordinates and then give three different expressions for the point such that (a) $r < 0$ and $0 \le \theta \le 2\pi$, (b) $r > 0$ and $\theta \le 0$ (c) $r > 0$ and $\theta \ge 2\pi$

1. $\left(2, \frac{\pi}{3}\right)$
2. $\left(5, \frac{7\pi}{4}\right)$
3. $\left(\frac{1}{3}, \frac{3\pi}{2}\right)$
4. $\left(\frac{5}{2}, \frac{5\pi}{6}\right)$
5. $\left(12, -\frac{7\pi}{6}\right)$
6. $\left(3, -\frac{5\pi}{4}\right)$
7. $(2\sqrt{2}, -\pi)$
8. $\left(\frac{7}{2}, -\frac{13\pi}{6}\right)$
9. $(-20, 3\pi)$
10. $\left(-4, \frac{5\pi}{4}\right)$
11. $\left(-1, \frac{2\pi}{3}\right)$
12. $\left(-3, \frac{\pi}{2}\right)$
13. $\left(-3, -\frac{11\pi}{6}\right)$
14. $\left(-2.5, -\frac{\pi}{4}\right)$
15. $\left(-\sqrt{5}, -\frac{4\pi}{3}\right)$
16. $(-\pi, -\pi)$

In Exercises 17 – 36, convert the point from polar coordinates into rectangular coordinates.

17. $\left(5, \frac{7\pi}{4}\right)$
18. $\left(2, \frac{\pi}{3}\right)$
19. $\left(11, -\frac{7\pi}{6}\right)$
20. $(-20, 3\pi)$
21. $\left(\frac{3}{5}, \frac{\pi}{2}\right)$
22. $\left(-4, \frac{5\pi}{6}\right)$
23. $\left(9, \frac{7\pi}{2}\right)$
24. $\left(-5, -\frac{9\pi}{4}\right)$
25. $\left(42, \frac{13\pi}{6}\right)$
26. $(-117, 117\pi)$
27. $(6, \arctan(2))$
28. $(10, \arctan(3))$
29. $\left(-3, \arctan\left(\frac{4}{3}\right)\right)$
30. $\left(5, \arctan\left(-\frac{4}{3}\right)\right)$
31. $\left(2, \pi - \arctan\left(\frac{1}{2}\right)\right)$
32. $\left(-\frac{1}{2}, \pi - \arctan(5)\right)$
33. $\left(-1, \pi + \arctan\left(\frac{3}{4}\right)\right)$
34. $\left(\frac{2}{3}, \pi + \arctan\left(2\sqrt{2}\right)\right)$
35. $(\pi, \arctan(\pi))$
36. $\left(13, \arctan\left(\frac{12}{5}\right)\right)$

In Exercises 37 – 56, convert the point from rectangular coordinates into polar coordinates with $r \ge 0$ and $0 \le \theta < 2\pi$.

37. $(0, 5)$
38. $(3, \sqrt{3})$

39. $(7, -7)$

40. $(-3, -\sqrt{3})$

41. $(-3, 0)$

42. $\left(-\sqrt{2}, \sqrt{2}\right)$

43. $\left(-4, -4\sqrt{3}\right)$

44. $\left(\frac{\sqrt{3}}{4}, -\frac{1}{4}\right)$

45. $\left(-\frac{3}{10}, -\frac{3\sqrt{3}}{10}\right)$

46. $\left(-\sqrt{5}, -\sqrt{5}\right)$

47. $(6, 8)$

48. $(\sqrt{5}, 2\sqrt{5})$

49. $(-8, 1)$

50. $(-2\sqrt{10}, 6\sqrt{10})$

51. $(-5, -12)$

52. $\left(-\frac{\sqrt{5}}{15}, -\frac{2\sqrt{5}}{15}\right)$

53. $(24, -7)$

54. $(12, -9)$

55. $\left(\frac{\sqrt{2}}{4}, \frac{\sqrt{6}}{4}\right)$

56. $\left(-\frac{\sqrt{65}}{5}, \frac{2\sqrt{65}}{5}\right)$

In Exercises 57 – 76, convert the equation from rectangular coordinates into polar coordinates. Solve for r in all but #60 through #63. In Exercises 60 - 63, you need to solve for θ.

57. $x = 6$

58. $x = -3$

59. $y = 7$

60. $y = 0$

61. $y = -x$

62. $y = x\sqrt{3}$

63. $y = 2x$

64. $x^2 + y^2 = 25$

65. $x^2 + y^2 = 117$

66. $y = 4x - 19$

67. $x = 3y + 1$

68. $y = -3x^2$

69. $4x = y^2$

70. $x^2 + y^2 - 2y = 0$

71. $x^2 - 4x + y^2 = 0$

72. $x^2 + y^2 = x$

73. $y^2 = 7y - x^2$

74. $(x + 2)^2 + y^2 = 4$

75. $x^2 + (y - 3)^2 = 9$

76. $4x^2 + 4\left(y - \frac{1}{2}\right)^2 = 1$

In Exercises 77 – 96, convert the equation from polar coordinates into rectangular coordinates.

77. $r = 7$

78. $r = -3$

79. $r = \sqrt{2}$

80. $\theta = \frac{\pi}{4}$

81. $\theta = \frac{2\pi}{3}$

82. $\theta = \pi$

83. $\theta = \frac{3\pi}{2}$

84. $r = 4\cos(\theta)$

85. $5r = \cos(\theta)$

86. $r = 3\sin(\theta)$

87. $r = -2\sin(\theta)$

88. $r = 7\sec(\theta)$

89. $12r = \csc(\theta)$

90. $r = -2\sec(\theta)$

91. $r = -\sqrt{5}\csc(\theta)$

92. $r = 2\sec(\theta)\tan(\theta)$

93. $r = -\csc(\theta)\cot(\theta)$

94. $r^2 = \sin(2\theta)$

95. $r = 1 - 2\cos(\theta)$

96. $r = 1 + \sin(\theta)$

97. Convert the origin $(0, 0)$ into polar coordinates in four different ways.

98. With the help of your classmates, use the Law of Cosines to develop a formula for the distance between two points in polar coordinates.

7.3 The Polar Form of Complex Numbers

In this section, we return to our study of complex numbers begun in Section 7.1. Recall that a **complex number** is a number of the form $z = a + bi$ where a and b are real numbers and i is the imaginary unit defined by $i = \sqrt{-1}$. The number a is called the **real part** of z, denoted $\text{Re}(z)$, while the real number b is called the **imaginary part** of z, denoted $\text{Im}(z)$. From Definition 54, we know that if $z = a + bi = c + di$ where a, b, c and d are real numbers, then $a = c$ and $b = d$, which means $\text{Re}(z)$ and $\text{Im}(z)$ are well-defined. To start off this section, we associate each complex number $z = a + bi$ with the point (a, b) on the coordinate plane. In this case, the x-axis is relabeled as the **real axis**, which corresponds to the real number line as usual, and the y-axis is relabeled as the **imaginary axis**, which is demarcated in increments of the imaginary unit i. The plane determined by these two axes is called the **complex plane**.

Saying that $\text{Re}(z)$ and $\text{Im}(z)$ are 'well-defined' means that no matter how we express z, the number $\text{Re}(z)$ is always the same, and the number $\text{Im}(z)$ is always the same. In other words, Re and Im are *functions* of complex numbers.

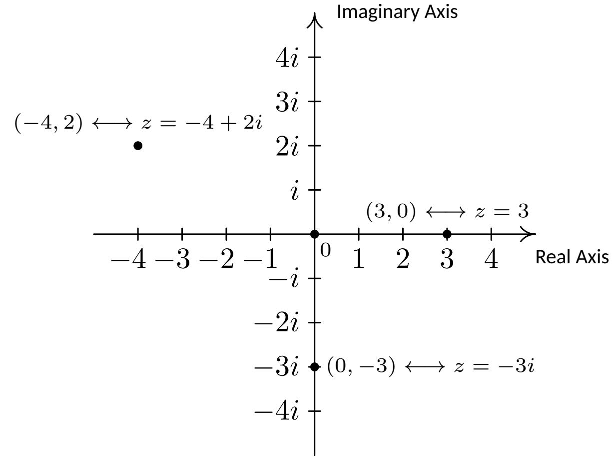


Figure 7.20: The complex plane

Since the ordered pair (a, b) gives the *rectangular* coordinates associated with the complex number $z = a + bi$, the expression $z = a + bi$ is called the **rectangular form** of z. Of course, we could just as easily associate z with a pair of *polar* coordinates (r, θ). Although it is not as straightforward as the definitions of $\text{Re}(z)$ and $\text{Im}(z)$, we can still give r and θ special names in relation to z.

Definition 55 The Modulus and Argument of Complex Numbers

Let $z = a + bi$ be a complex number with $a = \text{Re}(z)$ and $b = \text{Im}(z)$. Let (r, θ) be a polar representation of the point with rectangular coordinates (a, b) where $r \geq 0$.

- The **modulus** of z, denoted $|z|$, is defined by $|z| = r$.
- The angle θ is an **argument** of z. The set of all arguments of z is denoted $\text{arg}(z)$.
- If $z \neq 0$ and $-\pi < \theta \leq \pi$, then θ is the **principal argument** of z, written $\theta = \text{Arg}(z)$.

Some remarks about Definition 55 are in order. We know from Section 7.2 that every point in the plane has infinitely many polar coordinate representations (r, θ) which means it's worth our time to make sure the quantities 'modulus', 'argument' and 'principal argument' are well-defined. Concerning the modulus, if $z = 0$ then the point associated with z is the origin. In this case, the *only* r-value which can be used here is $r = 0$. Hence for $z = 0$, $|z| = 0$ is well-defined. If $z \neq 0$, then the point associated with z is not the origin, and there are *two* possibilities for r: one positive and one negative. However, we stipulated $r \geq 0$ in our definition so this pins down the value of $|z|$ to one and only one number. Thus the modulus is well-defined in this case, too. (In case you're wondering, the use of the absolute value notation $|z|$ for modulus will be explained shortly.) Even with the requirement $r \geq 0$, there are infinitely many angles θ which can be used in a polar representation of a point (r, θ). If $z \neq 0$ then the point in question is not the origin, so all of these angles θ are coterminal. Since coterminal angles are exactly 2π radians apart, we are guaranteed that only one of them lies in the interval $(-\pi, \pi]$, and this angle is what we call the principal argument of z, $\text{Arg}(z)$. In fact, the set $\arg(z)$ of all arguments of z can be described using set-builder notation as $\arg(z) = \{\text{Arg}(z) + 2\pi k \,|\, k \text{ is an integer}\}$. Note that since $\arg(z)$ is a *set*, we will write '$\theta \in \arg(z)$' to mean 'θ is in the set of arguments of z'. If $z = 0$ then the point in question is the origin, which we know can be represented in polar coordinates as $(0, \theta)$ for *any* angle θ. In this case, we have $\arg(0) = (-\infty, \infty)$ and since there is no one value of θ which lies $(-\pi, \pi]$, we leave $\text{Arg}(0)$ undefined. It is time for an example.

Example 147 Components of a complex number

For each of the following complex numbers find $\text{Re}(z)$, $\text{Im}(z)$, $|z|$, $\arg(z)$ and $\text{Arg}(z)$. Plot z in the complex plane.

1. $z = \sqrt{3} - i$
2. $z = -2 + 4i$
3. $z = 3i$
4. $z = -117$

Solution

1. For $z = \sqrt{3} - i = \sqrt{3} + (-1)i$, we have $\text{Re}(z) = \sqrt{3}$ and $\text{Im}(z) = -1$. To find $|z|$, $\arg(z)$ and $\text{Arg}(z)$, we need to find a polar representation (r, θ) with $r \geq 0$ for the point $P(\sqrt{3}, -1)$ associated with z. We know $r^2 = (\sqrt{3})^2 + (-1)^2 = 4$, so $r = \pm 2$. Since we require $r \geq 0$, we choose $r = 2$, so $|z| = 2$. Next, we find a corresponding angle θ. Since $r > 0$ and P lies in Quadrant IV, θ is a Quadrant IV angle. We know $\tan(\theta) = \frac{-1}{\sqrt{3}} = -\frac{\sqrt{3}}{3}$, so $\theta = -\frac{\pi}{6} + 2\pi k$ for integers k. Hence, $\arg(z) = \left\{-\frac{\pi}{6} + 2\pi k \,|\, k \text{ is an integer}\right\}$. Of these values, only $\theta = -\frac{\pi}{6}$ satisfies the requirement that $-\pi < \theta \leq \pi$, hence $\text{Arg}(z) = -\frac{\pi}{6}$.

2. The complex number $z = -2 + 4i$ has $\text{Re}(z) = -2$, $\text{Im}(z) = 4$, and is associated with the point $P(-2, 4)$. Our next task is to find a polar representation (r, θ) for P where $r \geq 0$. Running through the usual calculations gives $r = 2\sqrt{5}$, so $|z| = 2\sqrt{5}$. To find θ, we get $\tan(\theta) = -2$, and since $r > 0$ and P lies in Quadrant II, we know θ is a Quadrant II angle. We find $\theta = \pi + \arctan(-2) + 2\pi k$, or, more succinctly $\theta = \pi - \arctan(2) + 2\pi k$ for integers k. Hence $\arg(z) = \{\pi - \arctan(2) + 2\pi k \,|\, k \text{ is an integer}\}$. Only $\theta = \pi - \arctan(2)$ satisfies the requirement $-\pi < \theta \leq \pi$, so $\text{Arg}(z) = \pi - \arctan(2)$.

3. We rewrite $z = 3i$ as $z = 0 + 3i$ to find $\text{Re}(z) = 0$ and $\text{Im}(z) = 3$. The point in the plane which corresponds to z is $(0, 3)$ and while we could go through the usual calculations to find the required polar form of this point, we can

almost 'see' the answer. The point $(0, 3)$ lies 3 units away from the origin on the positive y-axis. Hence, $r = |z| = 3$ and $\theta = \frac{\pi}{2} + 2\pi k$ for integers k. We get $\arg(z) = \left\{\frac{\pi}{2} + 2\pi k \,|\, k \text{ is an integer}\right\}$ and $\operatorname{Arg}(z) = \frac{\pi}{2}$.

4. As in the previous problem, we write $z = -117 = -117 + 0i$ so $\operatorname{Re}(z) = -117$ and $\operatorname{Im}(z) = 0$. The number $z = -117$ corresponds to the point $(-117, 0)$, and this is another instance where we can determine the polar form 'by eye'. The point $(-117, 0)$ is 117 units away from the origin along the negative x-axis. Hence, $r = |z| = 117$ and $\theta = \pi + 2\pi = (2k + 1)\pi k$ for integers k. We have $\arg(z) = \{(2k + 1)\pi \,|\, k \text{ is an integer}\}$. Only one of these values, $\theta = \pi$, just barely lies in the interval $(-\pi, \pi]$ which means and $\operatorname{Arg}(z) = \pi$. We plot z along with the other numbers in this example in Figure 7.21 below.

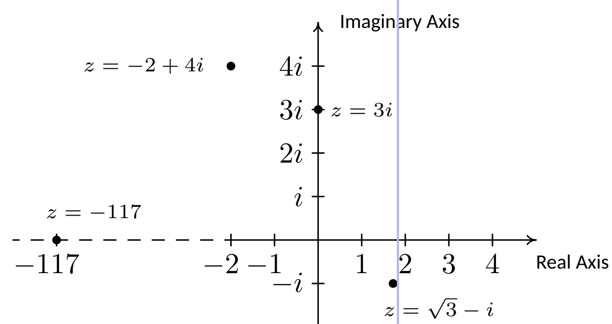


Figure 7.21: Plots of the four complex numbers in Example 146

Now that we've had some practice computing the modulus and argument of some complex numbers, it is time to explore their properties. We have the following theorem.

Theorem 34 Properties of the Modulus

Let z and w be complex numbers.

- $|z|$ is the distance from z to 0 in the complex plane
- $|z| \geq 0$ and $|z| = 0$ if and only if $z = 0$
- $|z| = \sqrt{\operatorname{Re}(z)^2 + \operatorname{Im}(z)^2}$
- **Product Rule:** $|zw| = |z|\,|w|$
- **Power Rule:** $|z^n| = |z|^n$ for all natural numbers, n
- **Quotient Rule:** $\left|\dfrac{z}{w}\right| = \dfrac{|z|}{|w|}$, provided $w \neq 0$

To prove the first three properties in Theorem 34, suppose $z = a + bi$ where a and b are real numbers. To determine $|z|$, we find a polar representation (r, θ) with $r \geq 0$ for the point (a, b). From Section 7.2, we know $r^2 = a^2 + b^2$ so that $r = \pm\sqrt{a^2 + b^2}$. Since we require $r \geq 0$, then it must be that $r = \sqrt{a^2 + b^2}$, which means $|z| = \sqrt{a^2 + b^2}$. Using the distance formula, we find the distance from $(0, 0)$ to (a, b) is also $\sqrt{a^2 + b^2}$, establishing the first property. For the second property, note that since $|z|$ is a distance, $|z| \geq 0$. Furthermore, $|z| = 0$ if and only if the distance from z to 0 is 0, and the latter happens if and only if $z = 0$, which is what we were asked to show. For the third property, we note that since $a = \text{Re}(z)$ and $b = \text{Im}(z)$, $z = \sqrt{a^2+b^2} = \sqrt{\text{Re}(z)^2 + \text{Im}(z)^2}$.

To prove the product rule, suppose $z = a + bi$ and $w = c + di$ for real numbers a, b, c and d. Then $zw = (a+bi)(c+di)$. After the usual arithmetic we get $zw = (ac - bd) + (ad + bc)i$. (See Example 140 in Section 7.1 for a review of complex number arithmetic.) Therefore,

In case you were not convinced by the argument for the second property in Theorem 34, we can work through the underlying Algebra to see this is true. We know $|z| = 0$ if and only if $\sqrt{a^2+b^2} = 0$ if and only if $a^2 + b^2 = 0$, which is true if and only if $a = b = 0$. The latter happens if and only if $z = a + bi = 0$. There.

$$\begin{aligned}
|zw| &= \sqrt{(ac-bd)^2 + (ad+bc)^2} & \\
&= \sqrt{a^2c^2 - 2abcd + b^2d^2 + a^2d^2 + 2abcd + b^2c^2} & \text{Expand} \\
&= \sqrt{a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2} & \text{Rearrange terms} \\
&= \sqrt{a^2\left(c^2+d^2\right) + b^2\left(c^2+d^2\right)} & \text{Factor} \\
&= \sqrt{\left(a^2+b^2\right)\left(c^2+d^2\right)} & \text{Factor} \\
&= \sqrt{a^2+b^2}\sqrt{c^2+d^2} & \text{Product Rule for Radicals} \\
&= |z|\,|w| & \text{Definition of } |z| \text{ and } |w|
\end{aligned}$$

Hence $|zw| = |z|\,|w|$ as required.

The Power Rule essentially follows from repeated application of the Product Rule, and again, a formal proof requires Mathematical Induction, so we leave it as a problem for you to consider when you take Math 2000.

Like the Power Rule, the Quotient Rule can also be established with the help of the Product Rule. We assume $w \neq 0$ (so $|w| \neq 0$) and we get

$$\begin{aligned}
\left|\frac{z}{w}\right| &= \left|(z)\left(\frac{1}{w}\right)\right| & \\
[3pt] &= |z|\left|\frac{1}{w}\right| & \text{Product Rule.}
\end{aligned}$$

Hence, the proof really boils down to showing $\left|\frac{1}{w}\right| = \frac{1}{|w|}$. This is left as an exercise.

Next, we characterize the argument of a complex number in terms of its real and imaginary parts.

Theorem 35 Properties of the Argument

Let z be a complex number.

- If $\text{Re}(z) \neq 0$ and $\theta \in \arg(z)$, then $\tan(\theta) = \frac{\text{Im}(z)}{\text{Re}(z)}$.
- If $\text{Re}(z) = 0$ and $\text{Im}(z) > 0$, then $\arg(z) = \left\{\frac{\pi}{2} + 2\pi k \,|\, k \text{ is an integer}\right\}$.
- If $\text{Re}(z) = 0$ and $\text{Im}(z) < 0$, then $\arg(z) = \left\{-\frac{\pi}{2} + 2\pi k \,|\, k \text{ is an integer}\right\}$.
- If $\text{Re}(z) = \text{Im}(z) = 0$, then $z = 0$ and $\arg(z) = (-\infty, \infty)$.

Since the absolute value $|x|$ of a real number x can be viewed as the distance from x to 0 on the number line, the first property in Theorem 34 justifies the notation $|z|$ for modulus. We leave it to the reader to show that if z is real, then the definition of modulus coincides with absolute value so the notation $|z|$ is unambiguous.

To prove Theorem 35, suppose $z = a + bi$ for real numbers a and b. By definition, $a = \text{Re}(z)$ and $b = \text{Im}(z)$, so the point associated with z is $(a, b) =$

$(\text{Re}(z), \text{Im}(z))$. From Section 7.2, we know that if (r, θ) is a polar representation for $(\text{Re}(z), \text{Im}(z))$, then $\tan(\theta) = \frac{\text{Im}(z)}{\text{Re}(z)}$, provided $\text{Re}(z) \neq 0$. If $\text{Re}(z) = 0$ and $\text{Im}(z) > 0$, then z lies on the positive imaginary axis. Since we take $r > 0$, we have that θ is coterminal with $\frac{\pi}{2}$, and the result follows. If $\text{Re}(z) = 0$ and $\text{Im}(z) < 0$, then z lies on the negative imaginary axis, and a similar argument shows θ is coterminal with $-\frac{\pi}{2}$. The last property in the theorem was already discussed in the remarks following Definition 55.

Our next goal is to completely marry the Geometry and the Algebra of the complex numbers. To that end, consider Figure 7.22 below.

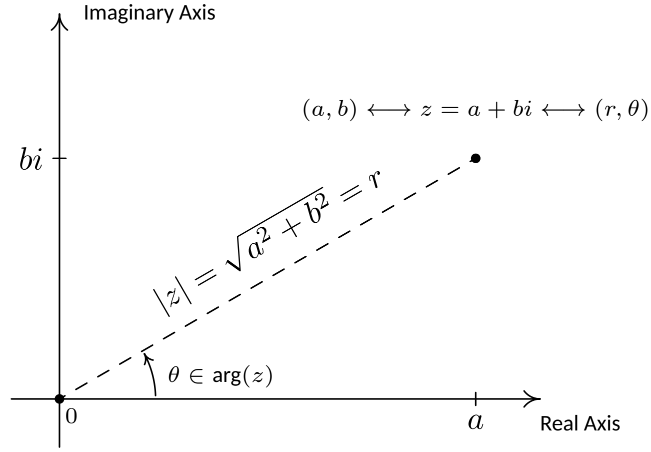


Figure 7.22: Polar coordinates, (r, θ) associated with $z = a + bi$ with $r \geq 0$.

