LINEAR ALGEBRA SUPPLEMENT FOR MATH 9

Acknowledgments. This material was developed specifically for Math 9 at Dartmouth. We are very grateful to Lyryx Learning for providing the source files to enable us to format the material in textbook style. We also used excerpts from the Lyryx text "A First Course in Linear Algebra" especially in Section 2.1 on matrix arithmetic.

Contents

0.1 Introduction

Long before you were introduced to calculus, you learned about straight lines and their slopes, and you learned how to specify a straight line via a function such as $f(x) = 5x$. Then in AB/BC calculus, you learned how to approximate graphs of complicated functions locally by straight lines.

What do we mean by locally? Suppose that you lived at a point A on a giant smooth curve. If you zoom in to a tiny neighborhood of your home at A, it appears almost to be a segment of a straight line as in Figure 1.1. (The second drawing in Figure 1.1 is a close-up view of what the detective is seeing through his magnifying glass.) It is not until you zoom out that you realize just how hilly your world really is.

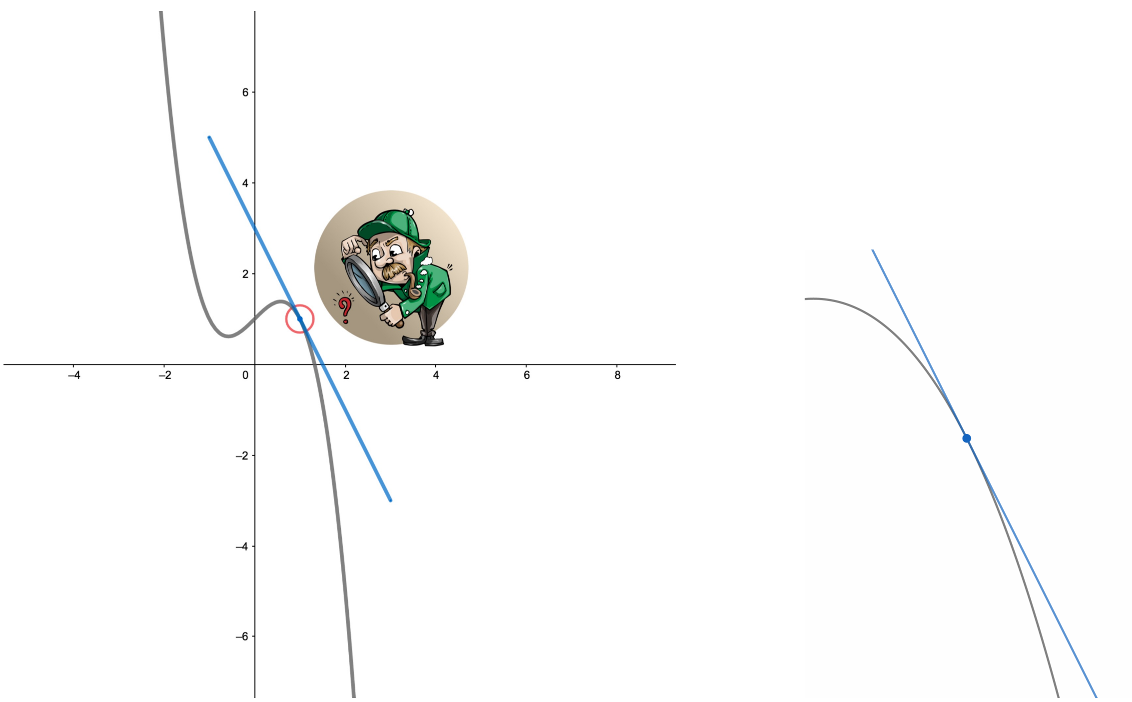


Figure 1

This is the premise of a differentiable function. At any point A, it is locally very close to being a line, specifically, the tangent line to the curve at A.

When we study functions of two variables, the graph can be thought of as a surface sitting in 3-dimensional space. Suppose that you live at a point A on this surface. If the function is differentiable (as we will define later), a tiny neighborhood of your home now resembles a **plane** as in Figure 2.