We know from Theorem 33 that $a = r\cos(\theta)$ and $b = r\sin(\theta)$. Making these substitutions for a and b gives $z = a + bi = r\cos(\theta) + r\sin(\theta)i = r[\cos(\theta) + i\sin(\theta)]$. The expression '$\cos(\theta) + i\sin(\theta)$' is abbreviated $\text{cis}(\theta)$ so we can write $z = r\,\text{cis}(\theta)$. Since $r = |z|$ and $\theta \in \text{arg}(z)$, we get

Definition 56 A Polar Form of a Complex Number

Suppose z is a complex number and $\theta \in \text{arg}(z)$. The expression:

$$|z|\,\text{cis}(\theta) = |z|\,[\cos(\theta) + i\sin(\theta)]$$

is called a polar form for z.

Since there are infinitely many choices for $\theta \in \text{arg}(z)$, there infinitely many polar forms for z, so we used the indefinite article 'a' in Definition 56. It is time for an example.

Example 148 Converting between rectangular and polar form

1. Find the rectangular form of the following complex numbers. Find $\text{Re}(z)$ and $\text{Im}(z)$.

   (a) $z = 4\,\text{cis}\left(\frac{2\pi}{3}\right)$
   (b) $z = 2\,\text{cis}\left(-\frac{3\pi}{4}\right)$
   (c) $z = 3\,\text{cis}(0)$
   (d) $z = \text{cis}\left(\frac{\pi}{2}\right)$

2. Use the results from Example 146 to find a polar form of the following complex numbers.

   (a) $z = \sqrt{3} - i$
   (b) $z = -2 + 4i$
   (c) $z = 3i$
   (d) $z = -117$

Solution

1. The key to this problem is to write out $\text{cis}(\theta)$ as $\cos(\theta) + i\sin(\theta)$.

   (a) By definition, $z = 4\,\text{cis}\left(\frac{2\pi}{3}\right) = 4\left[\cos\left(\frac{2\pi}{3}\right) + i\sin\left(\frac{2\pi}{3}\right)\right]$. After some simplifying, we get $z = -2 + 2i\sqrt{3}$, so that $\text{Re}(z) = -2$ and $\text{Im}(z) = 2\sqrt{3}$.

   (b) Expanding, we get $z = 2\,\text{cis}\left(-\frac{3\pi}{4}\right) = 2\left[\cos\left(-\frac{3\pi}{4}\right) + i\sin\left(-\frac{3\pi}{4}\right)\right]$. From this, we find $z = -\sqrt{2} - i\sqrt{2}$, so $\text{Re}(z) = -\sqrt{2} = \text{Im}(z)$.

   (c) We get $z = 3\,\text{cis}(0) = 3\left[\cos(0) + i\sin(0)\right] = 3$. Writing $3 = 3 + 0i$, we get $\text{Re}(z) = 3$ and $\text{Im}(z) = 0$, which makes sense seeing as 3 is a real number.

   (d) Lastly, we have $z = \text{cis}\left(\frac{\pi}{2}\right) = \cos\left(\frac{\pi}{2}\right) + i\sin\left(\frac{\pi}{2}\right) = i$. Since $i = 0 + 1i$, we get $\text{Re}(z) = 0$ and $\text{Im}(z) = 1$. Since i is called the 'imaginary unit,' these answers make perfect sense.

2. To write a polar form of a complex number z, we need two pieces of information: the modulus $|z|$ and an argument (not necessarily the principal argument) of z. We shamelessly mine our solution to Example 146 to find what we need.

   (a) For $z = \sqrt{3} - i$, $|z| = 2$ and $\theta = -\frac{\pi}{6}$, so $z = 2\,\text{cis}\left(-\frac{\pi}{6}\right)$. We can check our answer by converting it back to rectangular form to see that it simplifies to $z = \sqrt{3} - i$.

   (b) For $z = -2 + 4i$, $|z| = 2\sqrt{5}$ and $\theta = \pi - \arctan(2)$. Hence, $z = 2\sqrt{5}\,\text{cis}(\pi - \arctan(2))$. It is a good exercise to actually show that this polar form reduces to $z = -2 + 4i$.

   (c) For $z = 3i$, $|z| = 3$ and $\theta = \frac{\pi}{2}$. In this case, $z = 3\,\text{cis}\left(\frac{\pi}{2}\right)$. This can be checked geometrically. Head out 3 units from 0 along the positive real axis. Rotating $\frac{\pi}{2}$ radians counter-clockwise lands you exactly 3 units above 0 on the imaginary axis at $z = 3i$.

   (d) Last but not least, for $z = -117$, $|z| = 117$ and $\theta = \pi$. We get $z = 117\,\text{cis}(\pi)$. As with the previous problem, our answer is easily checked geometrically.

The following theorem summarizes the advantages of working with complex numbers in polar form.

Theorem 36 Products, Powers and Quotients Complex Numbers in Polar Form

Suppose z and w are complex numbers with polar forms $z = |z|\,\text{cis}(\alpha)$ and $w = |w|\,\text{cis}(\beta)$. Then

- **Product Rule:** $zw = |z|\,|w|\,\text{cis}(\alpha+\beta)$
- **Power Rule (DeMoivre's Theorem) :** $z^n = |z|^n\,\text{cis}(n\theta)$ for every natural number n
- **Quotient Rule:** $\dfrac{z}{w} = \dfrac{|z|}{|w|}\,\text{cis}(\alpha-\beta)$, provided $|w| \neq 0$

The proof of Theorem 36 requires a healthy mix of definition, arithmetic and identities. We first start with the product rule.

$$\begin{aligned} zw &= [|z|\,\text{cis}(\alpha)]\,[|w|\,\text{cis}(\beta)] \\ &= |z||w|\,[\cos(\alpha) + i\sin(\alpha)]\,[\cos(\beta) + i\sin(\beta)] \end{aligned}$$

We now focus on the quantity in brackets on the right hand side of the equation.

$$\begin{aligned} &[\cos(\alpha) + i\sin(\alpha)]\,[\cos(\beta) + i\sin(\beta)] \\ &\quad= \cos(\alpha)\cos(\beta) + i\cos(\alpha)\sin(\beta) \\ &\qquad + i\sin(\alpha)\cos(\beta) + i^2\sin(\alpha)\sin(\beta) \\ &\quad= \cos(\alpha)\cos(\beta) + i^2\sin(\alpha)\sin(\beta) && \text{Rearranging terms} \\ &\qquad + i\sin(\alpha)\cos(\beta) + i\cos(\alpha)\sin(\beta) \\ &\quad= (\cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)) && \text{Since } i^2 = -1 \\ &\quad+ \quad i\,(\sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta)) && \text{Factor out } i \\ &\quad= \cos(\alpha+\beta) + i\sin(\alpha+\beta) && \text{Sum identities} \\ &\quad= \text{cis}(\alpha+\beta) && \text{Definition of `cis'} \end{aligned}$$

Putting this together with our earlier work, we get $zw = |z|\,|w|\,\text{cis}(\alpha+\beta)$, as required.

As with with the proof of the Power Rule in Theorem 34, the proof of the Power Rule (better known as DeMoivre's Theorem) requires Mathematical Induction, and is therefore omitted from this text.

The last property in Theorem 36 to prove is the quotient rule. Assuming $|w| \neq 0$ we have

$$\begin{aligned} \frac{z}{w} &= \frac{|z|\,\text{cis}(\alpha)}{|w|\,\text{cis}(\beta)} \\ [3pt] &= \left(\frac{|z|}{|w|}\right)\frac{\cos(\alpha) + i\sin(\alpha)}{\cos(\beta) + i\sin(\beta)} \end{aligned}$$

Next, we multiply both the numerator and denominator of the right hand side by $(\cos(\beta) - i\sin(\beta))$ which is the complex conjugate of $(\cos(\beta) + i\sin(\beta))$ to get

While the notation $\text{cis}(\theta) = \cos(\theta) + i\sin(\theta)$ is not uncommon, most authors prefer to use Euler's exponential notation. In light of Theorem 36, one can make sense of the polar form using **Euler's formula**

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

The appearance of the exponential function in this context might seem strange, but note that the three properties in Theorem 36 can then be understood in terms of laws of exponents. If $z = re^{i\alpha}$ and $w = se^{i\beta}$, we have

$$zw = (rs)\left(e^{i\alpha}e^{i\beta}\right) = (rs)e^{i(\alpha+\beta)},$$

$$z^n = r^n(e^{i\alpha})^n = r^n e^{in\alpha},$$

and so on. For more details, see Exercise 82.

$$\frac{z}{w} = \left(\frac{|z|}{|w|}\right) \frac{\cos(\alpha) + i\sin(\alpha)}{\cos(\beta) + i\sin(\beta)} \cdot \frac{\cos(\beta) - i\sin(\beta)}{\cos(\beta) - i\sin(\beta)}$$

If we let the numerator be $N = [\cos(\alpha) + i\sin(\alpha)][\cos(\beta) - i\sin(\beta)]$ and simplify we get

$$\begin{aligned}
N &= [\cos(\alpha) + i\sin(\alpha)][\cos(\beta) - i\sin(\beta)] \\
&= \cos(\alpha)\cos(\beta) - i\cos(\alpha)\sin(\beta) \\
&\qquad + i\sin(\alpha)\cos(\beta) - i^2\sin(\alpha)\sin(\beta) && \text{Expand} \\
&= [\cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)] \\
&\qquad + i[\sin(\alpha)\cos(\beta) - \cos(\alpha)\sin(\beta)] && \text{Rearrange and Factor} \\
&= \cos(\alpha - \beta) + i\sin(\alpha - \beta) && \text{Difference Identities} \\
&= \text{cis}(\alpha - \beta) && \text{Definition of 'cis'}
\end{aligned}$$

If we call the denominator D then we get

$$\begin{aligned}
D &= [\cos(\beta) + i\sin(\beta)][\cos(\beta) - i\sin(\beta)] \\
&= \cos^2(\beta) - i\cos(\beta)\sin(\beta) \\
&\qquad + i\cos(\beta)\sin(\beta) - i^2\sin^2(\beta) && \text{Expand} \\
&= \cos^2(\beta) - i^2\sin^2(\beta) && \text{Simplify} \\
&= \cos^2(\beta) + \sin^2(\beta) && \text{Again, } i^2 = -1 \\
&= 1 && \text{Pythagorean Identity}
\end{aligned}$$

Putting it all together, we get

$$\begin{aligned}
\frac{z}{w} &= \left(\frac{|z|}{|w|}\right) \frac{\cos(\alpha) + i\sin(\alpha)}{\cos(\beta) + i\sin(\beta)} \cdot \frac{\cos(\beta) - i\sin(\beta)}{\cos(\beta) - i\sin(\beta)} \\
&= \left(\frac{|z|}{|w|}\right) \frac{\text{cis}(\alpha - \beta)}{1} \\
&= \frac{|z|}{|w|}\text{cis}(\alpha - \beta)
\end{aligned}$$

and we are done. The next example makes good use of Theorem 36.

Example 149 Complex arithmetic using the polar form

Let $z = 2\sqrt{3} + 2i$ and $w = -1 + i\sqrt{3}$. Use Theorem 36 to find the following.

1. zw
2. w^5
3. $\frac{z}{w}$

Write your final answers in rectangular form.

Solution In order to use Theorem 36, we need to write z and w in polar form. For $z = 2\sqrt{3} + 2i$, we find $|z| = \sqrt{(2\sqrt{3})^2 + (2)^2} = \sqrt{16} = 4$. If $\theta \in \arg(z)$, we know $\tan(\theta) = \frac{\text{Im}(z)}{\text{Re}(z)} = \frac{2}{2\sqrt{3}} = \frac{\sqrt{3}}{3}$. Since z lies in Quadrant I, we have $\theta = \frac{\pi}{6} + 2\pi k$ for integers k. Hence, $z = 4\,\text{cis}\left(\frac{\pi}{6}\right)$. For $w = -1 + i\sqrt{3}$, we have $|w| = \sqrt{(-1)^2 + (\sqrt{3})^2} = 2$. For an argument θ of w, we have $\tan(\theta) = \frac{\sqrt{3}}{-1} = -\sqrt{3}$. Since w lies in Quadrant II, $\theta = \frac{2\pi}{3} + 2\pi k$ for integers k and $w = 2\,\text{cis}\left(\frac{2\pi}{3}\right)$. We can now proceed.

1. We get $zw = \left(4\operatorname{cis}\left(\frac{\pi}{6}\right)\right)\left(2\operatorname{cis}\left(\frac{2\pi}{3}\right)\right) = 8\operatorname{cis}\left(\frac{\pi}{6}+\frac{2\pi}{3}\right) = 8\operatorname{cis}\left(\frac{5\pi}{6}\right) = 8\left[\cos\left(\frac{5\pi}{6}\right)+i\sin\left(\frac{5\pi}{6}\right)\right]$. After simplifying, we get $zw = -4\sqrt{3}+4i$.

2. We use DeMoivre's Theorem which yields

$$w^5 = \left[2\operatorname{cis}\left(\frac{2\pi}{3}\right)\right]^5 = 2^5\operatorname{cis}\left(5\cdot\frac{2\pi}{3}\right) = 32\operatorname{cis}\left(\frac{10\pi}{3}\right).$$

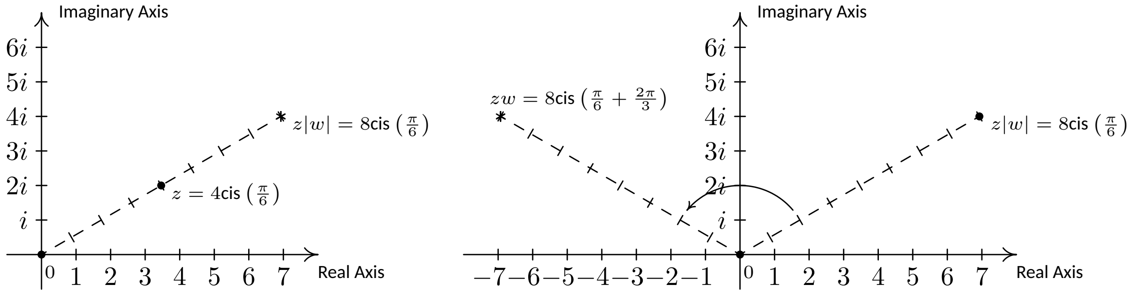
Since $\frac{10\pi}{3}$ is coterminal with $\frac{4\pi}{3}$, we get

$$w^5 = 32\left[\cos\left(\frac{4\pi}{3}\right)+i\sin\left(\frac{4\pi}{3}\right)\right] = -16-16i\sqrt{3}.$$

3. Last, but not least, we have $\frac{z}{w} = \frac{4\operatorname{cis}\left(\frac{\pi}{6}\right)}{2\operatorname{cis}\left(\frac{2\pi}{3}\right)} = \frac{4}{2}\operatorname{cis}\left(\frac{\pi}{6}-\frac{2\pi}{3}\right) = 2\operatorname{cis}\left(-\frac{\pi}{2}\right)$. Since the angle $-\frac{\pi}{2}$ lies along the negative y-axis, we can 'see' the rectangular form by moving out 2 units along the positive real axis, then rotating $\frac{\pi}{2}$ radians *clockwise* to arrive at the point 2 units below 0 on the imaginary axis. The long and short of it is that $\frac{z}{w} = -2i$.

Some remarks are in order. First, the reader may not be sold on using the polar form of complex numbers to multiply complex numbers – especially if they aren't given in polar form to begin with. Indeed, a lot of work was needed to convert the numbers z and w in Example 148 into polar form, compute their product, and convert back to rectangular form – certainly more work than is required to multiply out $zw = (2\sqrt{3}+2i)(-1+i\sqrt{3})$ the old-fashioned way. However, Theorem 36 pays huge dividends when computing powers of complex numbers. Consider how we computed w^5 above and compare that to using the Binomial Theorem to accomplish the same feat by expanding $(-1+i\sqrt{3})^5$. Division is tricky in the best of times, and we saved ourselves a lot of time and effort using Theorem 36 to find and simplify $\frac{z}{w}$ using their polar forms as opposed to starting with $\frac{2\sqrt{3}+2i}{-1+i\sqrt{3}}$, rationalizing the denominator, and so forth.

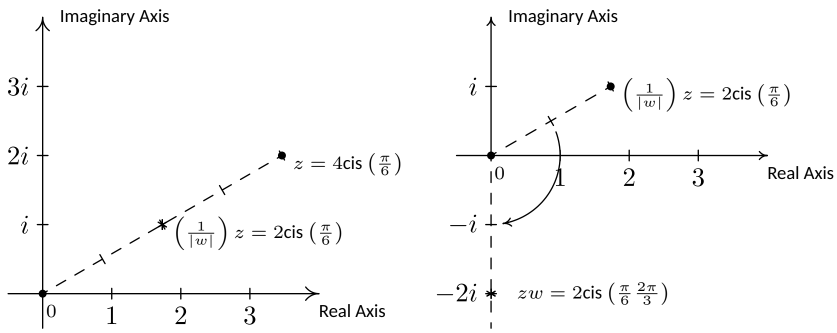
There is geometric reason for studying these polar forms and we would be derelict in our duties if we did not mention the Geometry hidden in Theorem 36. Take the product rule, for instance. If $z = |z|\operatorname{cis}(\alpha)$ and $w = |w|\operatorname{cis}(\beta)$, the formula $zw = |z|\,|w|\operatorname{cis}(\alpha+\beta)$ can be viewed geometrically as a two step process. The multiplication of $|z|$ by $|w|$ can be interpreted as magnifying the distance $|z|$ from z to 0, by the factor $|w|$. (Assuming $|w| > 1$.) Adding the argument of w to the argument of z can be interpreted geometrically as a rotation of β radians counter-clockwise. (Assuming $\beta > 0$.) Focusing on z and w from Example 148, we can arrive at the product zw by plotting z, doubling its distance from 0 (since $|w| = 2$), and rotating $\frac{2\pi}{3}$ radians counter-clockwise. The sequence of diagrams in Figure 7.23 below attempts to describe this process geometrically.



Multiplying z by $|w| = 2$. Rotating counter-clockwise by $\text{Arg}(w) = \frac{2\pi}{3}$ radians.

Figure 7.23: Visualizing zw for $z = 4\text{ cis}\left(\frac{\pi}{6}\right)$ and $w = 2\text{ cis}\left(\frac{2\pi}{3}\right)$.

We may also visualize division similarly. Here, the formula $\frac{z}{w} = \frac{|z|}{|w|}\text{ cis}(\alpha - \beta)$ may be interpreted as shrinking (again, assuming $|w| > 1$) the distance from 0 to z by the factor $|w|$, followed up by a *clockwise* rotation (again, assuming $\beta > 0$) of β radians. In the case of z and w from Example 148, we arrive at $\frac{z}{w}$ by first halving the distance from 0 to z, then rotating clockwise $\frac{2\pi}{3}$ radians.



Dividing z by $|w| = 2$. Rotating clockwise by $\text{Arg}(w) = \frac{2\pi}{3}$ radians.

Figure 7.24: Visualizing $\dfrac{z}{w}$ for $z = 4\text{ cis}\left(\frac{\pi}{6}\right)$ and $w = 2\text{ cis}\left(\frac{2\pi}{3}\right)$.

Our last goal of the section is to reverse DeMoivre's Theorem to extract roots of complex numbers.

Definition 57 Complex n^{th} roots

Let z and w be complex numbers. If there is a natural number n such that $w^n = z$, then w is an n^{th} **root** of z.

Recall that when taking an *even* root of a positive *real* number, there are two possible values: a positive root and a negative root. The principal root is taken to be the positive value. On the other hand, for *odd* roots of real numbers, there is only ever one possible value.

Unlike with real numbers, we do not specify one particular *prinicpal* n^{th} root, hence the use of the indefinite article 'an' as in 'an n^{th} root of z'. Using this definition, both 4 and -4 are square roots of 16, while $\sqrt{16}$ means the principal square root of 16 as in $\sqrt{16} = 4$. Suppose we wish to find all complex third (cube) roots of 8. Algebraically, we are trying to solve $w^3 = 8$. We know that there is only one *real* solution to this equation, namely $w = \sqrt[3]{8} = 2$, but if we take the time to rewrite this equation as $w^3 - 8 = 0$ and factor, we get $(w - 2)\left(w^2 + 2w + 4\right) = 0$. The quadratic factor gives two more cube roots $w = -1 \pm i\sqrt{3}$, for a total of three cube roots of 8. In accordance with Theorem 32, since the degree of $p(w) = w^3 - 8$ is three, there are three complex zeros, counting multiplicity. Since we have found three distinct zeros, we know these are all of the zeros, so there are exactly three distinct cube roots of 8. Let us now solve this same problem using the machinery developed in this section. To do so, we express $z = 8$ in polar form. Since $z = 8$ lies 8 units away on the positive real axis, we get $z = 8\,\text{cis}(0)$. If we let $w = |w|\,\text{cis}(\alpha)$ be a polar form of w, the equation $w^3 = 8$ becomes

$$\begin{aligned} w^3 &= 8 \\ (|w|\,\text{cis}(\alpha))^3 &= 8\,\text{cis}(0) \\ |w|^3\,\text{cis}(3\alpha) &= 8\,\text{cis}(0) && \text{DeMoivre's Theorem} \end{aligned}$$

The complex number on the left hand side of the equation corresponds to the point with polar coordinates $\left(|w|^3, 3\alpha\right)$, while the complex number on the right hand side corresponds to the point with polar coordinates $(8, 0)$. Since $|w| \geq 0$, so is $|w|^3$, which means $\left(|w|^3, 3\alpha\right)$ and $(8, 0)$ are two polar representations corresponding to the same complex number, both with positive r values. From Section 7.2, we know $|w|^3 = 8$ and $3\alpha = 0 + 2\pi k$ for integers k. Since $|w|$ is a real number, we solve $|w|^3 = 8$ by extracting the principal cube root to get $|w| = \sqrt[3]{8} = 2$. As for α, we get $\alpha = \frac{2\pi k}{3}$ for integers k. This produces three distinct points with polar coordinates corresponding to $k = 0$, 1 and 2: specifically $(2, 0)$, $\left(2, \frac{2\pi}{3}\right)$ and $\left(2, \frac{4\pi}{3}\right)$. These correspond to the complex numbers $w_0 = 2\,\text{cis}(0)$, $w_1 = 2\,\text{cis}\left(\frac{2\pi}{3}\right)$ and $w_2 = 2\,\text{cis}\left(\frac{4\pi}{3}\right)$, respectively. Writing these out in rectangular form yields $w_0 = 2$, $w_1 = -1 + i\sqrt{3}$ and $w_2 = -1 - i\sqrt{3}$. While this process seems a tad more involved than our previous factoring approach, this procedure can be generalized to find, for example, all of the fifth roots of 32. If we start with a generic complex number in polar form $z = |z|\,\text{cis}(\theta)$ and solve $w^n = z$ in the same manner as above, we arrive at the following theorem.

Theorem 37 The n^{th} roots of a complex number

Let $z \neq 0$ be a complex number with polar form $z = r\,\text{cis}(\theta)$. For each natural number n, z has n distinct n^{th} roots, which we denote by w_0, w_1, ..., w_{n-1}, and they are given by the formula

$$w_k = \sqrt[n]{r}\,\text{cis}\left(\frac{\theta}{n} + \frac{2\pi}{n}k\right)$$

The proof of Theorem 37 breaks into to two parts: first, showing that each w_k is an n^{th} root, and second, showing that the set $\{w_k \mid k = 0, 1, \ldots, (n-1)\}$ consists of n different complex numbers. To show w_k is an n^{th} root of z, we use DeMoivre's Theorem to show $(w_k)^n = z$.

$$\begin{aligned} (w_k)^n &= \left(\sqrt[n]{r}\,\text{cis}\left(\frac{\theta}{n} + \frac{2\pi}{n}k\right)\right)^n \\ &= \left(\sqrt[n]{r}\right)^n \text{cis}\left(n \cdot \left[\frac{\theta}{n} + \frac{2\pi}{n}k\right]\right) && \text{DeMoivre's Theorem} \\ &= r\,\text{cis}\left(\theta + 2\pi k\right) \end{aligned}$$

Since k is a whole number, $\cos(\theta + 2\pi k) = \cos(\theta)$ and $\sin(\theta + 2\pi k) = \sin(\theta)$. Hence, it follows that $\operatorname{cis}(\theta + 2\pi k) = \operatorname{cis}(\theta)$, so $(w_k)^n = r\operatorname{cis}(\theta) = z$, as required. To show that the formula in Theorem 37 generates n distinct numbers, we assume $n \geq 2$ (or else there is nothing to prove) and note that the modulus of each of the w_k is the same, namely $\sqrt[n]{r}$. Therefore, the only way any two of these polar forms correspond to the same number is if their arguments are coterminal – that is, if the arguments differ by an integer multiple of 2π. Suppose k and j are whole numbers between 0 and $(n-1)$, inclusive, with $k \neq j$. Since k and j are different, let's assume for the sake of argument that $k > j$. Then $\left(\frac{\theta}{n} + \frac{2\pi}{n}k\right) - \left(\frac{\theta}{n} + \frac{2\pi}{n}j\right) = 2\pi\left(\frac{k-j}{n}\right)$. For this to be an integer multiple of 2π, $(k-j)$ must be a multiple of n. But because of the restrictions on k and j, $0 < k - j \leq n - 1$. (Think this through.) Hence, $(k-j)$ is a positive number less than n, so it cannot be a multiple of n. As a result, w_k and w_j are different complex numbers, and we are done. By Theorem 32, we know there at most n distinct solutions to $w^n = z$, and we have just found all of them. We illustrate Theorem 37 in the next example.

Example 150 Finding complex roots

Use Theorem 37 to find the following:

1. both square roots of $z = -2 + 2i\sqrt{3}$
2. the four fourth roots of $z = -16$
3. the three cube roots of $z = \sqrt{2} + i\sqrt{2}$
4. the five fifth roots of $z = 1$.

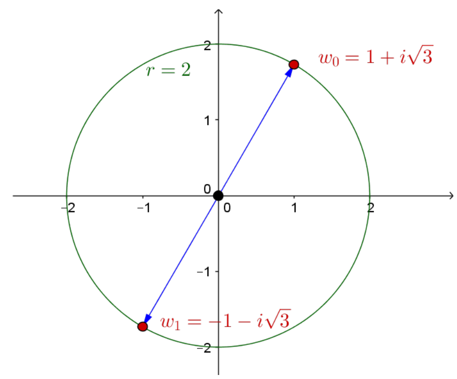


Figure 7.25: The two square roots of $z = -2 + 2\sqrt{3}i$

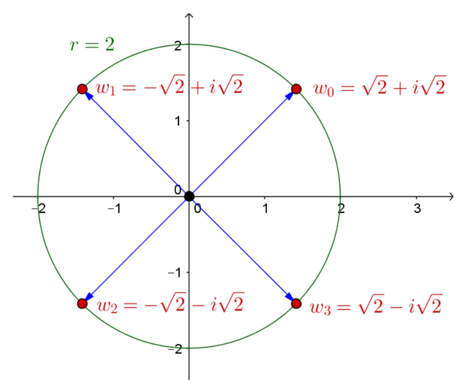


Figure 7.26: The four fourth roots of $z = -16$

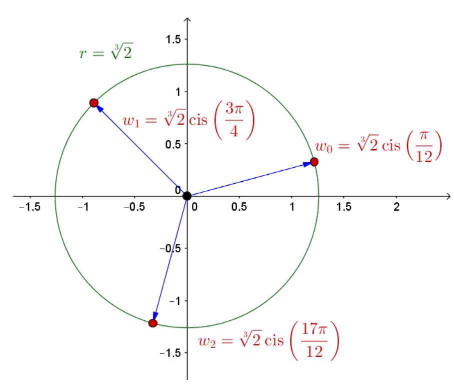


Figure 7.27: The three third roots of $z = \sqrt{2} + i\sqrt{2}$

Solution

1. We start by writing $z = -2 + 2i\sqrt{3} = 4\operatorname{cis}\left(\frac{2\pi}{3}\right)$. To use Theorem 37, we identify $r = 4$, $\theta = \frac{2\pi}{3}$ and $n = 2$. We know that z has two square roots, and in keeping with the notation in Theorem 37, we'll call them w_0 and w_1. We get $w_0 = \sqrt{4}\operatorname{cis}\left(\frac{(2\pi/3)}{2} + \frac{2\pi}{2}(0)\right) = 2\operatorname{cis}\left(\frac{\pi}{3}\right)$ and $w_1 = \sqrt{4}\operatorname{cis}\left(\frac{(2\pi/3)}{2} + \frac{2\pi}{2}(1)\right) = 2\operatorname{cis}\left(\frac{4\pi}{3}\right)$. In rectangular form, the two square roots of z are $w_0 = 1 + i\sqrt{3}$ and $w_1 = -1 - i\sqrt{3}$. We can check our answers by squaring them and showing that we get $z = -2 + 2i\sqrt{3}$. We've plotted the position of the two square roots along the circle $r = 2$ in Figure 7.25.

2. Proceeding as above, we get $z = -16 = 16\operatorname{cis}(\pi)$. With $r = 16$, $\theta = \pi$ and $n = 4$, we get the four fourth roots of z to be $w_0 = \sqrt[4]{16}\operatorname{cis}\left(\frac{\pi}{4} + \frac{2\pi}{4}(0)\right) = 2\operatorname{cis}\left(\frac{\pi}{4}\right)$, $w_1 = \sqrt[4]{16}\operatorname{cis}\left(\frac{\pi}{4} + \frac{2\pi}{4}(1)\right) = 2\operatorname{cis}\left(\frac{3\pi}{4}\right)$, $w_2 = \sqrt[4]{16}\operatorname{cis}\left(\frac{\pi}{4} + \frac{2\pi}{4}(2)\right) = 2\operatorname{cis}\left(\frac{5\pi}{4}\right)$ and $w_3 = \sqrt[4]{16}\operatorname{cis}\left(\frac{\pi}{4} + \frac{2\pi}{4}(3)\right) = 2\operatorname{cis}\left(\frac{7\pi}{4}\right)$. Converting these to rectangular form gives $w_0 = \sqrt{2} + i\sqrt{2}$, $w_1 = -\sqrt{2} + i\sqrt{2}$, $w_2 = -\sqrt{2} - i\sqrt{2}$ and $w_3 = \sqrt{2} - i\sqrt{2}$. We've plotted the four roots in Figure 7.26. Note how the roots are placed symmetrically about the circle $r = 2$.

3. For $z = \sqrt{2} + i\sqrt{2}$, we have $z = 2\operatorname{cis}\left(\frac{\pi}{4}\right)$. With $r = 2$, $\theta = \frac{\pi}{4}$ and $n = 3$ the usual computations yield $w_0 = \sqrt[3]{2}\operatorname{cis}\left(\frac{\pi}{12}\right)$, $w_1 = \sqrt[3]{2}\operatorname{cis}\left(\frac{9\pi}{12}\right) = \sqrt[3]{2}\operatorname{cis}\left(\frac{3\pi}{4}\right)$ and $w_2 = \sqrt[3]{2}\operatorname{cis}\left(\frac{17\pi}{12}\right)$. If we were to convert these to rectangular form, we would need to use either the Sum and Difference Identities or the Half-Angle Identities to evaluate w_0 and w_2. Since we are not explicitly told to do so, we leave this as a good, but messy, exercise, and plot the points in Figure 7.27.

4. To find the five fifth roots of 1, we write $1 = 1\,\mathrm{cis}(0)$. We have $r = 1$, $\theta = 0$ and $n = 5$. Since $\sqrt[5]{1} = 1$, the roots are $w_0 = \mathrm{cis}(0) = 1$, $w_1 = \mathrm{cis}\left(\frac{2\pi}{5}\right)$, $w_2 = \mathrm{cis}\left(\frac{4\pi}{5}\right)$, $w_3 = \mathrm{cis}\left(\frac{6\pi}{5}\right)$ and $w_4 = \mathrm{cis}\left(\frac{8\pi}{5}\right)$. The situation here is even graver than in the previous example, since we have not developed any identities to help us determine the cosine or sine of $\frac{2\pi}{5}$. At this stage, we could approximate our answers using a calculator, and we leave this as an exercise. Once more, we plot the roots, which in this case all lie on the unit circle.

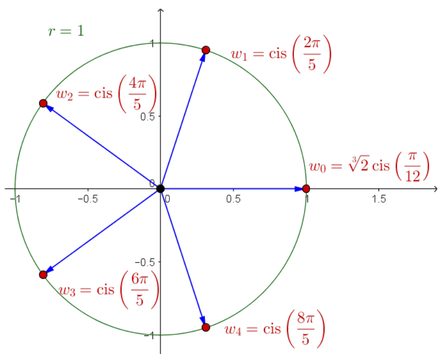


Figure 7.28: The five fifth roots of 1

Notice the geometric interpretation given in Figures 7.25-7.28. Essentially, Theorem 37 says that to find the n^{th} roots of a complex number, we first take the n^{th} root of the modulus and divide the argument by n. This gives the first root w_0. Each successive root is found by adding $\frac{2\pi}{n}$ to the argument, which amounts to rotating w_0 by $\frac{2\pi}{n}$ radians. This results in n roots, spaced equally around the complex plane.

We have only glimpsed at the beauty of the complex numbers in this section. The complex plane is without a doubt one of the most important mathematical constructs ever devised. Coupled with Calculus, it is the venue for incredibly important Science and Engineering applications. For now, the following exercises will have to suffice.

Exercises 7.3

Problems

In Exercises 1 – 20, find a polar representation for the complex number z and then identify $\mathrm{Re}(z)$, $\mathrm{Im}(z)$, $|z|$, $\arg(z)$ and $\mathrm{Arg}(z)$.

1. $z = 9 + 9i$
2. $z = 5 + 5i\sqrt{3}$
3. $z = 6i$
4. $z = -3\sqrt{2} + 3i\sqrt{2}$
5. $z = -6\sqrt{3} + 6i$
6. $z = -2$
7. $z = -\frac{\sqrt{3}}{2} - \frac{1}{2}i$
8. $z = -3 - 3i$
9. $z = -5i$
10. $z = 2\sqrt{2} - 2i\sqrt{2}$
11. $z = 6$
12. $z = i\sqrt[3]{7}$
13. $z = 3 + 4i$
14. $z = \sqrt{2} + i$
15. $z = -7 + 24i$
16. $z = -2 + 6i$
17. $z = -12 - 5i$
18. $z = -5 - 2i$
19. $z = 4 - 2i$
20. $z = 1 - 3i$

In Exercises 21 – 40, find the rectangular form of the given complex number. Use whatever identities are necessary to find the exact values.

21. $z = 6\,\mathrm{cis}(0)$
22. $z = 2\,\mathrm{cis}\left(\frac{\pi}{6}\right)$
23. $z = 7\sqrt{2}\,\mathrm{cis}\left(\frac{\pi}{4}\right)$
24. $z = 3\,\mathrm{cis}\left(\frac{\pi}{2}\right)$
25. $z = 4\,\mathrm{cis}\left(\frac{2\pi}{3}\right)$
26. $z = \sqrt{6}\,\mathrm{cis}\left(\frac{3\pi}{4}\right)$
27. $z = 9\,\mathrm{cis}\left(\pi\right)$
28. $z = 3\,\mathrm{cis}\left(\frac{4\pi}{3}\right)$
29. $z = 7\,\mathrm{cis}\left(-\frac{3\pi}{4}\right)$
30. $z = \sqrt{13}\,\mathrm{cis}\left(\frac{3\pi}{2}\right)$
31. $z = \frac{1}{2}\,\mathrm{cis}\left(\frac{7\pi}{4}\right)$
32. $z = 12\,\mathrm{cis}\left(-\frac{\pi}{3}\right)$
33. $z = 8\,\mathrm{cis}\left(\frac{\pi}{12}\right)$
34. $z = 2\,\mathrm{cis}\left(\frac{7\pi}{8}\right)$
35. $z = 5\,\mathrm{cis}\left(\arctan\left(\frac{4}{3}\right)\right)$
36. $z = \sqrt{10}\,\mathrm{cis}\left(\arctan\left(\frac{1}{3}\right)\right)$
37. $z = 15\,\mathrm{cis}\left(\arctan\left(-2\right)\right)$
38. $z = \sqrt{3}\left(\arctan\left(-\sqrt{2}\right)\right)$
39. $z = 50\,\mathrm{cis}\left(\pi - \arctan\left(\frac{7}{24}\right)\right)$
40. $z = \frac{1}{2}\,\mathrm{cis}\left(\pi + \arctan\left(\frac{5}{12}\right)\right)$

In Exercises 41 – 52, use $z = -\frac{3\sqrt{3}}{2} + \frac{3}{2}i$ and $w = 3\sqrt{2} - 3i\sqrt{2}$ to compute the quantity. Express your answers in polar form using the principal argument.

41. zw
42. $\frac{z}{w}$
43. $\frac{w}{z}$

44. z^4

45. w^3

46. z^5w^2

47. z^3w^2

48. $\dfrac{z^2}{w}$

49. $\dfrac{w}{z^2}$

50. $\dfrac{z^3}{w^2}$

51. $\dfrac{w^2}{z^3}$

52. $\left(\dfrac{w}{z}\right)^6$

In Exercises 53 – 64, use DeMoivre's Theorem to find the indicated power of the given complex number. Express your final answers in rectangular form.

53. $\left(-2+2i\sqrt{3}\right)^3$

54. $(-\sqrt{3}-i)^3$

55. $(-3+3i)^4$

56. $(\sqrt{3}+i)^4$

57. $\left(\dfrac{5}{2}+\dfrac{5}{2}i\right)^3$

58. $\left(-\dfrac{1}{2}-\dfrac{\sqrt{3}}{2}i\right)^6$

59. $\left(\dfrac{3}{2}-\dfrac{3}{2}i\right)^3$

60. $\left(\dfrac{\sqrt{3}}{3}-\dfrac{1}{3}i\right)^4$

61. $\left(\dfrac{\sqrt{2}}{2}+\dfrac{\sqrt{2}}{2}i\right)^4$

62. $(2+2i)^5$

63. $(\sqrt{3}-i)^5$

64. $(1-i)^8$

In Exercises 65 – 76, find the indicated complex roots. Express your answers in polar form and then convert them into rectangular form.

65. the two square roots of $z = 4i$

66. the two square roots of $z = -25i$

67. the two square roots of $z = 1 + i\sqrt{3}$

68. the two square roots of $\frac{5}{2} - \frac{5\sqrt{3}}{2}i$

69. the three cube roots of $z = 64$

70. the three cube roots of $z = -125$

71. the three cube roots of $z = i$

72. the three cube roots of $z = -8i$

73. the four fourth roots of $z = 16$

74. the four fourth roots of $z = -81$

75. the six sixth roots of $z = 64$

76. the six sixth roots of $z = -729$

77. Use trigonometric identities to express the three cube roots of $z = \sqrt{2} + i\sqrt{2}$ in rectangular form. (See Example 149, number 3.)

78. Use a calculator or computer to approximate the five fifth roots of 1. (See Example 149, number 4.)

79. Complete the proof of Theorem 34 by showing that if $w \neq 0$ than $\left|\frac{1}{w}\right| = \frac{1}{|w|}$.

80. Recall from Section 7.1 that given a complex number $z = a + bi$ its complex conjugate, denoted $\bar{z}$, is given by $\bar{z} = a - bi$.
    (a) Prove that $|\bar{z}| = |z|$.
    (b) Prove that $|z| = \sqrt{z\bar{z}}$
    (c) Show that $\text{Re}(z) = \dfrac{z+\bar{z}}{2}$ and $\text{Im}(z) = \dfrac{z-\bar{z}}{2i}$
    (d) Show that if $\theta \in \arg(z)$ then $-\theta \in \arg(\bar{z})$. Interpret this result geometrically.
    (e) Is it always true that $\text{Arg}(\bar{z}) = -\text{Arg}(z)$?

81. Given any natural number $n \geq 2$, the n complex n^{th} roots of the number $z = 1$ are called the **n^{th} Roots of Unity**. In the following exercises, assume that n is a fixed, but arbitrary, natural number such that $n \geq 2$.
    (a) Show that $w = 1$ is an n^{th} root of unity.
    (b) Show that if both w_j and w_k are n^{th} roots of unity then so is their product w_jw_k.
    (c) Show that if w_j is an n^{th} root of unity then there exists another n^{th} root of unity $w_{j'}$ such that $w_jw_{j'} = 1$. Hint: If $w_j = \text{cis}(\theta)$ let $w_{j'} = \text{cis}(2\pi - \theta)$. You'll need to verify that $w_{j'} = \text{cis}(2\pi - \theta)$ is indeed an n^{th} root of unity.

82. Another way to express the polar form of a complex number is to use the exponential function. For real numbers t, Euler's Formula defines $e^{it} = \cos(t) + i\sin(t)$.

    (a) Use Theorem 36 to show that $e^{ix}e^{iy} = e^{i(x+y)}$ for all real numbers x and y.

    (b) Use Theorem 36 to show that $\left(e^{ix}\right)^n = e^{i(nx)}$ for any real number x and any natural number n.

    (c) Use Theorem 36 to show that $\dfrac{e^{ix}}{e^{iy}} = e^{i(x-y)}$ for all real numbers x and y.

    (d) If $z = r\,\mathrm{cis}(\theta)$ is the polar form of z, show that $z = re^{it}$ where $\theta = t$ radians.

    (e) Show that $e^{i\pi} + 1 = 0$. (This famous equation relates the five most important constants in all of Mathematics with the three most fundamental operations in Mathematics.)

    (f) Show that $\cos(t) = \dfrac{e^{it} + e^{-it}}{2}$ and that $\sin(t) = \dfrac{e^{it} - e^{-it}}{2i}$ for all real numbers t.

8: Eigenvalues and Eigenvectors

We have often explored new ideas in matrix algebra by making connections to our previous algebraic experience. Adding two numbers, $x+y$, led us to adding vectors $\vec{x}+\vec{y}$ and adding matrices $A+B$. We explored multiplication, which then led us to solving the matrix equation $A\vec{x}=\vec{b}$, which was reminiscent of solving the algebra equation $ax=b$.

This chapter is motivated by another analogy. Consider: when we multiply an unknown number x by another number such as 5, what do we know about the result? Unless, $x=0$, we know that in some sense $5x$ will be "5 times bigger than x." Applying this to vectors, we would readily agree that $5\vec{x}$ gives a vector that is "5 times bigger than $\vec{x}$." Each entry in $\vec{x}$ is multiplied by 5.

Within the matrix algebra context, though, we have two types of multiplication: scalar and matrix multiplication. What happens to $\vec{x}$ when we multiply it by a matrix A? Our first response is likely along the lines of "You just get another vector. There is no definable relationship." We might wonder if there is ever the case where a matrix – vector multiplication is very similar to a scalar – vector multiplication. That is, do we ever have the case where $A\vec{x}=a\vec{x}$, where a is some scalar? That is the motivating question of this chapter.

8.1 Eigenvalues and Eigenvectors

AS YOU READ . . .

1. T/F: Given any matrix A, we can always find a vector $\vec{x}$ where $A\vec{x}=\vec{x}$.
2. When is the zero vector an eigenvector for a matrix?
3. If $\vec{v}$ is an eigenvector of a matrix A with eigenvalue of 2, then what is $A\vec{v}$?
4. T/F: If A is a 5×5 matrix, to find the eigenvalues of A, we would need to find the roots of a 5^{th} degree polynomial.

We start by considering the matrix A and vector $\vec{x}$ as given below. (Recall this matrix and vector were used in Example 66 on page 108.)

$$A=\begin{bmatrix}1&4\\2&3\end{bmatrix}\qquad \vec{x}=\begin{bmatrix}1\\1\end{bmatrix}$$

Multiplying $A\vec{x}$ gives:

$$\begin{aligned}A\vec{x}&=\begin{bmatrix}1&4\\2&3\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\\&=\begin{bmatrix}5\\5\end{bmatrix}\\&=5\begin{bmatrix}1\\1\end{bmatrix}!\end{aligned}$$

Wow! It looks like multiplying $A\vec{x}$ is the same as $5\vec{x}$! This makes us wonder lots of things: Is this the only case in the world where something like this happens? (Probably not.) Is A somehow a special matrix, and $A\vec{x}=5\vec{x}$ for any vector

$\vec{x}$ we pick? (Probably not.) Or maybe $\vec{x}$ was a special vector, and no matter what 2×2 matrix A we picked, we would have $A\vec{x} = 5\vec{x}$. (Again, probably not.)