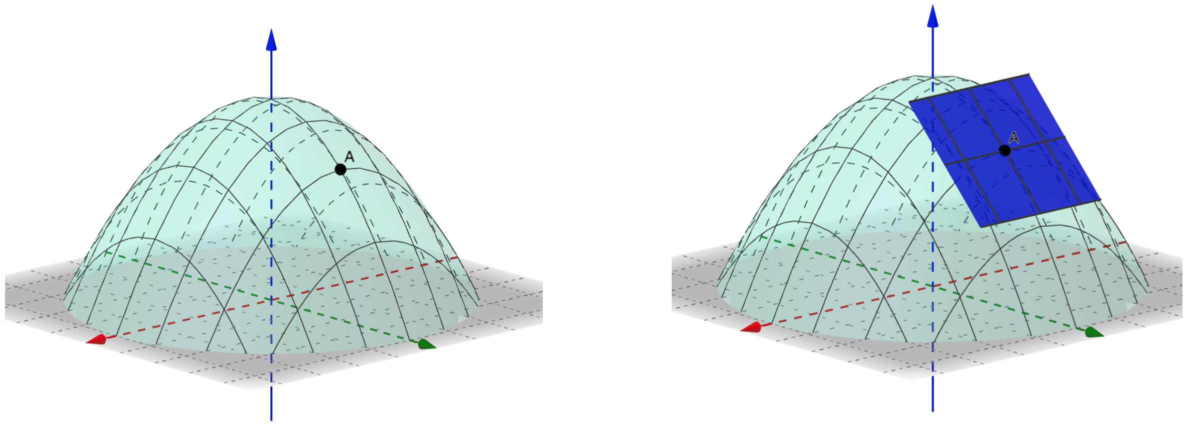


Figure 2

Analogous to the tangent line of a curve, there are tangent planes for surfaces. Just as we needed to understand lines – in particular their slopes – to define the derivative of a real-valued function of one variable, we need what are called *linear functions* or *linear transformations* to define the derivative of functions involving more variables. In these notes, we will develop the necessary framework for these constructions: **Linear Algebra**.

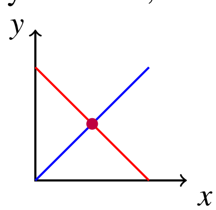
1. Systems of Equations

1.1 Systems of Equations, Geometry

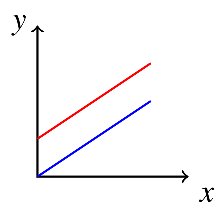
An equation of the form $ax+by=c$ where a, b and c are constants is called a linear equation in two variables. As you recall, equations of this form represent straight lines in the xy-plane. A system of linear equations consists of two or more linear equations, e.g.,

$$\begin{array}{rcrcr} x & + & y & = & 7 \\ 2x & - & 3y & = & -6 \end{array}$$

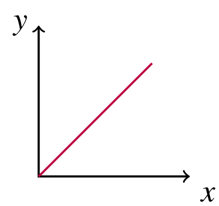
A solution of a system of linear equations is an ordered pair (x,y) that satisfies all the equations simultaneously. When we solve the system, we are finding the points of intersection of the lines represented by the equations. A system of 2 linear equations may have exactly one solution, no solutions or infinitely many solutions, as illustrated below:



One Solution



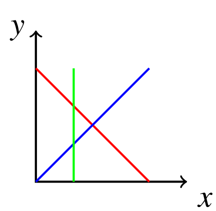
No Solutions



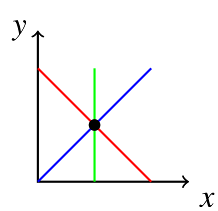
Infinitely Many Solutions

(In the third picture, the two equations represent the same line, so every point on the line is a solution. This will happen if the two equations are equivalent, i.e., one is a multiple of the other. For example $x+2y=3$ and $2x+4y=6$ are equivalent.)

If we have a system of three linear equations in two variables, then any solution must be a point of intersection of all three lines. Once again, the only possibilities are that the system has exactly one solution, no solutions or infinitely many solutions. The third possibility only occurs if all three equations are equivalent. The first two possibilities are depicted below.



No Solution



One Solution

Pushing further, we can consider systems of equations in *three* variables x, y, and z. This requires us to incorporate a third axis for the third variable, to think in three-dimensional space. Just as equations in two variables like $2x+3y=6$ correspond to lines, equations in three variables like $2x+4y-5z=8$ correspond to planes (as we will see later). If two such planes intersect, they do so along an entire line that contains all points (x,y,z) that satisfy both equations, as shown below:

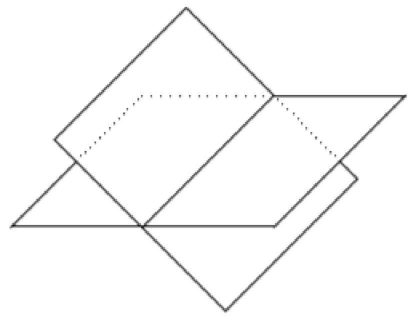


Figure 1.1

It could also happen that the two planes fail to intersect (if they are parallel). However, it is not possible for two planes to intersect at a single point. (Take a minute to convince yourself of this fact.) Therefore, systems of two equations in three variables can either have zero or infinitely many solutions.