A more likely explanation is this: given the matrix A, the number 5 and the vector $\vec{x}$ formed a special pair that happened to work together in a nice way. It is then natural to wonder if other "special" pairs exist. For instance, could we find a vector $\vec{x}$ where $A\vec{x} = 3\vec{x}$?

This equation is hard to solve *at first*; we are not used to matrix equations where $\vec{x}$ appears on both sides of "$=$." Therefore we put off solving this for just a moment to state a definition and make a few comments.

Definition 58 Eigenvalues and Eigenvectors

Let A be an $n \times n$ matrix, $\vec{x}$ a nonzero $n \times 1$ column vector and λ a scalar. If

$$A\vec{x} = \lambda\vec{x},$$

then $\vec{x}$ is an *eigenvector* of A and λ is an *eigenvalue* of A.

The word "eigen" is German for "proper" or "characteristic." Therefore, an *eigenvector* of A is a "characteristic vector of A." This vector tells us something about A.

Why do we use the Greek letter λ (lambda)? It is pure tradition. Above, we used a to represent the unknown scalar, since we are used to that notation. We now switch to λ because that is how everyone else does it. (An example of mathematical peer pressure.) Don't get hung up on this; λ is just a number.

Note that our definition requires that A be a square matrix. If A isn't square then $A\vec{x}$ and $\lambda\vec{x}$ will have different sizes, and so they cannot be equal. Also note that $\vec{x}$ must be nonzero. Why? What if $\vec{x} = \vec{0}$? Then *no matter what* λ is, $A\vec{x} = \lambda\vec{x}$. This would then imply that *every number* is an eigenvalue; if every number is an eigenvalue, then we wouldn't need a definition for it. (Recall note 21 on page 203.) Therefore we specify that $\vec{x} \neq \vec{0}$.

Our last comment before trying to find eigenvalues and eigenvectors for given matrices deals with "why we care." Did we stumble upon a mathematical curiosity, or does this somehow help us build better bridges, heal the sick, send astronauts into orbit, design optical equipment, and understand quantum mechanics? The answer, of course, is "Yes." (Except for the "understand quantum mechanics" part. Nobody truly understands that stuff; they just *probably* understand it.) This is a wonderful topic in and of itself: we need no external application to appreciate its worth. At the same time, it has many, many applications to "the real world." A simple Internet seach on "applications of eigenvalues" with confirm this.

Back to our math. Given a square matrix A, we want to find a nonzero vector $\vec{x}$ and a scalar λ such that $A\vec{x} = \lambda\vec{x}$. We will solve this using the skills we developed in Chapter 3.

$$\begin{aligned} A\vec{x} &= \lambda\vec{x} && \text{original equation} \\ A\vec{x} - \lambda\vec{x} &= \vec{0} && \text{subtract } \lambda\vec{x} \text{ from both sides} \\ (A - \lambda I)\vec{x} &= \vec{0} && \text{factor out } \vec{x} \end{aligned}$$

Think about this last factorization. We are likely tempted to say

$$A\vec{x} - \lambda\vec{x} = (A - \lambda)\vec{x},$$

but this really doesn't make sense. After all, what does "a matrix minus a number" mean? We need the identity matrix in order for this to be logical.

Let us now think about the equation $(A - \lambda I)\vec{x} = \vec{0}$. While it looks complicated, it really is just matrix equation of the type we solved in Section 5.1. We are just trying to solve $B\vec{x} = \vec{0}$, where $B = (A - \lambda I)$.

We know from our previous work that this type of equation always has a solution, namely, $\vec{x} = \vec{0}$. (Recall this is a *homogeneous* system of equations.) However, we want $\vec{x}$ to be an eigenvector and, by the definition, eigenvectors cannot be $\vec{0}$.

This means that we want solutions to $(A - \lambda I)\vec{x} = \vec{0}$ other than $\vec{x} = \vec{0}$. Recall that Theorem 17 says that if the matrix $(A - \lambda I)$ is invertible, then the *only* solution to $(A - \lambda I)\vec{x} = \vec{0}$ is $\vec{x} = \vec{0}$. Therefore, in order to have other solutions, we need $(A - \lambda I)$ to not be invertible.

Finally, recall from Theorem 25 that noninvertible matrices all have a determinant of 0. Therefore, if we want to find eigenvalues λ and eigenvectors $\vec{x}$, we need $\det(A - \lambda I) = 0$.

Let's start our practice of this theory by finding λ such that $\det(A - \lambda I) = 0$; that is, let's find the eigenvalues of a matrix.

Example 151 **Computing the eigenvalues of a matrix**

Find the eigenvalues of A, that is, find λ such that $\det(A - \lambda I) = 0$, where

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

Solution (Note that this is the matrix we used at the beginning of this section.) First, we write out what $A - \lambda I$ is:

$$\begin{aligned} A - \lambda I &= \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \\ &= \begin{bmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{bmatrix} \end{aligned}$$

Therefore,

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{vmatrix} \\ &= (1-\lambda)(3-\lambda) - 8 \\ &= \lambda^2 - 4\lambda - 5 \end{aligned}$$

Since we want $\det(A - \lambda I) = 0$, we want $\lambda^2 - 4\lambda - 5 = 0$. This is a simple quadratic equation that is easy to factor:

$$\begin{aligned} \lambda^2 - 4\lambda - 5 &= 0 \\ (\lambda - 5)(\lambda + 1) &= 0 \\ \lambda &= -1,\ 5 \end{aligned}$$

According to our above work, $\det(A - \lambda I) = 0$ when $\lambda = -1,\ 5$. Thus, the eigenvalues of A are -1 and 5.

Earlier, when looking at the same matrix as used in our example, we wondered if we could find a vector $\vec{x}$ such that $A\vec{x} = 3\vec{x}$. According to this example, the answer is "No." With this matrix A, the only values of λ that work are -1 and 5.

Let's restate the above in a different way: It is pointless to try to find $\vec{x}$ where $A\vec{x} = 3\vec{x}$, for there is no such $\vec{x}$. There are only 2 equations of this form that have a solution, namely

$$A\vec{x} = -\vec{x} \qquad \text{and} \qquad A\vec{x} = 5\vec{x}.$$

As we introduced this section, we gave a vector $\vec{x}$ such that $A\vec{x} = 5\vec{x}$. Is this the only one? Let's find out while calling our work an example; this will amount to finding the eigenvectors of A that correspond to the eigenvector of 5.

Example 152 **Computing an eigenvector corresponding to a given eigenvalue**

Find $\vec{x}$ such that $A\vec{x} = 5\vec{x}$, where

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

Solution Recall that our algebra from before showed that if

$$A\vec{x} = \lambda\vec{x} \quad \text{then} \quad (A - \lambda I)\vec{x} = \vec{0}.$$

Therefore, we need to solve the equation $(A - \lambda I)\vec{x} = \vec{0}$ for $\vec{x}$ when $\lambda = 5$.

$$\begin{aligned} A - 5I &= \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} - 5\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -4 & 4 \\ 2 & -2 \end{bmatrix} \end{aligned}$$

To solve $(A-5I)\vec{x} = \vec{0}$, we form the augmented matrix and put it into reduced row echelon form:

$$\begin{bmatrix} -4 & 4 & 0 \\ 2 & -2 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus

$$\begin{aligned} x_1 &= x_2 \\ x_2 &\text{ is free} \end{aligned}$$

and

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

We have infinite solutions to the equation $A\vec{x} = 5\vec{x}$; any nonzero scalar multiple of the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is a solution. We can do a few examples to confirm this:

$$\begin{aligned} \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ 2 \end{bmatrix} &= \begin{bmatrix} 10 \\ 10 \end{bmatrix} = 5\begin{bmatrix} 2 \\ 2 \end{bmatrix}; \\ \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 7 \\ 7 \end{bmatrix} &= \begin{bmatrix} 35 \\ 35 \end{bmatrix} = 5\begin{bmatrix} 7 \\ 7 \end{bmatrix}; \\ \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} -3 \\ -3 \end{bmatrix} &= \begin{bmatrix} -15 \\ -15 \end{bmatrix} = 5\begin{bmatrix} -3 \\ -3 \end{bmatrix}. \end{aligned}$$

Our method of finding the eigenvalues of a matrix A boils down to determining which values of λ give the matrix $(A - \lambda I)$ a determinant of 0. In computing

$\det(A - \lambda I)$, we get a polynomial in λ whose roots are the eigenvalues of A. This polynomial is important and so it gets its own name.

Definition 59 Characteristic Polynomial

Let A be an $n \times n$ matrix. The *characteristic polynomial* of A is the n^{th} degree polynomial $p(\lambda) = \det(A - \lambda I)$.

Our definition just states *what* the characteristic polynomial is. We know from our work so far *why* we care: the roots of the characteristic polynomial of an $n \times n$ matrix A are the eigenvalues of A.

In Examples 150 and 151, we found eigenvalues and eigenvectors, respectively, of a given matrix. That is, given a matrix A, we found values λ and vectors $\vec{x}$ such that $A\vec{x} = \lambda\vec{x}$. The steps that follow outline the general procedure for finding eigenvalues and eigenvectors; we'll follow this up with some examples.

Key Idea 25 Finding Eigenvalues and Eigenvectors

Let A be an $n \times n$ matrix.

1. To find the eigenvalues of A, compute $p(\lambda)$, the characteristic polynomial of A, set it equal to 0, then solve for λ.
2. To find the eigenvectors of A, *for each eigenvalue* solve the homogeneous system $(A - \lambda I)\vec{x} = \vec{0}$.

Example 153 Computing eigenvalues and eigenvectors

Find the eigenvalues of A, and for each eigenvalue, find an eigenvector where

$$A = \begin{bmatrix} -3 & 15 \\ 3 & 9 \end{bmatrix}.$$

Solution To find the eigenvalues, we must compute $\det(A - \lambda I)$ and set it equal to 0.

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -3-\lambda & 15 \\ 3 & 9-\lambda \end{vmatrix} \\ &= (-3-\lambda)(9-\lambda) - 45 \\ &= \lambda^2 - 6\lambda - 27 - 45 \\ &= \lambda^2 - 6\lambda - 72 \\ &= (\lambda - 12)(\lambda + 6) \end{aligned}$$

Therefore, $\det(A - \lambda I) = 0$ when $\lambda = -6$ and 12; these are our eigenvalues. (We should note that $p(\lambda) = \lambda^2 - 6\lambda - 72$ is our characteristic polynomial.) It sometimes helps to give them "names," so we'll say $\lambda_1 = -6$ and $\lambda_2 = 12$. Now we find eigenvectors.

For $\lambda_1 = -6$:

We need to solve the equation $(A - (-6)I)\vec{x} = \vec{0}$. To do this, we form the appropriate augmented matrix and put it into reduced row echelon form.

$$\begin{bmatrix} 3 & 15 & 0 \\ 3 & 15 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 5 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Our solution is

$$\begin{aligned} x_1 &= -5x_2 \\ x_2 &\text{ is free;} \end{aligned}$$

in vector form, we have

$$\vec{x} = x_2 \begin{bmatrix} -5 \\ 1 \end{bmatrix}.$$

We may pick any nonzero value for x_2 to get an eigenvector; a simple option is $x_2 = 1$. Thus we have the eigenvector

$$\vec{x_1} = \begin{bmatrix} -5 \\ 1 \end{bmatrix}.$$

(We used the notation $\vec{x_1}$ to associate this eigenvector with the eigenvalue λ_1.)

We now repeat this process to find an eigenvector for $\lambda_2 = 12$:
In solving $(A - 12I)\vec{x} = \vec{0}$, we find

$$\begin{bmatrix} -15 & 15 & 0 \\ 3 & -3 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In vector form, we have

$$\vec{x} = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Again, we may pick any nonzero value for x_2, and so we choose $x_2 = 1$. Thus an eigenvector for λ_2 is

$$\vec{x_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

To summarize, we have:

$$\text{eigenvalue } \lambda_1 = -6 \text{ with eigenvector } \vec{x_1} = \begin{bmatrix} -5 \\ 1 \end{bmatrix}$$

and

$$\text{eigenvalue } \lambda_2 = 12 \text{ with eigenvector } \vec{x_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

We should take a moment and check our work: is it true that $A\vec{x_1} = \lambda_1\vec{x_1}$?

$$\begin{aligned} A\vec{x_1} &= \begin{bmatrix} -3 & 15 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} -5 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 30 \\ -6 \end{bmatrix} \\ &= (-6) \begin{bmatrix} -5 \\ 1 \end{bmatrix} \\ &= \lambda_1 \vec{x_1}. \end{aligned}$$

Yes; it appears we have truly found an eigenvalue/eigenvector pair for the matrix A.

Let's do another example.

Example 154 **Computing eigenvalues and eigenvectors**

Let $A = \begin{bmatrix} -3 & 0 \\ 5 & 1 \end{bmatrix}$. Find the eigenvalues of A and an eigenvector for each eigenvalue.

Solution We first compute the characteristic polynomial, set it equal to 0, then solve for λ.

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -3-\lambda & 0 \\ 5 & 1-\lambda \end{vmatrix} \\ &= (-3-\lambda)(1-\lambda) \end{aligned}$$

From this, we see that $\det(A - \lambda I) = 0$ when $\lambda = -3, 1$. We'll set $\lambda_1 = -3$ and $\lambda_2 = 1$.

Finding an eigenvector for λ_1:
We solve $(A - (-3)I)\vec{x} = \vec{0}$ for $\vec{x}$ by row reducing the appropriate matrix:

$$\begin{bmatrix} 0 & 0 & 0 \\ 5 & 4 & 0 \end{bmatrix} \quad \xrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 5/4 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Our solution, in vector form, is

$$\vec{x} = x_2 \begin{bmatrix} -5/4 \\ 1 \end{bmatrix}.$$

Again, we can pick any nonzero value for x_2; a nice choice would eliminate the fraction. Therefore we pick $x_2 = 4$, and find

$$\vec{x_1} = \begin{bmatrix} -5 \\ 4 \end{bmatrix}.$$

Finding an eigenvector for λ_2:
We solve $(A - (1)I)\vec{x} = \vec{0}$ for $\vec{x}$ by row reducing the appropriate matrix:

$$\begin{bmatrix} -4 & 0 & 0 \\ 5 & 0 & 0 \end{bmatrix} \quad \xrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

We've seen a matrix like this before, but we may need a bit of a refreshing. (See page 165. Our future need of knowing how to handle this situation is foretold in note 41.) Our first row tells us that $x_1 = 0$, and we see that no rows/equations involve x_2. We conclude that x_2 is free. Therefore, our solution, in vector form, is

$$\vec{x} = x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

We pick $x_2 = 1$, and find

$$\vec{x_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

To summarize, we have:

$$\text{eigenvalue } \lambda_1 = -3 \text{ with eigenvector } \vec{x_1} = \begin{bmatrix} -5 \\ 4 \end{bmatrix}$$

and

$$\text{eigenvalue } \lambda_2 = 1 \text{ with eigenvector } \vec{x_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

So far, our examples have involved 2×2 matrices. Let's do an example with a 3×3 matrix.

Example 155 **Eigenvalues and eigenvectors for a 3×3 matrix**
Find the eigenvalues of A, and for each eigenvalue, give one eigenvector, where

$$A = \begin{bmatrix} -7 & -2 & 10 \\ -3 & 2 & 3 \\ -6 & -2 & 9 \end{bmatrix}.$$

Solution We first compute the characteristic polynomial, set it equal to 0, then solve for λ. A warning: this process is rather long. We'll use cofactor expansion along the first row; don't get bogged down with the arithmetic that comes from each step; just try to get the basic idea of what was done from step to step.

You should have learned how to do this in high school. As a reminder, possible roots can be found by factoring the constant term (in this case, -6) of the polynomial. That is, the roots of this equation could be $\pm1, \pm2, \pm3$ and ±6. That's 12 things to check.
One could also graph this polynomial to find the roots. Graphing will show us that $\lambda = 3$ *looks* like a root, and a simple calculation will confirm that it is.
If your factoring skills are a bit rusty, you may want to consult the resources on the *Math Basics* Moodle page.

$$\begin{aligned}
\det(A - \lambda I) &= \begin{vmatrix} -7-\lambda & -2 & 10 \\ -3 & 2-\lambda & 3 \\ -6 & -2 & 9-\lambda \end{vmatrix} \\
&= (-7-\lambda)\begin{vmatrix} 2-\lambda & 3 \\ -2 & 9-\lambda \end{vmatrix} - (-2)\begin{vmatrix} -3 & 3 \\ -6 & 9-\lambda \end{vmatrix} + 10\begin{vmatrix} -3 & 2-\lambda \\ -6 & -2 \end{vmatrix} \\
&= (-7-\lambda)(\lambda^2 - 11\lambda + 24) + 2(3\lambda - 9) + 10(-6\lambda + 18) \\
&= -\lambda^3 + 4\lambda^2 - \lambda - 6 \\
&= -(\lambda + 1)(\lambda - 2)(\lambda - 3)
\end{aligned}$$

In the last step we factored the characteristic polynomial $-\lambda^3 + 4\lambda^2 - \lambda - 6$. Factoring polynomials of degree > 2 is not trivial; we'll assume the reader has access to methods for doing this accurately.

Our eigenvalues are $\lambda_1 = -1$, $\lambda_2 = 2$ and $\lambda_3 = 3$. We now find corresponding eigenvectors.

For $\lambda_1 = -1$:

We need to solve the equation $(A - (-1)I)\vec{x} = \vec{0}$. To do this, we form the appropriate augmented matrix and put it into reduced row echelon form.

$$\begin{bmatrix} -6 & -2 & 10 & 0 \\ -3 & 3 & 3 & 0 \\ -6 & -2 & 10 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & -1.5 & 0 \\ 0 & 1 & -.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Our solution, in vector form, is

$$\vec{x} = x_3 \begin{bmatrix} 3/2 \\ 1/2 \\ 1 \end{bmatrix}.$$

We can pick any nonzero value for x_3; a nice choice would get rid of the fractions. So we'll set $x_3 = 2$ and choose $\vec{x_1} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ as our eigenvector.

For $\lambda_2 = 2$:

We need to solve the equation $(A - 2I)\vec{x} = \vec{0}$. To do this, we form the appropriate augmented matrix and put it into reduced row echelon form.

$$\begin{bmatrix} -9 & -2 & 10 & 0 \\ -3 & 0 & 3 & 0 \\ -6 & -2 & 7 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -.5 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Our solution, in vector form, is

$$\vec{x} = x_3 \begin{bmatrix} 1 \\ 1/2 \\ 1 \end{bmatrix}.$$

We can pick any nonzero value for x_3; again, a nice choice would get rid of the fractions. So we'll set $x_3 = 2$ and choose $\vec{x_2} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ as our eigenvector.

For $\lambda_3 = 3$:

We need to solve the equation $(A - 3I)\vec{x} = \vec{0}$. To do this, we form the appropriate augmented matrix and put it into reduced row echelon form.

$$\begin{bmatrix} -10 & -2 & 10 & 0 \\ -3 & -1 & 3 & 0 \\ -6 & -2 & 6 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Our solution, in vector form, is (note that $x_2 = 0$):

$$\vec{x} = x_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

We can pick any nonzero value for x_3; an easy choice is $x_3 = 1$, so $\vec{x_3} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ as our eigenvector.

To summarize, we have the following eigenvalue/eigenvector pairs:

$$\text{eigenvalue } \lambda_1 = -1 \text{ with eigenvector } \vec{x_1} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$$

$$\text{eigenvalue } \lambda_2 = 2 \text{ with eigenvector } \vec{x_2} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

$$\text{eigenvalue } \lambda_3 = 3 \text{ with eigenvector } \vec{x_3} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Let's practice once more.

Example 156 **Computing eigenvalues and eigenvectors**

Find the eigenvalues of A, and for each eigenvalue, give one eigenvector, where

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

Solution We first compute the characteristic polynomial, set it equal to 0, then solve for λ. We'll use cofactor expansion down the first column (since it has lots of zeros).

$$\begin{aligned}\det(A-\lambda I) &= \begin{vmatrix} 2-\lambda & -1 & 1 \\ 0 & 1-\lambda & 6 \\ 0 & 3 & 4-\lambda \end{vmatrix} \\ &= (2-\lambda)\begin{vmatrix} 1-\lambda & 6 \\ 3 & 4-\lambda \end{vmatrix} \\ &= (2-\lambda)(\lambda^2-5\lambda-14) \\ &= (2-\lambda)(\lambda-7)(\lambda+2)\end{aligned}$$

Notice that while the characteristic polynomial is cubic, we never actually saw a cubic; we never distributed the $(2-\lambda)$ across the quadratic. Instead, we realized that this was a factor of the cubic, and just factored the remaining quadratic. (This makes this example quite a bit simpler than the previous example.)

Our eigenvalues are $\lambda_1 = -2$, $\lambda_2 = 2$ and $\lambda_3 = 7$. We now find corresponding eigenvectors.

For $\lambda_1 = -2$:

We need to solve the equation $(A-(-2)I)\vec{x} = \vec{0}$. To do this, we form the appropriate augmented matrix and put it into reduced row echelon form.

$$\begin{bmatrix} 4 & -1 & 1 & 0 \\ 0 & 3 & 6 & 0 \\ 0 & 3 & 6 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 3/4 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Our solution, in vector form, is

$$\vec{x} = x_3 \begin{bmatrix} -3/4 \\ -2 \\ 1 \end{bmatrix}.$$

We can pick any nonzero value for x_3; a nice choice would get rid of the fractions. So we'll set $x_3 = 4$ and choose $\vec{x_1} = \begin{bmatrix} -3 \\ -8 \\ 4 \end{bmatrix}$ as our eigenvector.

For $\lambda_2 = 2$:

We need to solve the equation $(A-2I)\vec{x} = \vec{0}$. To do this, we form the appropriate augmented matrix and put it into reduced row echelon form.

$$\begin{bmatrix} 0 & -1 & 1 & 0 \\ 0 & -1 & 6 & 0 \\ 0 & 3 & 2 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

This looks funny, so we'll look remind ourselves how to solve this. The first two rows tell us that $x_2 = 0$ and $x_3 = 0$, respectively. Notice that no row/equation uses x_1; we conclude that it is free. Therefore, our solution in vector form is

$$\vec{x} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

We can pick any nonzero value for x_1; an easy choice is $x_1 = 1$ and choose $\vec{x_2} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ as our eigenvector.

For $\lambda_3 = 7$:

We need to solve the equation $(A - 7I)\vec{x} = \vec{0}$. To do this, we form the appropriate augmented matrix and put it into reduced row echelon form.

$$\begin{bmatrix} -5 & -1 & 1 & 0 \\ 0 & -6 & 6 & 0 \\ 0 & 3 & -3 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Our solution, in vector form, is (note that $x_1 = 0$):

$$\vec{x} = x_3 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

We can pick any nonzero value for x_3; an easy choice is $x_3 = 1$, so $\vec{x_3} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ as our eigenvector.

To summarize, we have the following eigenvalue/eigenvector pairs:

$$\text{eigenvalue } \lambda_1 = -2 \text{ with eigenvector } \vec{x_1} = \begin{bmatrix} -3 \\ -8 \\ 4 \end{bmatrix}$$

$$\text{eigenvalue } \lambda_2 = 2 \text{ with eigenvector } \vec{x_2} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\text{eigenvalue } \lambda_3 = 7 \text{ with eigenvector } \vec{x_3} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

In this section we have learned about a new concept: given a matrix A we can find certain values λ and vectors $\vec{x}$ where $A\vec{x} = \lambda\vec{x}$. In the next section we will continue to the pattern we have established in this text: after learning a new concept, we see how it interacts with other concepts we know about. That is, we'll look for connections between eigenvalues and eigenvectors and things like the inverse, determinants, the trace, the transpose, etc.

Exercises 8.1

Problems

In Exercises 1 – 6, a matrix A and one of its eigenvectors are given. Find the eigenvalue of A for the given eigenvector.

1. $A = \begin{bmatrix} 9 & 8 \\ -6 & -5 \end{bmatrix}$
   $\vec{x} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$

2. $A = \begin{bmatrix} 19 & -6 \\ 48 & -15 \end{bmatrix}$
   $\vec{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$

3. $A = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}$
   $\vec{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

4. $A = \begin{bmatrix} -11 & -19 & 14 \\ -6 & -8 & 6 \\ -12 & -22 & 15 \end{bmatrix}$
   $\vec{x} = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}$

5. $A = \begin{bmatrix} -7 & 1 & 3 \\ 10 & 2 & -3 \\ -20 & -14 & 1 \end{bmatrix}$
   $\vec{x} = \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}$

6. $A = \begin{bmatrix} -12 & -10 & 0 \\ 15 & 13 & 0 \\ 15 & 18 & -5 \end{bmatrix}$
   $\vec{x} = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$

In Exercises 7 – 11, a matrix A and one of its eigenvalues are given. Find an eigenvector of A for the given eigenvalue.

7. $A = \begin{bmatrix} 16 & 6 \\ -18 & -5 \end{bmatrix}$
   $\lambda = 4$

8. $A = \begin{bmatrix} -2 & 6 \\ -9 & 13 \end{bmatrix}$
   $\lambda = 7$

9. $A = \begin{bmatrix} -16 & -28 & -19 \\ 42 & 69 & 46 \\ -42 & -72 & -49 \end{bmatrix}$
   $\lambda = 5$

10. $A = \begin{bmatrix} 7 & -5 & -10 \\ 6 & 2 & -6 \\ 2 & -5 & -5 \end{bmatrix}$
    $\lambda = -3$

11. $A = \begin{bmatrix} 4 & 5 & -3 \\ -7 & -8 & 3 \\ 1 & -5 & 8 \end{bmatrix}$
    $\lambda = 2$

In Exercises 12 – 28, find the eigenvalues of the given matrix. For each eigenvalue, give an eigenvector.

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

13. $\begin{bmatrix} -4 & 72 \\ -1 & 13 \end{bmatrix}$

14. $\begin{bmatrix} 2 & -12 \\ 2 & -8 \end{bmatrix}$

15. $\begin{bmatrix} 3 & 12 \\ 1 & -1 \end{bmatrix}$

16. $\begin{bmatrix} 5 & 9 \\ -1 & -5 \end{bmatrix}$

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

18. $\begin{bmatrix} 0 & 1 \\ 25 & 0 \end{bmatrix}$

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

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

21. $\begin{bmatrix} 5 & -2 & 3 \\ 0 & 4 & 0 \\ 0 & -1 & 3 \end{bmatrix}$

22. $\begin{bmatrix} 1 & 0 & 12 \\ 2 & -5 & 0 \\ 1 & 0 & 2 \end{bmatrix}$

23. $\begin{bmatrix} 1 & 0 & -18 \\ -4 & 3 & -1 \\ 1 & 0 & -8 \end{bmatrix}$

24. $\begin{bmatrix} -1 & 18 & 0 \\ 1 & 2 & 0 \\ 5 & -3 & -1 \end{bmatrix}$

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

26. $\begin{bmatrix} 2 & -1 & 1 \\ 0 & 3 & 6 \\ 0 & 0 & 7 \end{bmatrix}$

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

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

8.2 Properties of Eigenvalues and Eigenvectors

AS YOU READ . . .

1. T/F: A and A^T have the same eigenvectors.
2. T/F: A and A^{-1} have the same eigenvalues.
3. T/F: Marie Ennemond Camille Jordan was a guy.
4. T/F: Matrices with a trace of 0 are important, although we haven't seen why.
5. T/F: A matrix A is invertible only if 1 is an eigenvalue of A.

In this section we'll explore how the eigenvalues and eigenvectors of a matrix relate to other properties of that matrix. This section is essentially a hodgepodge of interesting facts about eigenvalues; the goal here is not to memorize various facts about matrix algebra, but to again be amazed at the many connections between mathematical concepts.

We'll begin our investigations with an example that will give a foundation for other discoveries.

Example 157 **Eigenvalues of a triangular matrix**

Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$. Find the eigenvalues of A.

Solution To find the eigenvalues, we compute $\det(A - \lambda I)$:

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 1-\lambda & 2 & 3 \\ 0 & 4-\lambda & 5 \\ 0 & 0 & 6-\lambda \end{vmatrix} \\ &= (1-\lambda)(4-\lambda)(6-\lambda) \end{aligned}$$

Since our matrix is triangular, the determinant is easy to compute; it is just the product of the diagonal elements. Therefore, we found (and factored) our characteristic polynomial very easily, and we see that we have eigenvalues of $\lambda = 1, 4$, and 6.

This examples demonstrates a wonderful fact for us: the eigenvalues of a triangular matrix are simply the entries on the diagonal. Finding the corresponding eigenvectors still takes some work, but finding the eigenvalues is easy.

With that fact in the backs of our minds, let us proceed to the next example where we will come across some more interesting facts about eigenvalues and eigenvectors.

Example 158 **Exploring properties of eigenvalues**

Let $A = \begin{bmatrix} -3 & 15 \\ 3 & 9 \end{bmatrix}$ and let $B = \begin{bmatrix} -7 & -2 & 10 \\ -3 & 2 & 3 \\ -6 & -2 & 9 \end{bmatrix}$ (as used in Examples 152 and 154, respectively). Find the following:

1. eigenvalues and eigenvectors of A and B
2. eigenvalues and eigenvectors of A^{-1} and B^{-1}
3. eigenvalues and eigenvectors of A^T and B^T
4. The trace of A and B
5. The determinant of A and B

Solution We'll answer each in turn.

1. We already know the answer to these for we did this work in previous examples. Therefore we just list the answers.

   For A, we have eigenvalues $\lambda = -6$ and 12, with eigenvectors

   $$\vec{x} = x_2 \begin{bmatrix} -5 \\ 1 \end{bmatrix} \text{ and } x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \text{ respectively.}$$

   For B, we have eigenvalues $\lambda = -1,\ 2,$ and 3 with eigenvectors

   $$\vec{x} = x_3 \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}, \ x_3 \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \text{ and } x_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \text{ respectively.}$$

2. We first compute the inverses of A and B. They are:

   $$A^{-1} = \begin{bmatrix} -1/8 & 5/24 \\ 1/24 & 1/24 \end{bmatrix} \quad \text{and} \quad B^{-1} = \begin{bmatrix} -4 & 1/3 & 13/3 \\ -3/2 & 1/2 & 3/2 \\ -3 & 1/3 & 10/3 \end{bmatrix}.$$

   Finding the eigenvalues and eigenvectors of these matrices is not terribly hard, but it is not "easy," either. Therefore, we omit showing the intermediate steps and go right to the conclusions.

   For A^{-1}, we have eigenvalues $\lambda = -1/6$ and $1/12$, with eigenvectors

   $$\vec{x} = x_2 \begin{bmatrix} -5 \\ 1 \end{bmatrix} \text{ and } x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \text{ respectively.}$$

   For B^{-1}, we have eigenvalues $\lambda = -1$, $1/2$ and $1/3$ with eigenvectors

   $$\vec{x} = x_3 \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}, \ x_3 \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} \text{ and } x_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \text{ respectively.}$$

3. Of course, computing the transpose of A and B is easy; computing their eigenvalues and eigenvectors takes more work. Again, we omit the intermediate steps.

For A^T, we have eigenvalues $\lambda = -6$ and 12 with eigenvectors

$$\vec{x} = x_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} \text{ and } x_2 \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \text{ respectively.}$$

For B^T, we have eigenvalues $\lambda = -1,\ 2$ and 3 with eigenvectors

$$\vec{x} = x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix},\ x_3 \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \text{ and } x_3 \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}, \text{ respectively.}$$

4. The trace of A is 6; the trace of B is 4.

5. The determinant of A is -72; the determinant of B is -6.

Now that we have completed the "grunt work," let's analyze the results of the previous example. We are looking for any patterns or relationships that we can find.

The eigenvalues and eigenvectors of A and A^{-1}.

In our example, we found that the eigenvalues of A are -6 and 12; the eigenvalues of A^{-1} are $-1/6$ and $1/12$. Also, the eigenvalues of B are -1, 2 and 3, whereas the eigenvalues of B^{-1} are -1, $1/2$ and $1/3$. There is an obvious relationship here; it seems that if λ is an eigenvalue of A, then $1/\lambda$ will be an eigenvalue of A^{-1}. We can also note that the corresponding eigenvectors matched, too.

Why is this the case? Consider an invertible matrix A with eigenvalue λ and eigenvector $\vec{x}$. Then, by definition, we know that $A\vec{x} = \lambda\vec{x}$. Now multiply both sides by A^{-1}:

$$\begin{aligned} A\vec{x} &= \lambda\vec{x} \\ A^{-1}A\vec{x} &= A^{-1}\lambda\vec{x} \\ \vec{x} &= \lambda A^{-1}\vec{x} \\ \frac{1}{\lambda}\vec{x} &= A^{-1}\vec{x} \end{aligned}$$

We have just shown that $A^{-1}\vec{x} = 1/\lambda\vec{x}$; this, by definition, shows that $\vec{x}$ is an eigenvector of A^{-1} with eigenvalue $1/\lambda$. This explains the result we saw above.

The eigenvalues and eigenvectors of A and A^T.

Our example showed that A and A^T had the same eigenvalues but different (but somehow similar) eigenvectors; it also showed that B and B^T had the same eigenvalues but unrelated eigenvectors. Why is this?

We can answer the eigenvalue question relatively easily; it follows from the properties of the determinant and the transpose. Recall the following two facts:

1. $(A + B)^T = A^T + B^T$ (Theorem 20) and

2. $\det(A) = \det(A^T)$ (Theorem 25).

We find the eigenvalues of a matrix by computing the characteristic polynomial; that is, we find $\det\left(A-\lambda I\right)$. What is the characteristic polynomial of A^T? Consider:

$$\begin{aligned}
\det\left(A^T-\lambda I\right) &= \det\left(A^T-\lambda I^T\right) && \text{since } I = I^T \\
&= \det\left((A-\lambda I)^T\right) && \text{Theorem 20} \\
&= \det\left(A-\lambda I\right) && \text{Theorem 25}
\end{aligned}$$

So we see that the characteristic polynomial of A^T is the same as that for A. Therefore they have the same eigenvalues.

What about their respective eigenvectors? Is there any relationship? The simple answer is "No."

The eigenvalues and eigenvectors of *A* and The Trace.

Note that the eigenvalues of A are -6 and 12, and the trace is 6; the eigenvalues of B are -1, 2 and 3, and the trace of B is 4. Do we notice any relationship?

It seems that the sum of the eigenvalues is the trace! Why is this the case?

The answer to this is a bit out of the scope of this text; we can justify part of this fact, and another part we'll just state as being true without justification.

Who in the world thinks up this stuff? It seems that the answer is Marie Ennemond Camille Jordan, who, despite having at least two girl names, was a guy.

First, recall from Theorem 22 that $\text{tr}(AB)=\text{tr}(BA)$. Secondly, we state without justification that given a square matrix A, we can find a square matrix P such that $P^{-1}AP$ is an upper triangular matrix with the eigenvalues of A on the diagonal.

Thus $\text{tr}(P^{-1}AP)$ is the sum of the eigenvalues; also, using our Theorem 22, we know that $\text{tr}(P^{-1}AP)=\text{tr}(P^{-1}PA)=\text{tr}(A)$. Thus the trace of A is the sum of the eigenvalues.

The eigenvalues and eigenvectors of *A* and The Determinant.

Again, the eigenvalues of A are -6 and 12, and the determinant of A is -72. The eigenvalues of B are -1, 2 and 3; the determinant of B is -6. It seems as though the product of the eigenvalues is the determinant.

This is indeed true; we defend this with our argument from above. We know that the determinant of a triangular matrix is the product of the diagonal elements. Therefore, given a matrix A, we can find P such that $P^{-1}AP$ is upper triangular with the eigenvalues of A on the diagonal. Thus $\det\left(P^{-1}AP\right)$ is the product of the eigenvalues. Using Theorem 25, we know that $\det\left(P^{-1}AP\right)=\det\left(P^{-1}PA\right)=\det\left(A\right)$. Thus the determinant of A is the product of the eigenvalues.

We summarize the results of our example with the following theorem.

Theorem 38 Properties of Eigenvalues and Eigenvectors

Let A be an $n \times n$ invertible matrix. The following are true:

1. If A is triangular, then the diagonal elements of A are the eigenvalues of A.
2. If λ is an eigenvalue of A with eigenvector $\vec{x}$, then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1} with eigenvector $\vec{x}$.
3. If λ is an eigenvalue of A then λ is an eigenvalue of A^T.
4. The sum of the eigenvalues of A is equal to $\text{tr}(A)$, the trace of A.
5. The product of the eigenvalues of A is the equal to $\det(A)$, the determinant of A.

There is one more concept concerning eigenvalues and eigenvectors that we will explore. We do so in the context of an example.

Example 159 Eigenvalues of a non-invertible matrix

Find the eigenvalues and eigenvectors of the matrix $A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$.

Since 0 is a "special" number, we might think so – afterall, we found that having a determinant of 0 is important. Then again, a matrix with a trace of 0 isn't all that important. (Well, as far as we have seen; it actually *is*). So, having an eigenvalue of 0 may or may not be significant, but we would be doing well if we recognized the possibility of significance and decided to investigate further.

Solution To find the eigenvalues, we compute $\det(A - \lambda I)$:

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 1-\lambda & 2 \\ 1 & 2-\lambda \end{vmatrix} \\ &= (1-\lambda)(2-\lambda) - 2 \\ &= \lambda^2 - 3\lambda \\ &= \lambda(\lambda - 3) \end{aligned}$$

Our eigenvalues are therefore $\lambda = 0, 3$.
For $\lambda = 0$, we find the eigenvectors:

$$\begin{bmatrix} 1 & 2 & 0 \\ 1 & 2 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This shows that $x_1 = -2x_2$, and so our eigenvectors $\vec{x}$ are

$$\vec{x} = x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

For $\lambda = 3$, we find the eigenvectors:

$$\begin{bmatrix} -2 & 2 & 0 \\ 1 & -1 & 0 \end{bmatrix} \quad \overrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This shows that $x_1 = x_2$, and so our eigenvectors $\vec{x}$ are

$$\vec{x} = x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

One interesting thing about the above example is that we see that 0 is an eigenvalue of A; we have not officially encountered this before. Does this mean anything significant?

Think about what an eigenvalue of 0 means: there exists an nonzero vector $\vec{x}$ where $A\vec{x} = 0\vec{x} = \vec{0}$. That is, we have a nontrivial solution to $A\vec{x} = \vec{0}$. We know this only happens when A is not invertible.

So if A is invertible, there is no nontrivial solution to $A\vec{x} = \vec{0}$, and hence 0 *is not* an eigenvalue of A. If A is not invertible, then there is a nontrivial solution to $A\vec{x} = \vec{0}$, and hence 0 *is* an eigenvalue of A. This leads us to our final addition to the Invertible Matrix Theorem.

Theorem 39 Invertible Matrix Theorem

Let A be an $n \times n$ matrix. The following statements are equivalent.

(a) A is invertible.

(h) A does not have an eigenvalue of 0.

This section is about the properties of eigenvalues and eigenvectors. Of course, we have not investigated all of the numerous properties of eigenvalues and eigenvectors; we have just surveyed some of the most common (and most important) concepts. Here are four quick examples of the many things that still exist to be explored.

First, recall the matrix

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

that we used in Example 150. It's characteristic polynomial is $p(\lambda) = \lambda^2 - 4\lambda - 5$. Compute $p(A)$; that is, compute $A^2 - 4A - 5I$. You should get something "interesting," and you should wonder "does this always work?" (Yes.)

Second, in all of our examples, we have considered matrices where eigenvalues "appeared only once." Since we know that the eigenvalues of a triangular matrix appear on the diagonal, we know that the eigenvalues of

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

are "1 and 1;" that is, the eigenvalue $\lambda = 1$ appears twice. What does that mean when we consider the eigenvectors of $\lambda = 1$? Compare the result of this to the matrix

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

which also has the eigenvalue $\lambda = 1$ appearing twice.

To direct further study, it helps to know that mathematicians refer to this as the *multiplicity* of an eigenvalue. In each of these two examples, A has the eigenvalue $\lambda = 1$ with multiplicity of 2.

Third, consider the matrix

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

What are the eigenvalues? (Be careful; this matrix is *not* triangular.) We quickly compute the characteristic polynomial to be $p(\lambda) = \lambda^2 + 1$. Therefore the eigenvalues are $\pm\sqrt{-1} = \pm i$. What does this mean?

Finally, we have found the eigenvalues of matrices by finding the roots of the characteristic polynomial. We have limited our examples to quadratic and cubic polynomials; one would expect for larger sized matrices that a computer would be used to factor the characteristic polynomials. However, in general, this is *not* how the eigenvalues are found. Factoring high order polynomials is too unreliable, even with a computer – round off errors can cause unpredictable results.

Also, to even compute the characteristic polynomial, one needs to compute the determinant, which is also expensive (as discussed in the previous chapter).

So how are eigenvalues found? There are *iterative* processes that can progressively transform a matrix A into another matrix that is *almost* an upper triangular matrix (the entries below the diagonal are almost zero) where the entries on the diagonal are the eigenvalues. The more iterations one performs, the better the approximation is.

These methods are so fast and reliable that some computer programs convert polynomial root finding problems into eigenvalue problems!

Most textbooks on Linear Algebra will provide direction on exploring the above topics and give further insight to what is going on. We have mentioned all the eigenvalue and eigenvector properties in this section for the same reasons we gave in the previous section. First, knowing these properties helps us solve numerous real world problems, and second, it is fascinating to see how rich and deep the theory of matrices is.

Exercises 8.2

Problems

In Exercises 1 – 6, a matrix A is given. For each,

(a) **Find the eigenvalues of A, and for each eigenvalue, find an eigenvector.**

(b) **Do the same for A^T.**

(c) **Do the same for A^{-1}.**

(d) **Find tr(A).**

(e) **Find det (A).**

Use Theorem 38 to verify your results.

1. $\begin{bmatrix} 0 & 4 \\ -1 & 5 \end{bmatrix}$

2. $\begin{bmatrix} -2 & -14 \\ -1 & 3 \end{bmatrix}$

3. $\begin{bmatrix} 5 & 30 \\ -1 & -6 \end{bmatrix}$

4. $\begin{bmatrix} -4 & 72 \\ -1 & 13 \end{bmatrix}$

5. $\begin{bmatrix} 5 & -9 & 0 \\ 1 & -5 & 0 \\ 2 & 4 & 3 \end{bmatrix}$

6. $\begin{bmatrix} 0 & 25 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & -3 \end{bmatrix}$

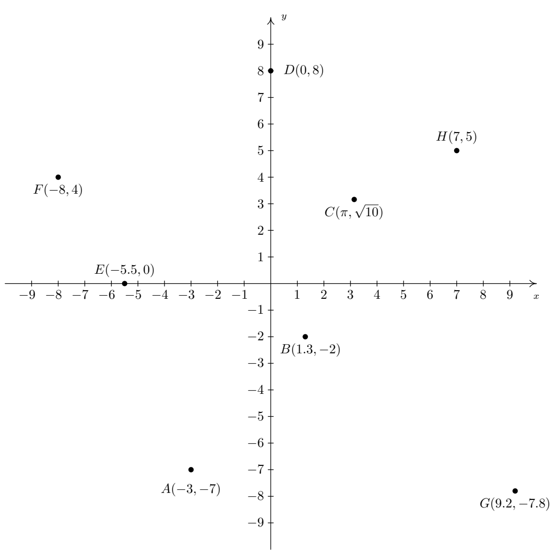
A: Answers To Selected Problems

Section 1.2

1. 6
3. $\frac{2}{21}$
5. $-\frac{1}{3}$
7. $\frac{3}{5}$
9. $-\frac{7}{8}$
11. 0
13. $\frac{23}{9}$
15. $-\frac{24}{7}$
17. $\frac{243}{32}$
19. $\frac{9}{22}$

Section 1.3

1. The required points $A(-3,-7)$, $B(1.3,-2)$, $C(\pi,\sqrt{10})$, $D(0,8)$, $E(-5.5,0)$, $F(-8,4)$, $G(9.2,-7.8)$, and $H(7,5)$ are plotted in the Cartesian Coordinate Plane below.