To consider a system of three equations in three variables we must add a third equation and its respective plane. As shown below, the solution can be a single point, a single line, or nonexistent.

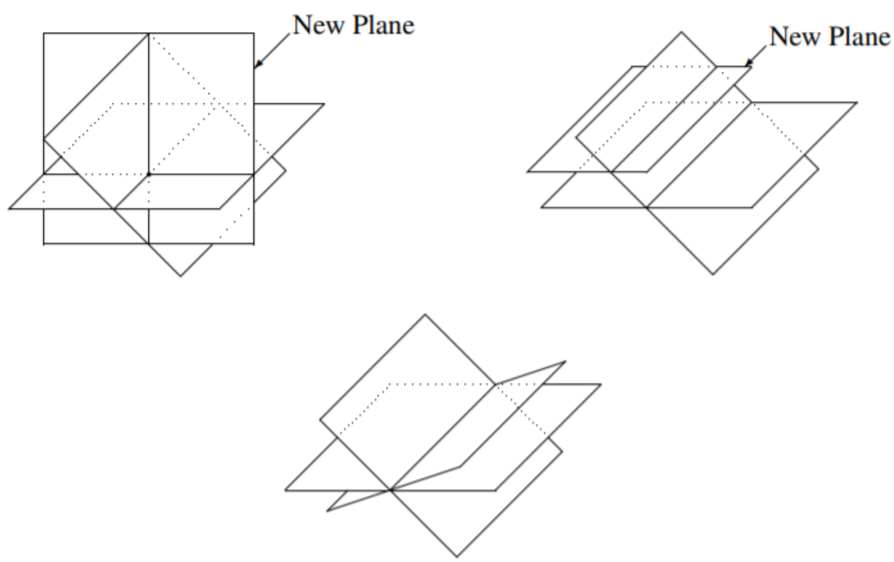


Figure 1.2

One can also consider systems of linear equations in more than 3 variables. While an equation in four or more variables such as

$$3x_1+2x_2+5x_3+4x_4=100$$

no longer has a geometric interpretation, such equations arise in practically every area.. The variables may represent, for example, amounts of various commodities or perhaps different types of molecules with the coefficients being, say, the cost of each commodity or the number of hydrogen atoms in each type of molecule.

We say a system of equations is consistent if it has at least one solution and inconsistent if it has no solutions. In the next section, we will develop a method for checking whether a system of linear equations is inconsistent and finding all solutions if so. A common feature of all systems of linear equations in any number of variables is that if the system is consistent, it either has only one solution or else infinitely many solutions.

1.2 Systems Of Equations, Algebraic Procedures

Outcomes

1. *Encode systems of linear equations into augmented matrices*
2. *Be able to recognize when an augmented matrix is in row-echelon form and when it is in reduced row-echelon form .*
3. *Understand elementary row operations and apply them to convert any augmented matrix into reduced row-echelon form .*
4. *Use these tools to solve systems of linear equations.*

We will introduce a widely-used systematic procedure for solving systems of linear equations. It is especially useful when dealing with systems of linear equations involving several variables and/or several equations. The method consists of:

1. Write down a rectangular array of numbers, called an augmented matrix, that encodes all the information in the given system of linear equations.
2. Use "elementary row operations" to convert the augmented matrix into nicer forms called row-echelon form and reduced row-echelon form .
3. Read off the solution of the system from the augmented matrix in row-echelon form or reduced row-echelon form .

Outline of this section:

- In Subsection 1.2.1, we introduce augmented matrices and elementary row operations and motivate the ideas using examples involving only two equations in two variables.
- In Subsection 1.2.2, we first define row-echelon form and reduced row-echelon form . We then explain how to read off the solution of a system of linear equations once the augmented matrix has been converted to one of thse forms.
- In Subsection 1.2.3, we give a systematic method for using elementary row operations in order to reach row-echelon form and reduced row-echelon form . We then put all the steps together to solve systems of linear equations.

1.2.1. Augmented matrices and elementary row operations

. To motivate the technique, we begin with systems of two equations in two variables. We will review the familiar steps for solving such systems and then see how to encode the system and how to carry out the analogous steps on the encoded version.