3. $d=4\sqrt{10}$, $M=(1,-4)$
5. $d=\frac{\sqrt{37}}{2}$, $M=\left(\frac{5}{6},\frac{7}{4}\right)$
7. $d=3\sqrt{5}$, $M=\left(-\frac{\sqrt{2}}{2},-\frac{\sqrt{3}}{2}\right)$
9. $d=\sqrt{x^2+y^2}$, $M=\left(\frac{x}{2},\frac{y}{2}\right)$
11. $(0,3)$
13. $\left(\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2}\right)$, $\left(-\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right)$
15.
17. (a) The distance from A to B is $|AB|=\sqrt{13}$, the distance from A to C is $|AC|=\sqrt{52}$, and the distance from B to C is $|BC|=\sqrt{65}$. Since $\left(\sqrt{13}\right)^2+\left(\sqrt{52}\right)^2=\left(\sqrt{65}\right)^2$, we are guaranteed by the converse of the Pythagorean Theorem that the triangle is a right triangle.

    (b) Show that $|AC|^2+|BC|^2=|AB|^2$
19.

Chapter 2

Section 2.1

1. right hand
3. curve (a parabola); surface (a cylinder)
5. a hyperboloid of two sheets

7. $\| \overline{AB} \| = \sqrt{6}$; $\| \overline{BC} \| = \sqrt{17}$; $\| \overline{AC} \| = \sqrt{11}$. Yes, it is a right triangle as $\| \overline{AB} \|^2 + \| \overline{AC} \|^2 = \| \overline{BC} \|^2$.

9. Center at $(4, -1, 0)$; radius = 3

11. Interior of a sphere with radius 1 centered at the origin.

13. The first octant of space; all points (x, y, z) where each of x, y and z are positive. (Analogous to the first quadrant in the plane.)

Section 2.2

1. Answers will vary.

3. A vector with magnitude 1.

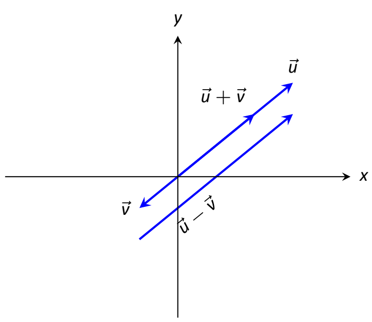
5. It stretches the vector by a factor of 2, and points it in the opposite direction.

7. $\vec{v}PQ = \langle -4, 4\rangle = -4\vec{i} + 4\vec{j}$

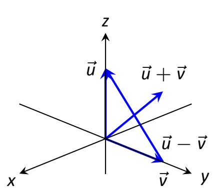
9. $\vec{v}PQ = \langle 2, 2, 0\rangle = 2\vec{i} + 2\vec{j}$

11. (a) $\vec{u}+\vec{v} = \langle 3, 2, 1\rangle$; $\vec{u}-\vec{v} = \langle -1, 0, -3\rangle$; $\pi\vec{u} - \sqrt{2}\vec{v} = \langle \pi - 2\sqrt{2}, \pi - \sqrt{2}, -\pi - 2\sqrt{2}\rangle$.

    (c) $\vec{x} = \langle -1, 0, -3\rangle$.

13. 

Sketch of $\vec{u}-\vec{v}$ shifted for clarity.

15. 

17. $\| \vec{u} \| = \sqrt{17}$, $\| \vec{v} \| = \sqrt{3}$, $\| \vec{u}+\vec{v} \| = \sqrt{14}$, $\| \vec{u}-\vec{v} \| = \sqrt{26}$

19. $\| \vec{u} \| = 7$, $\| \vec{v} \| = 35$, $\| \vec{u}+\vec{v} \| = 42$, $\| \vec{u}-\vec{v} \| = 28$

21. $\vec{u} = \langle 3/\sqrt{30}, 7/\sqrt{30}\rangle$

23. $\vec{u} = \langle 1/3, -2/3, 2/3\rangle$

25. $\vec{u} = \langle \cos 50^\circ, \sin 50^\circ\rangle \approx \langle 0.643, 0.766\rangle$.

27.
$$\begin{aligned}\| \vec{u} \| &= \sqrt{\sin^2\theta\cos^2\varphi + \sin^2\theta\sin^2\varphi + \cos^2\theta}\\ &= \sqrt{\sin^2\theta(\cos^2\varphi + \sin^2\varphi) + \cos^2\theta}\\ &= \sqrt{\sin^2\theta + \cos^2\theta}\\ &= 1.\end{aligned}$$

29. The force on each chain is 100lb.

31. The force on each chain is 50lb.

33. $\theta = 5.71^\circ$; the weight is lifted 0.005 ft (about 1/16th of an inch).

35. $\theta = 84.29^\circ$; the weight is lifted 9 ft.

Section 2.3

1. Scalar

3. By considering the sign of the dot product of the two vectors. If the dot product is positive, the angle is acute; if the dot product is negative, the angle is obtuse.

5. -22

7. 3

9. not defined

11. Answers will vary.

13. $\theta = 0.3218 \approx 18.43^\circ$

15. $\theta = \pi/4 = 45^\circ$

17. Answers will vary; two possible answers are $\langle -7, 4\rangle$ and $\langle 14, -8\rangle$.

19. Answers will vary; two possible answers are $\langle 1, 0, -1\rangle$ and $\langle 4, 5, -9\rangle$.

21. $\text{proj}_{\vec{v}}\,\vec{u} = \langle -1/2, 3/2\rangle$.

23. $\text{proj}_{\vec{v}}\,\vec{u} = \langle -1/2, -1/2\rangle$.

25. $\text{proj}_{\vec{v}}\,\vec{u} = \langle 1, 2, 3\rangle$.

27. $\vec{u} = \langle -1/2, 3/2\rangle + \langle 3/2, 1/2\rangle$.

29. $\vec{u} = \langle -1/2, -1/2\rangle + \langle -5/2, 5/2\rangle$.

31. $\vec{u} = \langle 1, 2, 3\rangle + \langle 0, 3, -2\rangle$.

33. 1.96lb

35. 141.42ft–lb

37. 500ft–lb

39. 500ft–lb

Section 2.4

1. vector

3. "Perpendicular" is one answer.

5. Torque

7. $\vec{u}\times\vec{v} = \langle 11, 1, -17\rangle$

9. $\vec{u}\times\vec{v} = \langle 47, -36, -44\rangle$

11. $\vec{u}\times\vec{v} = \langle 0, 0, 0\rangle$

13. $\vec{i}\times\vec{k} = -\vec{j}$

15. Answers will vary.

17. 5

19. 0

21. $\sqrt{14}$

23. 3

25. $5\sqrt{2}/2$

27. 1

29. 7

31. 2

33. $\pm\frac{1}{\sqrt{6}}\langle 1, 1, -2\rangle$

35. $\langle 0, \pm 1, 0\rangle$

37. 87.5ft–lb

39. $200/3 \approx 66.67$ft–lb

41. With $\vec{u} = \langle u_1, u_2, u_3\rangle$ and $\vec{v} = \langle v_1, v_2, v_3\rangle$, we have
$$\begin{aligned}\vec{u}\cdot(\vec{u}\times\vec{v}) &= \langle u_1, u_2, u_3\rangle \cdot (\langle u_2v_3 - u_3v_2, -(u_1v_3 - u_3v_1), u_1v_2 - u_2v_1\rangle)\\ &= u_1(u_2v_3 - u_3v_2) - u2(u_1v_3 - u_3v_1) + u3(u_1v_2 - u_2v_1)\\ &= 0.\end{aligned}$$

Section 2.5

1. A point on the line and the direction of the line.
3. parallel, skew
5. vector: $\ell(t) = \langle 2, -4, 1\rangle + t\,\langle 9, 2, 5\rangle$
parametric: $x = 2 + 9t$, $y = -4 + 2t$, $z = 1 + 5t$
symmetric: $(x-2)/9 = (y+4)/2 = (z-1)/5$
7. Answers can vary: vector: $\ell(t) = \langle 2, 1, 5\rangle + t\,\langle 5, -3, -1\rangle$
parametric: $x = 2 + 5t$, $y = 1 - 3t$, $z = 5 - t$
symmetric: $(x-2)/5 = -(y-1)/3 = -(z-5)$
9. Answers can vary; here the direction is given by $\vec{d}_1 \times \vec{d}_2$: vector: $\ell(t) = \langle 0, 1, 2\rangle + t\,\langle -10, 43, 9\rangle$
parametric: $x = -10t$, $y = 1 + 43t$, $z = 2 + 9t$
symmetric: $-x/10 = (y-1)/43 = (z-2)/9$
11. Answers can vary; here the direction is given by $\vec{d}_1 \times \vec{d}_2$: vector: $\ell(t) = \langle 7, 2, -1\rangle + t\,\langle 1, -1, 2\rangle$
parametric: $x = 7 + t$, $y = 2 - t$, $z = -1 + 2t$
symmetric: $x - 7 = 2 - y = (z+1)/2$
13. vector: $\ell(t) = \langle 1, 1\rangle + t\,\langle 2, 3\rangle$
parametric: $x = 1 + 2t$, $y = 1 + 3t$
symmetric: $(x-1)/2 = (y-1)/3$
15. parallel
17. intersecting; $\vec{\ell}_1(3) = \vec{\ell}_2(4) = \langle 9, -5, 13\rangle$
19. skew
21. same
23. $\sqrt{41}/3$
25. $5\sqrt{2}/2$
27. $3/\sqrt{2}$
29. Since both P and Q are on the line, $\vec{v}PQ$ is parallel to $\vec{d}$. Thus $\vec{v}PQ \times \vec{d} = \vec{0}$, giving a distance of 0.
31.
   (a) The distance formula cannot be used because since $\vec{d}_1$ and $\vec{d}_2$ are parallel, $\vec{c}$ is $\vec{0}$ and we cannot divide by $\parallel \vec{0} \parallel$.
   (b) Since $\vec{d}_1$ and $\vec{d}_2$ are parallel, $\vec{v}P_1P_2$ lies in the plane formed by the two lines. Thus $\vec{v}P_1P_2 \times \vec{d}_2$ is orthogonal to this plane, and $\vec{c} = (\vec{v}P_1P_2 \times \vec{d}_2) \times \vec{d}_2$ is parallel to the plane, but still orthogonal to both $\vec{d}_1$ and $\vec{d}_2$. We desire the length of the projection of $\vec{v}P_1P_2$ onto $\vec{c}$, which is what the formula provides.
   (c) Since the lines are parallel, one can measure the distance between the lines at any location on either line (just as to find the distance between straight railroad tracks, one can use a measuring tape anywhere along the track, not just at one specific place.) Let $P = P_1$ and $Q = P_2$ as given by the equations of the lines, and apply the formula for distance between a point and a line.

Section 2.6

1. A point in the plane and a normal vector (i.e., a direction orthogonal to the plane).
3. Answers will vary.
5. Answers will vary.
7. Standard form: $3(x-2) - (y-3) + 7(z-4) = 0$
general form: $3x - y + 7z = 31$
9. Answers may vary;
Standard form: $8(x-1) + 4(y-2) - 4(z-3) = 0$
general form: $8x + 4y - 4z = 4$
11. Answers may vary;
Standard form: $-7(x-2) + 2(y-1) + (z-2) = 0$
general form: $-7x + 2y + z = -10$
13. Answers may vary;
Standard form: $2(x-1) - (y-1) = 0$
general form: $2x - y = 1$
15. Answers may vary;
Standard form: $2(x-2) - (y+6) - 4(z-1) = 0$
general form: $2x - y - 4z = 6$
17. Answers may vary;
Standard form: $(x-5) + (y-7) + (z-3) = 0$
general form: $x + y + z = 15$
19. Answers may vary;
Standard form: $3(x+4) + 8(y-7) - 10(z-2) = 0$
general form: $3x + 8y - 10z = 24$
21. Answers may vary:
$\ell = \begin{cases} x = 14t \\ y = -1 - 10t \\ z = 2 - 8t \end{cases}$
23. $(-3, -7, -5)$
25. No point of intersection; the plane and line are parallel.
27. $\sqrt{5/7}$
29. $1/\sqrt{3}$
31. If P is any point in the plane, and Q is also in the plane, then $\vec{v}PQ$ lies parallel to the plane and is orthogonal to $\vec{n}$, the normal vector. Thus $\vec{n} \cdot \vec{v}PQ = 0$, giving the distance as 0.

Chapter 3

Section 3.1

1. $\begin{bmatrix} -2 & -1 \\ 12 & 13 \end{bmatrix}$
3. $\begin{bmatrix} 2 & -2 \\ 14 & 8 \end{bmatrix}$
5. $\begin{bmatrix} 9 & -7 \\ 11 & -6 \end{bmatrix}$
7. $\begin{bmatrix} -14 \\ 6 \end{bmatrix}$
9. $\begin{bmatrix} -15 \\ -25 \end{bmatrix}$
11. $X = \begin{bmatrix} -5 & 9 \\ -1 & -14 \end{bmatrix}$
13. $X = \begin{bmatrix} -5 & -2 \\ -9/2 & -19/2 \end{bmatrix}$
15. $a = 2$, $b = 1$
17. $a = 5/2 + 3/2b$
19. No solution.
21. No solution.

Section 3.2

1. -22
3. 0
5. 5
7. 15
9. -2
11. Not possible.
13. $AB = \begin{bmatrix} 8 & 3 \\ 10 & -9 \end{bmatrix}$
$BA = \begin{bmatrix} -3 & 24 \\ 4 & 2 \end{bmatrix}$

15. $AB = \begin{bmatrix} -1 & -2 & 12 \\ 10 & 4 & 32 \end{bmatrix}$

BA is not possible.

17. AB is not possible.

$BA = \begin{bmatrix} 27 & -33 & 39 \\ -27 & -3 & -15 \end{bmatrix}$

19. $AB = \begin{bmatrix} -32 & 34 & -24 \\ -32 & 38 & -8 \\ -16 & 21 & 4 \end{bmatrix}$

$BA = \begin{bmatrix} 22 & -14 \\ -4 & -12 \end{bmatrix}$

21. $AB = \begin{bmatrix} -56 & 2 & -36 \\ 20 & 19 & -30 \\ -50 & -13 & 0 \end{bmatrix}$

$BA = \begin{bmatrix} -46 & 40 \\ 72 & 9 \end{bmatrix}$

23. $AB = \begin{bmatrix} -15 & -22 & -21 & -1 \\ 16 & -53 & -59 & -31 \end{bmatrix}$

BA is not possible.

25. $AB = \begin{bmatrix} 0 & 0 & 4 \\ 6 & 4 & -2 \\ 2 & -4 & -6 \end{bmatrix}$

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

27. $AB = \begin{bmatrix} 21 & -17 & -5 \\ 19 & 5 & 19 \\ 5 & 9 & 4 \end{bmatrix}$

$BA = \begin{bmatrix} 19 & 5 & 23 \\ 5 & -7 & -1 \\ -14 & 6 & 18 \end{bmatrix}$

29. $DA = \begin{bmatrix} 4 & -6 \\ 4 & -6 \end{bmatrix}$

$AD = \begin{bmatrix} 4 & 8 \\ -3 & -6 \end{bmatrix}$

31. $DA = \begin{bmatrix} 2 & 2 & 2 \\ -6 & -6 & -6 \\ -15 & -15 & -15 \end{bmatrix}$ $AD = \begin{bmatrix} 2 & -3 & 5 \\ 4 & -6 & 10 \\ -6 & 9 & -15 \end{bmatrix}$

33. $DA = \begin{bmatrix} d_1a & d_1b & d_1c \\ d_2d & d_2e & d_2f \\ d_3g & d_3h & d_3i \end{bmatrix}$ $AD = \begin{bmatrix} d_1a & d_2b & d_3c \\ d_1d & d_2e & d_3f \\ d_1g & d_2h & d_3i \end{bmatrix}$

35. $A\vec{x} = \begin{bmatrix} -6 \\ 11 \end{bmatrix}$

37. $A\vec{x} = \begin{bmatrix} -5 \\ 5 \\ 21 \end{bmatrix}$

39. $A\vec{x} = \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ x_1 + 2x_3 \\ 2x_1 + 3x_2 + x_3 \end{bmatrix}$

41. $A^2 = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}; A^3 = \begin{bmatrix} 8 & 0 \\ 0 & 27 \end{bmatrix}$

43. $A^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}; A^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

45. (a) $\begin{bmatrix} 0 & -2 \\ -5 & -1 \end{bmatrix}$

(b) $\begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix}$

(c) $\begin{bmatrix} -11 & -15 \\ 37 & 32 \end{bmatrix}$

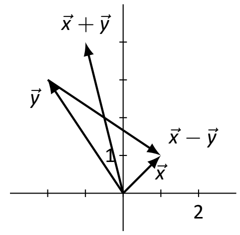
(d) No.

(e) $(A+B)(A+B) = AA + AB + BA + BB = A^2 + AB + BA + B^2$.

Section 3.3

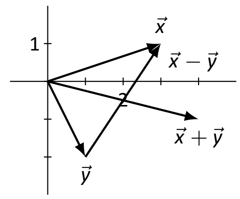
1. $\vec{x} + \vec{y} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}, \vec{x} - \vec{y} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$

Sketches will vary depending on choice of origin of each vector.

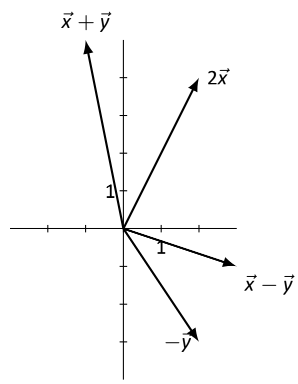


3. $\vec{x} + \vec{y} = \begin{bmatrix} -3 \\ 3 \end{bmatrix}, \vec{x} - \vec{y} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

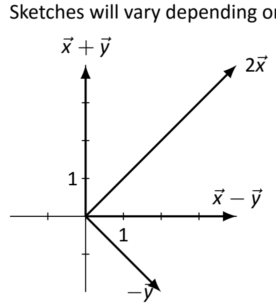
Sketches will vary depending on choice of origin of each vector.



5. Sketches will vary depending on choice of origin of each vector.



7. Sketches will vary depending on choice of origin of each vector.



9. $||\vec{x}|| = \sqrt{5}$; $||a\vec{x}|| = \sqrt{45} = 3\sqrt{5}$. The vector $a\vec{x}$ is 3 times as long as $\vec{x}$.

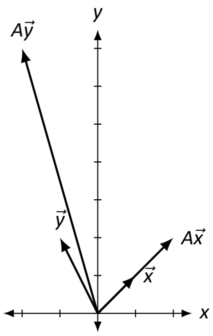
11. $||\vec{x}|| = \sqrt{34}$; $||a\vec{x}|| = \sqrt{34}$. The vectors $a\vec{x}$ and $\vec{x}$ are the same length (they just point in opposite directions).

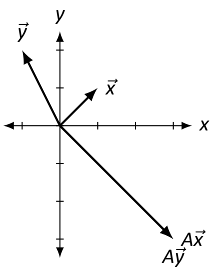
13. (a) $||\vec{x}|| = \sqrt{2}$; $||\vec{y}|| = \sqrt{13}$; $||\vec{x}+\vec{y}|| = 5$.

(b) $||\vec{x}|| = \sqrt{5}$; $||\vec{y}|| = 3\sqrt{5}$; $||\vec{x}+\vec{y}|| = 4\sqrt{5}$.

(c) $||\vec{x}|| = \sqrt{10}$; $||\vec{y}|| = \sqrt{29}$; $||\vec{x}+\vec{y}|| = \sqrt{65}$.

(d) $||\vec{x}|| = \sqrt{5}$; $||\vec{y}|| = 2\sqrt{5}$; $||\vec{x}+\vec{y}|| = \sqrt{5}$.
The equality holds sometimes; only when $\vec{x}$ and $\vec{y}$ point along the same line, in the same direction.

15. 

17. 

Section 3.4

1. $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$
3. $A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$
5. $A = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}$
7. $A = \begin{bmatrix} 0 & 1 \\ 3 & 0 \end{bmatrix}$
9. $A = \begin{bmatrix} 0 & -1 \\ -1 & -1 \end{bmatrix}$
11. Yes, these are the same; the transformation matrix in each is $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$.
13. Yes, these are the same. Each produces the transformation matrix $\begin{bmatrix} 1/2 & 0 \\ 0 & 3 \end{bmatrix}$.

Section 3.5

1. Yes
3. No; cannot add a constant.
5. Yes.
7. $[\,T\,] = \begin{bmatrix} 1 & 2 \\ 3 & -5 \\ 0 & 2 \end{bmatrix}$
9. $[\,T\,] = \begin{bmatrix} 1 & 0 & 3 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \end{bmatrix}$
11. $[\,T\,] = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$

Chapter 4

Section 4.1

1. y
3. y
5. n
7. y
9. y
11. $x = 1, y = -2$
13. $x = -1, y = 0, z = 2$
15. 29 chickens and 33 pigs

Section 4.2

1. $\begin{bmatrix} 3 & 4 & 5 & 7 \\ -1 & 1 & -3 & 1 \\ 2 & -2 & 3 & 5 \end{bmatrix}$
3. $\begin{bmatrix} 1 & 3 & -4 & 5 & 17 \\ -1 & 0 & 4 & 8 & 1 \\ 2 & 3 & 4 & 5 & 6 \end{bmatrix}$
5. $\begin{aligned} x_1 + 2x_2 &= 3 \\ -x_1 + 3x_2 &= 9 \end{aligned}$
7. $\begin{aligned} x_1 + x_2 - x_3 - x_4 &= 2 \\ 2x_1 + x_2 + 3x_3 + 5x_4 &= 7 \end{aligned}$
9. $\begin{aligned} x_1 + x_3 + 7x_5 &= 2 \\ x_2 + 3x_3 + 2x_4 &= 5 \end{aligned}$
11. $\begin{bmatrix} 2 & -1 & 7 \\ 5 & 0 & 3 \\ 0 & 4 & -2 \end{bmatrix}$
13. $\begin{bmatrix} 2 & -1 & 7 \\ 0 & 4 & -2 \\ 5 & 8 & -1 \end{bmatrix}$
15. $\begin{bmatrix} 2 & -1 & 7 \\ 0 & 4 & -2 \\ 0 & 5/2 & -29/2 \end{bmatrix}$
17. $R_1 + R_2 \rightarrow R_2$
19. $R_1 \leftrightarrow R_2$
21. $x = 2, y = 1$
23. $x = -1, y = 0$
25. $x_1 = -2, x_2 = 1, x_3 = 2$

Section 4.3

1. (a) yes (b) no (c) no (d) yes
3. (a) no (b) yes (c) yes (d) yes
5. $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
7. $\begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}$
9. $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 7 \end{bmatrix}$
11. $\begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$
13. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

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

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

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

21. $\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & 1 & 4 \end{bmatrix}$

Section 4.4

1. $x_1 = 1 - 2x_2$; x_2 is free. Possible solutions: $x_1 = 1, x_2 = 0$ and $x_1 = -1, x_2 = 1$.
3. $x_1 = 1$; $x_2 = 2$
5. No solution; the system is inconsistent.
7. $x_1 = -11 + 10x_3$; $x_2 = -4 + 4x_3$; x_3 is free. Possible solutions: $x_1 = -11, x_2 = -4, x_3 = 0$ and $x_1 = -1, x_2 = 0$ and $x_3 = 1$.
9. $x_1 = 1 - x_2 - x_4$; x_2 is free; $x_3 = 1 - 2x_4$; x_4 is free. Possible solutions: $x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 0$ and $x_1 = -2, x_2 = 1, x_3 = -3, x_4 = 2$
11. No solution; the system is inconsistent.
13. $x_1 = \frac{1}{3} - \frac{4}{3}x_3$; $x_2 = \frac{1}{3} - \frac{1}{3}x_3$; x_3 is free. Possible solutions: $x_1 = \frac{1}{3}, x_2 = \frac{1}{3}, x_3 = 0$ and $x_1 = -1, x_2 = 0, x_3 = 1$
15. Never exactly 1 solution; infinite solutions if $k = 2$; no solution if $k \neq 2$.
17. Exactly 1 solution if $k \neq 2$; no solution if $k = 2$; never infinite solutions.

Section 4.5

1. 29 chickens and 33 pigs
3. 42 grande tables, 22 venti tables
5. 30 men, 15 women, 20 kids
7. $f(x) = -2x + 10$
9. $f(x) = \frac{1}{2}x^2 + 3x + 1$
11. $f(x) = 3$
13. $f(x) = x^3 + 1$
15. $f(x) = \frac{3}{2}x + 1$
17. The augmented matrix from this system is $\begin{bmatrix} 1 & 1 & 1 & 1 & 8 \\ 6 & 1 & 2 & 3 & 24 \\ 0 & 1 & -1 & 0 & 0 \end{bmatrix}$. From this we find the solution

$$\begin{aligned} t &= \frac{8}{3} - \frac{1}{3}f \\ x &= \frac{8}{3} - \frac{1}{3}f \\ w &= \frac{8}{3} - \frac{1}{3}f. \end{aligned}$$

The only time each of these variables are nonnegative integers is when $f = 2$ or $f = 8$. If $f = 2$, then we have 2 touchdowns, 2 extra points and 2 two point conversions (and 2 field goals); this doesn't make sense since the extra points and two point conversions follow touchdowns. If $f = 8$, then we have no touchdowns, extra points or two point conversions (just 8 field goals). This is the only solution; all points were scored from field goals.

19. Let x_1, x_2 and x_3 represent the number of free throws, 2 point and 3 point shots taken. The augmented matrix from this system is $\begin{bmatrix} 1 & 1 & 1 & 30 \\ 1 & 2 & 3 & 80 \end{bmatrix}$. From this we find the solution

$$\begin{aligned} x_1 &= -20 + x_3 \\ x_2 &= 50 - 2x_3. \end{aligned}$$

In order for x_1 and x_2 to be nonnegative, we need $20 \leq x_3 \leq 25$. Thus there are 6 different scenerios: the "first" is where 20 three point shots are taken, no free throws, and 10 two point shots; the "last" is where 25 three point shots are taken, 5 free throws, and no two point shots.

21. Let $y = ax + b$; all linear functions through (1,3) come in the form $y = (3 - b)x + b$. Examples: $b = 0$ yields $y = 3x$; $b = 2$ yields $y = x + 2$.
23. Let $y = ax^2 + bx + c$; we find that $a = -\frac{1}{2} + \frac{1}{2}c$ and $b = \frac{1}{2} - \frac{3}{2}c$. Examples: $c = 1$ yields $y = -x + 1$; $c = 3$ yields $y = x^2 - 4x + 3$.

Chapter 5

Section 5.1

1. Multiply $A\vec{u}$ and $A\vec{v}$ to verify.
3. Multiply $A\vec{u}$ and $A\vec{v}$ to verify.
5. Multiply $A\vec{u}$ and $A\vec{v}$ to verify.
7. Multiply $A\vec{u}$, $A\vec{v}$ and $A(\vec{u} + \vec{v})$ to verify.
9. Multiply $A\vec{u}$, $A\vec{v}$ and $A(\vec{u} + \vec{v})$ to verify.

11. (a) $\vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

    (b) $\vec{x} = \begin{bmatrix} 2/5 \\ -13/5 \end{bmatrix}$

13. (a) $\vec{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

    (b) $\vec{x} = \begin{bmatrix} -2 \\ -9/4 \end{bmatrix}$

15. (a) $\vec{x} = x_3 \begin{bmatrix} 5/4 \\ 1 \\ 1 \end{bmatrix}$

    (b) $\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 5/4 \\ 1 \\ 1 \end{bmatrix}$

17. (a) $\vec{x} = x_3 \begin{bmatrix} 14 \\ -10 \\ 0 \end{bmatrix}$

    (b) $\vec{x} = \begin{bmatrix} -4 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} 14 \\ -10 \\ 0 \end{bmatrix}$

19. (a) $\vec{x} = x_3 \begin{bmatrix} 2 \\ 2/5 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 2/5 \\ 0 \\ 1 \end{bmatrix}$

    (b) $\vec{x} = \begin{bmatrix} -2 \\ 2/5 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 2 \\ 2/5 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 2/5 \\ 0 \\ 1 \end{bmatrix}$

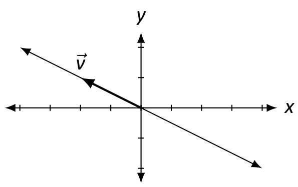
21. (a) $\vec{x} = x_2 \begin{bmatrix} -1/2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1/2 \\ 0 \\ -1/2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 13/2 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix}$

(b) $\vec{x} = \begin{bmatrix} -5 \\ 0 \\ 3/2 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -1/2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1/2 \\ 0 \\ -1/2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 13/2 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix}$

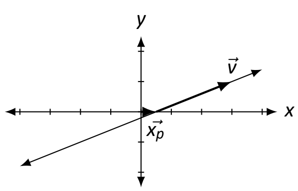
23. (a) $\vec{x} = x_4 \begin{bmatrix} 1 \\ 13/9 \\ -1/3 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{bmatrix}$

(b) $\vec{x} = \begin{bmatrix} 1 \\ 1/9 \\ 5/3 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 13/9 \\ -1/3 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{bmatrix}$

25. $\vec{x} = x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix} = x_2\vec{v}$



27. $\vec{x} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2.5 \\ 1 \end{bmatrix} = \vec{x_p} + x_2\vec{v}$



Section 5.2

1. $X = \begin{bmatrix} 1 & -9 \\ -4 & -5 \end{bmatrix}$

3. $X = \begin{bmatrix} -2 & -7 \\ 7 & -6 \end{bmatrix}$

5. $X = \begin{bmatrix} -5 & 2 & -3 \\ -4 & -3 & -2 \end{bmatrix}$

7. $X = \begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix}$

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

11. $X = \begin{bmatrix} 5/3 & 2/3 & 1 \\ -1/3 & 1/6 & 0 \\ 1/3 & 1/3 & 0 \end{bmatrix}$

Section 5.3

1. $\begin{bmatrix} -24 & -5 \\ 5 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 1/3 & 0 \\ 0 & 1/7 \end{bmatrix}$

5. A^{-1} does not exist.

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

9. $\begin{bmatrix} -5/13 & 3/13 \\ 1/13 & 2/13 \end{bmatrix}$

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

13. $\begin{bmatrix} 1 & 2 & -2 \\ 0 & 1 & -3 \\ 6 & 10 & -5 \end{bmatrix}$

15. $\begin{bmatrix} 1 & 0 & 0 \\ 52 & -48 & 7 \\ 8 & -7 & 1 \end{bmatrix}$

17. A^{-1} does not exist.

19. $\begin{bmatrix} 25 & 8 & 0 \\ 78 & 25 & 0 \\ -30 & -9 & 1 \end{bmatrix}$

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

23. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & -1 & 0 & -4 \\ -35 & -10 & 1 & -47 \\ -2 & -2 & 0 & -9 \end{bmatrix}$

25. $\begin{bmatrix} 28 & 18 & 3 & -19 \\ 5 & 1 & 0 & -5 \\ 4 & 5 & 1 & 0 \\ 52 & 60 & 12 & -15 \end{bmatrix}$

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

29. $\vec{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

31. $\vec{x} = \begin{bmatrix} -8 \\ 1 \end{bmatrix}$

33. $\vec{x} = \begin{bmatrix} -7 \\ 1 \\ -1 \end{bmatrix}$

35. $\vec{x} = \begin{bmatrix} 3 \\ -1 \\ -9 \end{bmatrix}$

Section 5.4

1. $(AB)^{-1} = \begin{bmatrix} -2 & 3 \\ 1 & -1.4 \end{bmatrix}$

3. $(AB)^{-1} = \begin{bmatrix} 29/5 & -18/5 \\ -11/5 & 7/5 \end{bmatrix}$

5. $A^{-1} = \begin{bmatrix} -2 & -5 \\ -1 & -3 \end{bmatrix}$,
$(A^{-1})^{-1} = \begin{bmatrix} -3 & 5 \\ 1 & -2 \end{bmatrix}$

7. $A^{-1} = \begin{bmatrix} -3 & 7 \\ 1 & -2 \end{bmatrix}$,
$(A^{-1})^{-1} = \begin{bmatrix} 2 & 7 \\ 1 & 3 \end{bmatrix}$

9. Solutions will vary.

11. Likely some entries that should be 0 will not be exactly 0, but rather very small values.

Chapter 6

Section 6.1

1. A is symmetric. $\begin{bmatrix} -7 & 4 \\ 4 & -6 \end{bmatrix}$

3. A is diagonal, as is A^T. $\begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix}$

5. $\begin{bmatrix} -5 & 3 & -10 \\ -9 & 1 & -8 \end{bmatrix}$

7. $\begin{bmatrix} 4 & -9 \\ -7 & 6 \\ -4 & 3 \\ -9 & -9 \end{bmatrix}$

9. $\begin{bmatrix} -7 \\ -8 \\ 2 \\ -3 \end{bmatrix}$

11. $\begin{bmatrix} -9 & 6 & -8 \\ 4 & -3 & 1 \\ 10 & -7 & -1 \end{bmatrix}$

13. A is symmetric. $\begin{bmatrix} 4 & 0 & -2 \\ 0 & 2 & 3 \\ -2 & 3 & 6 \end{bmatrix}$

15. $\begin{bmatrix} 2 & 5 & 7 \\ -5 & 5 & -4 \\ -3 & -6 & -10 \end{bmatrix}$

17. $\begin{bmatrix} 4 & 5 & -6 \\ 2 & -4 & 6 \\ -9 & -10 & 9 \end{bmatrix}$

19. A is upper triangular; A^T is lower triangular. $\begin{bmatrix} -3 & 0 & 0 \\ -4 & -3 & 0 \\ -5 & 5 & -3 \end{bmatrix}$

21. A is diagonal, as is A^T. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

23. A is skew symmetric. $\begin{bmatrix} 0 & -1 & 2 \\ 1 & 0 & -4 \\ -2 & 4 & 0 \end{bmatrix}$

Section 6.2

1. 6
3. 3
5. -9
7. 1
9. Not defined; the matrix must be square.
11. -23
13. 4
15. 0
17. (a) tr(A)=8; tr(B)=-2; tr($A+B$)=6
    (b) tr(AB) = 53 = tr(BA)
19. (a) tr(A)=-1; tr(B)=6; tr($A+B$)=5
    (b) tr(AB) = 201 = tr(BA)

Section 6.3

1. 34
3. -44
5. -44
7. 28
9. (a) The submatrices are $\begin{bmatrix} 7 & 6 \\ 6 & 10 \end{bmatrix}$, $\begin{bmatrix} 3 & 6 \\ 1 & 10 \end{bmatrix}$, and $\begin{bmatrix} 3 & 7 \\ 1 & 6 \end{bmatrix}$, respectively.
   (b) $C_{1,2} = 34$, $C_{1,2} = -24$, $C_{1,3} = 11$
11. (a) The submatrices are $\begin{bmatrix} 3 & 10 \\ 3 & 9 \end{bmatrix}$, $\begin{bmatrix} -3 & 10 \\ -9 & 9 \end{bmatrix}$, and $\begin{bmatrix} -3 & 3 \\ -9 & 3 \end{bmatrix}$, respectively.
    (b) $C_{1,2} = -3$, $C_{1,2} = -63$, $C_{1,3} = 18$
13. -59
15. 15
17. 3
19. 0
21. 0
23. -113
25. Hint: $C_{1,1} = d$.

Section 6.4

1. 84
3. 0
5. 10
7. 24
9. 175
11. -200
13. 34
15. (a) $\det(A) = 41$; $R_2 \leftrightarrow R_3$
    (b) $\det(B) = 164$; $-4R_3 \rightarrow R_3$
    (c) $\det(C) = -41$; $R_2 + R_1 \rightarrow R_1$
17. (a) $\det(A) = -16$; $R_1 \leftrightarrow R_2$ then $R_1 \leftrightarrow R_3$
    (b) $\det(B) = -16$; $-R_1 \rightarrow R_1$ and $-R_2 \rightarrow R_2$
    (c) $\det(C) = -432$; $C = 3 * M$
19. $\det(A) = 4$, $\det(B) = 4$, $\det(AB) = 16$
21. $\det(A) = -12$, $\det(B) = 29$, $\det(AB) = -348$
23. -59
25. 15
27. 3
29. 0

Section 6.5

1. (a) $\det(A) = 14$, $\det(A_1) = 70$, $\det(A_2) = 14$
   (b) $\vec{x} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$
3. (a) $\det(A) = 0$, $\det(A_1) = 0$, $\det(A_2) = 0$
   (b) Infinite solutions exist.
5. (a) $\det(A) = 16$, $\det(A_1) = -64$, $\det(A_2) = 80$
   (b) $\vec{x} = \begin{bmatrix} -4 \\ 5 \end{bmatrix}$
7. (a) $\det(A) = -123$, $\det(A_1) = -492$, $\det(A_2) = 123$, $\det(A_3) = 492$
   (b) $\vec{x} = \begin{bmatrix} 4 \\ -1 \\ -4 \end{bmatrix}$
9. (a) $\det(A) = 56$, $\det(A_1) = 224$, $\det(A_2) = 0$, $\det(A_3) = -112$

(b) $\vec{x} = \begin{bmatrix} 4 \\ 0 \\ -2 \end{bmatrix}$

11. (a) $\det(A) = 0$, $\det(A_1) = 147$, $\det(A_2) = -49$, $\det(A_3) = -49$

(b) No solution exists.

Chapter 7

Section 7.1

1. For $z = 2 + 3i$ and $w = 4i$

- $z + w = 2 + 7i$
- $zw = -12 + 8i$
- $z^2 = -5 + 12i$
- $\frac{1}{z} = \frac{2}{13} - \frac{3}{13}i$
- $\frac{z}{w} = \frac{3}{4} - \frac{1}{2}i$
- $\frac{w}{z} = \frac{12}{13} + \frac{8}{13}i$
- $\bar{z} = 2 - 3i$
- $z\bar{z} = 13$
- $(\bar{z})^2 = -5 - 12i$

3. For $z = i$ and $w = -1 + 2i$

- $z + w = -1 + 3i$
- $zw = -2 - i$
- $z^2 = -1$
- $\frac{1}{z} = -i$
- $\frac{z}{w} = \frac{2}{5} - \frac{1}{5}i$
- $\frac{w}{z} = 2 + i$
- $\bar{z} = -i$
- $z\bar{z} = 1$
- $(\bar{z})^2 = -1$

5. For $z = 3 - 5i$ and $w = 2 + 7i$

- $z + w = 5 + 2i$
- $zw = 41 + 11i$
- $z^2 = -16 - 30i$
- $\frac{1}{z} = \frac{3}{34} + \frac{5}{34}i$
- $\frac{z}{w} = -\frac{29}{53} - \frac{31}{53}i$
- $\frac{w}{z} = -\frac{29}{34} + \frac{31}{34}i$
- $\bar{z} = 3 + 5i$
- $z\bar{z} = 34$
- $(\bar{z})^2 = -16 + 30i$

7. For $z = \sqrt{2} - i\sqrt{2}$ and $w = \sqrt{2} + i\sqrt{2}$

- $z + w = 2\sqrt{2}$
- $zw = 4$
- $z^2 = -4i$
- $\frac{1}{z} = \frac{\sqrt{2}}{4} + \frac{\sqrt{2}}{4}i$
- $\frac{z}{w} = -i$
- $\frac{w}{z} = i$
- $\bar{z} = \sqrt{2} + i\sqrt{2}$
- $z\bar{z} = 4$
- $(\bar{z})^2 = 4i$

9. For $z = \frac{1}{2} + \frac{\sqrt{3}}{2}i$ and $w = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$

- $z + w = i\sqrt{3}$
- $zw = -1$
- $z^2 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$
- $\frac{1}{z} = \frac{1}{2} - \frac{\sqrt{3}}{2}i$
- $\frac{z}{w} = \frac{1}{2} - \frac{\sqrt{3}}{2}i$
- $\frac{w}{z} = \frac{1}{2} + \frac{\sqrt{3}}{2}i$
- $\bar{z} = \frac{1}{2} - \frac{\sqrt{3}}{2}i$
- $z\bar{z} = 1$
- $(\bar{z})^2 = -\frac{1}{2} - \frac{\sqrt{3}}{2}i$

11. $7i$

13. -10

15. -12

17. 3

19. $i^5 = i^4 \cdot i = 1 \cdot i = i$

21. $i^7 = i^4 \cdot i^3 = 1 \cdot (-i) = -i$

23. $i^{15} = \left(i^4\right)^3 \cdot i^3 = 1 \cdot (-i) = -i$

25. $i^{117} = \left(i^4\right)^{29} \cdot i = 1 \cdot i = i$

27. $x = \dfrac{2 \pm i\sqrt{14}}{3}$

29. $y = \pm 2, \pm i$

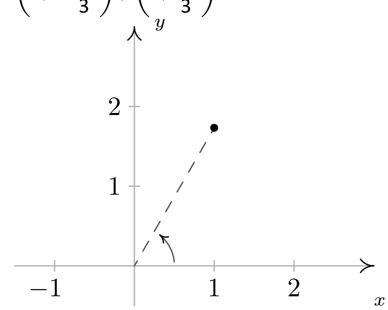
31. $y = \pm\dfrac{3i\sqrt{2}}{2}$

33. $x = \dfrac{\sqrt{5} \pm i\sqrt{3}}{2}$

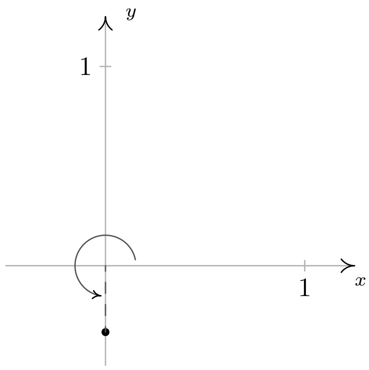
35. $z = \pm 2, \pm 2i$

Section 7.2

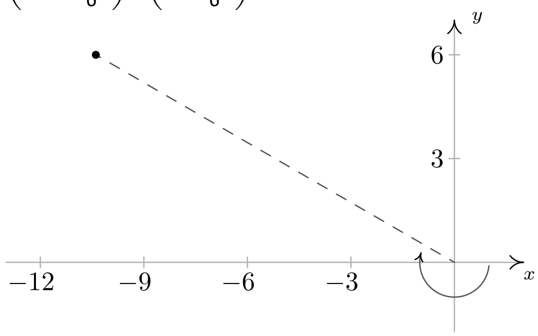
1. $\left(2, \dfrac{\pi}{3}\right)$, $\left(-2, \dfrac{4\pi}{3}\right)$
$\left(2, -\dfrac{5\pi}{3}\right)$, $\left(2, \dfrac{7\pi}{3}\right)$



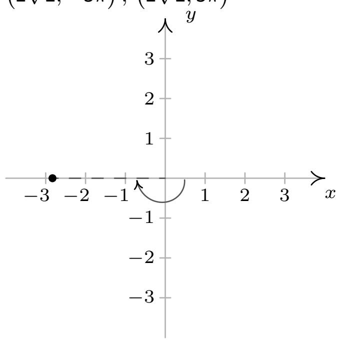
3. $\left(\dfrac{1}{3}, \dfrac{3\pi}{2}\right)$, $\left(-\dfrac{1}{3}, \dfrac{\pi}{2}\right)$
$\left(\dfrac{1}{3}, -\dfrac{\pi}{2}\right)$, $\left(\dfrac{1}{3}, \dfrac{7\pi}{2}\right)$



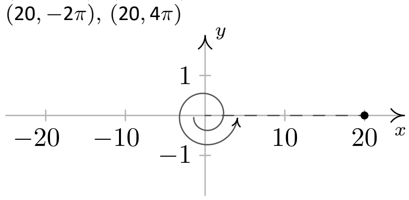
5. $\left(12, -\frac{7\pi}{6}\right)$, $\left(-12, \frac{11\pi}{6}\right)$
$\left(12, -\frac{19\pi}{6}\right)$, $\left(12, \frac{17\pi}{6}\right)$



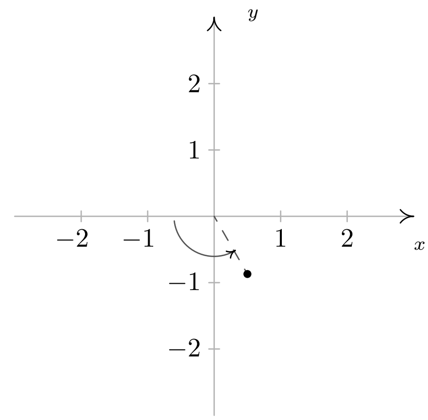
7. $\left(2\sqrt{2}, -\pi\right)$, $\left(-2\sqrt{2}, 0\right)$
$\left(2\sqrt{2}, -3\pi\right)$, $\left(2\sqrt{2}, 3\pi\right)$



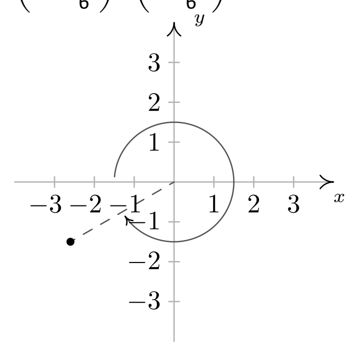
9. $(-20, 3\pi)$, $(-20, \pi)$
$(20, -2\pi)$, $(20, 4\pi)$



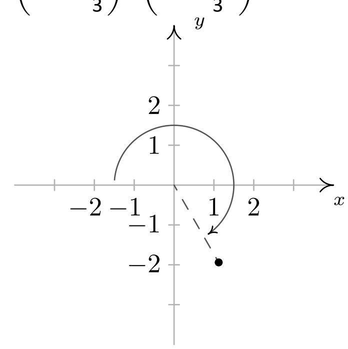
11. $\left(-1, \frac{2\pi}{3}\right)$, $\left(-1, \frac{2\pi}{3}\right)$
$\left(1, -\frac{\pi}{3}\right)$, $\left(1, \frac{11\pi}{3}\right)$



13. $\left(-3, -\frac{11\pi}{6}\right)$, $\left(-3, \frac{\pi}{6}\right)$
$\left(3, -\frac{5\pi}{6}\right)$, $\left(3, \frac{19\pi}{6}\right)$



15. $\left(-\sqrt{5}, -\frac{4\pi}{3}\right)$, $\left(-\sqrt{5}, \frac{2\pi}{3}\right)$
$\left(\sqrt{5}, -\frac{\pi}{3}\right)$, $\left(\sqrt{5}, \frac{11\pi}{3}\right)$



17. $\left(\frac{5\sqrt{2}}{2}, -\frac{5\sqrt{2}}{2}\right)$

19. $\left(-\frac{11\sqrt{3}}{2}, \frac{11}{2}\right)$

21. $\left(0, \frac{3}{5}\right)$

23. $(0, -9)$

25. $(21\sqrt{3}, 21)$

27. $\left(\frac{6\sqrt{5}}{5}, \frac{12\sqrt{5}}{5}\right)$

29. $\left(-\frac{9}{5}, -\frac{12}{5}\right)$

31. $\left(-\frac{4\sqrt{5}}{5}, \frac{2\sqrt{5}}{5}\right)$

33. $\left(\frac{4}{5}, \frac{3}{5}\right)$

35. $\left(\frac{\pi}{\sqrt{1+\pi^2}}, \frac{\pi^2}{\sqrt{1+\pi^2}}\right)$

37. $\left(5, \frac{\pi}{2}\right)$

39. $\left(7\sqrt{2}, \frac{7\pi}{4}\right)$

41. $(3, \pi)$

43. $\left(8, \frac{4\pi}{3}\right)$

45. $\left(\frac{3}{5}, \frac{4\pi}{3}\right)$

47. $\left(10, \arctan\left(\frac{4}{3}\right)\right)$

49. $\left(\sqrt{65}, \pi - \arctan\left(\frac{1}{8}\right)\right)$

51. $\left(13, \pi + \arctan\left(\frac{12}{5}\right)\right)$

53. $\left(25, 2\pi - \arctan\left(\frac{7}{24}\right)\right)$

55. $\left(\frac{\sqrt{2}}{2}, \frac{\pi}{3}\right)$

57. $r = 6\sec(\theta)$

59. $r = 7\csc(\theta)$

61. $\theta = \frac{3\pi}{4}$

63. $\theta = \arctan(2)$

65. $r = \sqrt{117}$

67. $x = \frac{1}{\cos(\theta) - 3\sin(\theta)}$

69. $r = 4\csc(\theta)\cot(\theta)$

71. $r = 4\cos(\theta)$

73. $r = 7\sin(\theta)$

75. $r = 6\sin(\theta)$

77. $x^2 + y^2 = 49$

79. $x^2 + y^2 = 2$

81. $y = -\sqrt{3}x$

83. $x = 0$

85. $5x^2 + 5y^2 = x$ or $\left(x - \frac{1}{10}\right)^2 + y^2 = \frac{1}{100}$

87. $x^2 + y^2 = -2y$ or $x^2 + (y+1)^2 = 1$

89. $y = \frac{1}{12}$

91. $y = -\sqrt{5}$

93. $y^2 = -x$

95. $\left(x^2 + 2x + y^2\right)^2 = x^2 + y^2$

97. Any point of the form $(0, \theta)$ will work, e.g. $(0, \pi)$, $(0, -117)$, $\left(0, \frac{23\pi}{4}\right)$ and $(0, 0)$.

Section 7.3

1. $z = 9 + 9i = 9\sqrt{2}\operatorname{cis}\left(\frac{\pi}{4}\right)$, $\operatorname{Re}(z) = 9$, $\operatorname{Im}(z) = 9$, $|z| = 9\sqrt{2}$, $\arg(z) = \left\{\frac{\pi}{4} + 2\pi k \mid k \text{ is an integer}\right\}$ and $\operatorname{Arg}(z) = \frac{\pi}{4}$.

3. $z = 6i = 6\operatorname{cis}\left(\frac{\pi}{2}\right)$, $\operatorname{Re}(z) = 0$, $\operatorname{Im}(z) = 6$, $|z| = 6$, $\arg(z) = \left\{\frac{\pi}{2} + 2\pi k \mid k \text{ is an integer}\right\}$ and $\operatorname{Arg}(z) = \frac{\pi}{2}$.

5. $z = -6\sqrt{3} + 6i = 12\operatorname{cis}\left(\frac{5\pi}{6}\right)$, $\operatorname{Re}(z) = -6\sqrt{3}$, $\operatorname{Im}(z) = 6$, $|z| = 12$, $\arg(z) = \left\{\frac{5\pi}{6} + 2\pi k \mid k \text{ is an integer}\right\}$ and $\operatorname{Arg}(z) = \frac{5\pi}{6}$.

7. $z = -\frac{\sqrt{3}}{2} - \frac{1}{2}i = \operatorname{cis}\left(\frac{7\pi}{6}\right)$, $\operatorname{Re}(z) = -\frac{\sqrt{3}}{2}$, $\operatorname{Im}(z) = -\frac{1}{2}$, $|z| = 1$, $\arg(z) = \left\{\frac{7\pi}{6} + 2\pi k \mid k \text{ is an integer}\right\}$ and $\operatorname{Arg}(z) = -\frac{5\pi}{6}$.

9. $z = -5i = 5\operatorname{cis}\left(\frac{3\pi}{2}\right)$, $\operatorname{Re}(z) = 0$, $\operatorname{Im}(z) = -5$, $|z| = 5$, $\arg(z) = \left\{\frac{3\pi}{2} + 2\pi k \mid k \text{ is an integer}\right\}$ and $\operatorname{Arg}(z) = -\frac{\pi}{2}$.

11. $z = 6 = 6\operatorname{cis}(0)$, $\operatorname{Re}(z) = 6$, $\operatorname{Im}(z) = 0$, $|z| = 6$, $\arg(z) = \{2\pi k \mid k \text{ is an integer}\}$ and $\operatorname{Arg}(z) = 0$.

13. $z = 3 + 4i = 5\operatorname{cis}\left(\arctan\left(\frac{4}{3}\right)\right)$, $\operatorname{Re}(z) = 3$, $\operatorname{Im}(z) = 4$, $|z| = 5$, $\arg(z) = \left\{\arctan\left(\frac{4}{3}\right) + 2\pi k \mid k \text{ is an integer}\right\}$ and $\operatorname{Arg}(z) = \arctan\left(\frac{4}{3}\right)$.

15. $z = -7 + 24i = 25\operatorname{cis}\left(\pi - \arctan\left(\frac{24}{7}\right)\right)$, $\operatorname{Re}(z) = -7$, $\operatorname{Im}(z) = 24$, $|z| = 25$, $\arg(z) = \left\{\pi - \arctan\left(\frac{24}{7}\right) + 2\pi k \mid k \text{ is an integer}\right\}$ and $\operatorname{Arg}(z) = \pi - \arctan\left(\frac{24}{7}\right)$.

17. $z = -12 - 5i = 13\operatorname{cis}\left(\pi + \arctan\left(\frac{5}{12}\right)\right)$, $\operatorname{Re}(z) = -12$, $\operatorname{Im}(z) = -5$, $|z| = 13$, $\arg(z) = \left\{\pi + \arctan\left(\frac{5}{12}\right) + 2\pi k \mid k \text{ is an integer}\right\}$ and $\operatorname{Arg}(z) = \arctan\left(\frac{5}{12}\right) - \pi$.

19. $z = 4 - 2i = 2\sqrt{5}\operatorname{cis}\left(\arctan\left(-\frac{1}{2}\right)\right)$, $\operatorname{Re}(z) = 4$, $\operatorname{Im}(z) = -2$, $|z| = 2\sqrt{5}$, $\arg(z) = \left\{\arctan\left(-\frac{1}{2}\right) + 2\pi k \mid k \text{ is an integer}\right\}$ and $\operatorname{Arg}(z) = \arctan\left(-\frac{1}{2}\right) = -\arctan\left(\frac{1}{2}\right)$.

21. $z = 6\operatorname{cis}(0) = 6$

23. $z = 7\sqrt{2}\operatorname{cis}\left(\frac{\pi}{4}\right) = 7 + 7i$

25. $z = 4\operatorname{cis}\left(\frac{2\pi}{3}\right) = -2 + 2i\sqrt{3}$

27. $z = 9\operatorname{cis}(\pi) = -9$

29. $z = 7\operatorname{cis}\left(-\frac{3\pi}{4}\right) = -\frac{7\sqrt{2}}{2} - \frac{7\sqrt{2}}{2}i$

31. $z = \frac{1}{2}\operatorname{cis}\left(\frac{7\pi}{4}\right) = \frac{\sqrt{2}}{4} - i\frac{\sqrt{2}}{4}$

33. $z = 8\operatorname{cis}\left(\frac{\pi}{12}\right) = 4\sqrt{2+\sqrt{3}} + 4i\sqrt{2-\sqrt{3}}$

35. $z = 5\operatorname{cis}\left(\arctan\left(\frac{4}{3}\right)\right) = 3 + 4i$

37. $z = 15\operatorname{cis}(\arctan(-2)) = 3\sqrt{5} - 6i\sqrt{5}$

39. $z = 50\operatorname{cis}\left(\pi - \arctan\left(\frac{7}{24}\right)\right) = -48 + 14i$

41. Since $z = -\frac{3\sqrt{3}}{2} + \frac{3}{2}i = 3\operatorname{cis}\left(\frac{5\pi}{6}\right)$ and $w = 3\sqrt{2} - 3i\sqrt{2} = 6\operatorname{cis}\left(-\frac{\pi}{4}\right)$, we have $zw = 18\operatorname{cis}\left(\frac{7\pi}{12}\right)$

43. Since $z = 3\operatorname{cis}\left(\frac{5\pi}{6}\right)$ and $w = 6\operatorname{cis}\left(-\frac{\pi}{4}\right)$, $\frac{w}{z} = 2\operatorname{cis}\left(\frac{11\pi}{12}\right)$

45. Since $z = 3\operatorname{cis}\left(\frac{5\pi}{6}\right)$ and $w = 6\operatorname{cis}\left(-\frac{\pi}{4}\right)$, $w^3 = 216\operatorname{cis}\left(-\frac{3\pi}{4}\right)$

47. Since $z = 3\operatorname{cis}\left(\frac{5\pi}{6}\right)$ and $w = 6\operatorname{cis}\left(-\frac{\pi}{4}\right)$, $z^3w^2 = 972\operatorname{cis}(0)$

49. Since $z = 3\operatorname{cis}\left(\frac{5\pi}{6}\right)$ and $w = 6\operatorname{cis}\left(-\frac{\pi}{4}\right)$, $\frac{w}{z^2} = \frac{2}{3}\operatorname{cis}\left(\frac{\pi}{12}\right)$

51. Since $z = 3\operatorname{cis}\left(\frac{5\pi}{6}\right)$ and $w = 6\operatorname{cis}\left(-\frac{\pi}{4}\right)$, $\frac{w^2}{z^3} = \frac{4}{3}\operatorname{cis}(\pi)$

53. $\left(-2 + 2i\sqrt{3}\right)^3 = 64$

55. $(-3 + 3i)^4 = -324$

57. $\left(\frac{5}{2} + \frac{5}{2}i\right)^3 = -\frac{125}{4} + \frac{125}{4}i$

59. $\left(\frac{3}{2} - \frac{3}{2}i\right)^3 = -\frac{27}{4} - \frac{27}{4}i$

61. $\left(\frac{\sqrt{2}}{2}+\frac{\sqrt{2}}{2}i\right)^4=-1$

63. $(\sqrt{3}-i)^5=-16\sqrt{3}-16i$

65. Since $z=4i=4\,\text{cis}\left(\frac{\pi}{2}\right)$ we have
$w_0=2\,\text{cis}\left(\frac{\pi}{4}\right)=\sqrt{2}+i\sqrt{2}$
$w_1=2\,\text{cis}\left(\frac{5\pi}{4}\right)=-\sqrt{2}-i\sqrt{2}$

67. Since $z=1+i\sqrt{3}=2\,\text{cis}\left(\frac{\pi}{3}\right)$ we have
$w_0=\sqrt{2}\,\text{cis}\left(\frac{\pi}{6}\right)=\frac{\sqrt{6}}{2}+\frac{\sqrt{2}}{2}i$
$w_1=\sqrt{2}\,\text{cis}\left(\frac{7\pi}{6}\right)=-\frac{\sqrt{6}}{2}-\frac{\sqrt{2}}{2}i$

69. Since $z=64=64\,\text{cis}\,(0)$ we have
$w_0=4\,\text{cis}\,(0)=4$
$w_1=4\,\text{cis}\left(\frac{2\pi}{3}\right)=-2+2i\sqrt{3}$
$w_2=4\,\text{cis}\left(\frac{4\pi}{3}\right)=-2-2i\sqrt{3}$

71. Since $z=i=\text{cis}\left(\frac{\pi}{2}\right)$ we have
$w_0=\text{cis}\left(\frac{\pi}{6}\right)=\frac{\sqrt{3}}{2}+\frac{1}{2}i$
$w_1=\text{cis}\left(\frac{5\pi}{6}\right)=-\frac{\sqrt{3}}{2}+\frac{1}{2}i$
$w_2=\text{cis}\left(\frac{3\pi}{2}\right)=-i$

73. Since $z=16=16\,\text{cis}\,(0)$ we have
$w_0=2\,\text{cis}\,(0)=2$
$w_1=2\,\text{cis}\left(\frac{\pi}{2}\right)=2i$
$w_2=2\,\text{cis}\,(\pi)=-2$
$w_3=2\,\text{cis}\left(\frac{3\pi}{2}\right)=-2i$

75. Since $z=64=64\,\text{cis}(0)$ we have
$w_0=2\,\text{cis}(0)=2$
$w_1=2\,\text{cis}\left(\frac{\pi}{3}\right)=1+\sqrt{3}i$
$w_2=2\,\text{cis}\left(\frac{2\pi}{3}\right)=-1+\sqrt{3}i$
$w_3=2\,\text{cis}\,(\pi)=-2$
$w_4=2\,\text{cis}\left(-\frac{2\pi}{3}\right)=-1-\sqrt{3}i$
$w_5=2\,\text{cis}\left(-\frac{\pi}{3}\right)=1-\sqrt{3}i$

77. Note: In the answers for w_0 and w_2 the first rectangular form comes from applying the appropriate Sum or Difference Identity ($\frac{\pi}{12}=\frac{\pi}{3}-\frac{\pi}{4}$ and $\frac{17\pi}{12}=\frac{2\pi}{3}+\frac{3\pi}{4}$, respectively) and the second comes from using the Half-Angle Identities.
$w_0=\sqrt[3]{2}\,\text{cis}\left(\frac{\pi}{12}\right)=\sqrt[3]{2}\left(\frac{\sqrt{6}+\sqrt{2}}{4}+i\left(\frac{\sqrt{6}-\sqrt{2}}{4}\right)\right)=$
$\sqrt[3]{2}\left(\frac{\sqrt{2+\sqrt{3}}}{2}+i\frac{\sqrt{2-\sqrt{3}}}{2}\right)$
$w_1=\sqrt[3]{2}\,\text{cis}\left(\frac{3\pi}{4}\right)=\sqrt[3]{2}\left(-\frac{\sqrt{2}}{2}+\frac{\sqrt{2}}{2}i\right)$
$w_2=\sqrt[3]{2}\,\text{cis}\left(\frac{17\pi}{12}\right)=\sqrt[3]{2}\left(\frac{\sqrt{2}-\sqrt{6}}{4}+i\left(\frac{-\sqrt{2}-\sqrt{6}}{4}\right)\right)=$
$\sqrt[3]{2}\left(\frac{\sqrt{2-\sqrt{3}}}{2}+i\frac{\sqrt{2+\sqrt{3}}}{2}\right)$

79.

81.

Chapter 8

Section 8.1

1. $\lambda=3$

3. $\lambda=0$

5. $\lambda=3$

7. $\vec{x}=\begin{bmatrix}-1\\2\end{bmatrix}$

9. $\vec{x}=\begin{bmatrix}3\\-7\\7\end{bmatrix}$

11. $\vec{x}=\begin{bmatrix}-1\\1\\1\end{bmatrix}$

13. $\lambda_1=4$ with $\vec{x_1}=\begin{bmatrix}9\\1\end{bmatrix}$;
$\lambda_2=5$ with $\vec{x_2}=\begin{bmatrix}8\\1\end{bmatrix}$

15. $\lambda_1=-3$ with $\vec{x_1}=\begin{bmatrix}-2\\1\end{bmatrix}$;
$\lambda_2=5$ with $\vec{x_2}=\begin{bmatrix}6\\1\end{bmatrix}$

17. $\lambda_1=2$ with $\vec{x_1}=\begin{bmatrix}1\\1\end{bmatrix}$;
$\lambda_2=4$ with $\vec{x_2}=\begin{bmatrix}-1\\1\end{bmatrix}$

19. $\lambda_1=-1$ with $\vec{x_1}=\begin{bmatrix}1\\2\end{bmatrix}$;
$\lambda_2=-3$ with $\vec{x_2}=\begin{bmatrix}1\\0\end{bmatrix}$

21. $\lambda_1=3$ with $\vec{x_1}=\begin{bmatrix}-3\\0\\2\end{bmatrix}$;
$\lambda_2=4$ with $\vec{x_2}=\begin{bmatrix}-5\\-1\\1\end{bmatrix}$
$\lambda_3=5$ with $\vec{x_3}=\begin{bmatrix}1\\0\\0\end{bmatrix}$

23. $\lambda_1=-5$ with $\vec{x_1}=\begin{bmatrix}24\\13\\8\end{bmatrix}$;
$\lambda_2=-2$ with $\vec{x_2}=\begin{bmatrix}6\\5\\1\end{bmatrix}$
$\lambda_3=3$ with $\vec{x_3}=\begin{bmatrix}0\\1\\0\end{bmatrix}$

25. $\lambda_1=-2$ with $\vec{x_1}=\begin{bmatrix}0\\0\\1\end{bmatrix}$;
$\lambda_2=1$ with $\vec{x_2}=\begin{bmatrix}0\\3\\5\end{bmatrix}$
$\lambda_3=5$ with $\vec{x_3}=\begin{bmatrix}28\\7\\1\end{bmatrix}$

27. $\lambda_1=-2$ with $\vec{x}=\begin{bmatrix}1\\0\\1\end{bmatrix}$;
$\lambda_2=3$ with $\vec{x}=\begin{bmatrix}1\\1\\1\end{bmatrix}$;
$\lambda_3=5$ with $\vec{x}=\begin{bmatrix}0\\1\\1\end{bmatrix}$

Section 8.2

1. (a) $\lambda_1=1$ with $\vec{x_1}=\begin{bmatrix}4\\1\end{bmatrix}$;
$\lambda_2=4$ with $\vec{x_2}=\begin{bmatrix}1\\1\end{bmatrix}$
(b) $\lambda_1=1$ with $\vec{x_1}=\begin{bmatrix}-1\\1\end{bmatrix}$;
$\lambda_2=4$ with $\vec{x_2}=\begin{bmatrix}-1\\4\end{bmatrix}$

(c) $\lambda_1 = 1/4$ with $\vec{x_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$;

$\lambda_2 = 1$ with $\vec{x_2} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$

(d) 5

(e) 4

3. (a) $\lambda_1 = -1$ with $\vec{x_1} = \begin{bmatrix} -5 \\ 1 \end{bmatrix}$;

$\lambda_2 = 0$ with $\vec{x_2} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$

(b) $\lambda_1 = -1$ with $\vec{x_1} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$;

$\lambda_2 = 0$ with $\vec{x_2} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$

(c) $A$is not invertible.

(d) -1

(e) 0

5. (a) $\lambda_1 = -4$ with $\vec{x_1} = \begin{bmatrix} -7 \\ -7 \\ 6 \end{bmatrix}$;

$\lambda_2 = 3$ with $\vec{x_2} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$\lambda_3 = 4$ with $\vec{x_3} = \begin{bmatrix} 9 \\ 1 \\ 22 \end{bmatrix}$

(b) $\lambda_1 = -4$ with $\vec{x_1} = \begin{bmatrix} -1 \\ 9 \\ 0 \end{bmatrix}$;

$\lambda_2 = 3$ with $\vec{x_2} = \begin{bmatrix} -20 \\ 26 \\ 7 \end{bmatrix}$

$\lambda_3 = 4$ with $\vec{x_3} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$

(c) $\lambda_1 = -1/4$ with $\vec{x_1} = \begin{bmatrix} -7 \\ -7 \\ 6 \end{bmatrix}$;

$\lambda_2 = 1/3$ with $\vec{x_2} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$\lambda_3 = 1/4$ with $\vec{x_3} = \begin{bmatrix} 9 \\ 1 \\ 22 \end{bmatrix}$

(d) 3

(e) -48

Index