Example 1.1: Solving a system of linear equations

Solve the following system of linear equations

$$\begin{array}{rcrcr} x & + & y & = & 7 \\ 2x & - & 3y & = & -6 \end{array}$$

Solution. *To begin, we can subtract twice the first equation from the second equation:*

$$\begin{array}{rrcrcrl} -2(& x & + & y & = & 7 &) \\ & 2x & - & 3y & = & -6 & \\ \hline & 0x & - & 5y & = & -20 & \end{array}$$

The system then becomes:

$$\begin{array}{rcrcr} x & + & y & = & 7 \\ & - & 5y & = & -20 \end{array}$$

Next, we can multiply the second equation by $-\frac{1}{5}$ *to get rid of the coefficients, yielding:*

$$\begin{array}{rcrcr} x & + & y & = & 7 \\ & & y & = & 4 \end{array}$$

Finally, we can subtract the second equation from the first:

$$\begin{array}{rrcrcrl} & x & + & y & = & 7 & \\ -1(& & & y & = & 4 &) \\ \hline & x & + & 0y & = & 3 & \end{array}$$

So we end up with:

$$\begin{array}{rcrcr} x & & & = & 3 \\ & & y & = & 4 \end{array}$$

Thus, we have arrived at the "simple" form of the system of equations, from which it is easy to see that the system has a unique solution $(3,4)$.

♠

The process described above is effective, although a little burdensome especially when we have more variables and equations. Notice that throughout the steps we kept variables and constants in the same relative position to each other. If we abstract from this concept and keep track of coefficients only, our work will be much easier. This is what augmented matrices provide:

Example 1.2: Encoding systems of linear equations by augmented matrices

Given a system of linear equations

$$\begin{array}{rcrcrl} 1x & + & 1y & = & 7 & \qquad (E_1) \\ 2x & - & 3y & = & -6 & \qquad (E_2) \end{array}$$

we create what is known as an augmented matrix that captures all the information given in the equations:

$$\left[\begin{array}{rr|r} 1 & 1 & 7 \\ 2 & -3 & -6 \end{array}\right] \qquad \begin{array}{l} (R_1) \\ (R_2) \end{array}$$

(A matrix is a rectangular array of numbers. The word "augmented" here refers to the fact that there is an extra column separated by a vertical bar from the other columns. You can think of the bar as representing the equal signs in the equations.)

Observe that

- *Each row of the augmented matrix corresponds to one of the equations in the system.*
- *Each column to the left of the bar corresponds to one of the variables (in this example, x or y). More precisely the column gives the coefficients of the corresponding variable in the various equations.*

In Example 1.1, we solved the system by a series of manipulations involving multiplying equations by constants and adding multiples of one equation to another. Sometimes it is also convenient to reorder the equations. The elementary row operations that we now introduce give the corresponding effect of these manipulations on the augmented matrices.

Definition 1.3: Elementary Operations

Elementary row operations are those operations of any of the following types:

1. *Interchange the order in which the rows are listed.*
2. *Multiply any row by a nonzero number.*
3. *Replace any row with itself added to a multiple of another row.*

We abbreviate these elementary row operations as follows:

- $R_i \leftrightarrow R_j$ denotes interchanging rows i and j. For example, $R_1 \leftrightarrow R_2$ denotes interchanging rows 1 and 2.
- $R_i \to cR_i$ denotes multiplying row i by the real number c. For example, $R_2 \to 5R_2$ means we're multiplying row 2 by 5.
- $R_i \to R_i + aR_j$ denotes adding a multiple of row j to row i. For example, $R_1 \to R_1 + 4R_2$ means we're adding 4 times row 2 to row 1.

Example 1.4: Elementary Operations at work

Consider the system in Example 1.1 and its augmented matrix:

$$\begin{array}{rcrcr} x & + & y & = & 7 \\ 2x & - & 3y & = & -6 \end{array} \qquad \left[\begin{array}{rr|r} 1 & 1 & 7 \\ 2 & -3 & -6 \end{array}\right]$$

Subtracting twice the first equation from the second corresponds to the type 3 elementary row operation $R_2 \to R_2 - 2R_1$:

$$\begin{array}{rcrcr} x & + & y & = & 7 \\ & - & 5y & = & -20 \end{array} \qquad \left[\begin{array}{rr|r} 1 & 1 & 7 \\ 0 & -5 & -20 \end{array}\right]$$

Multiplying the second equation by $-\frac{1}{5}$ is an instance of the type 2 elementary row operation ($R_2 \to -\frac{1}{5}R_2$):

$$\begin{array}{rcrcr} x & + & y & = & 7 \\ & & y & = & 4 \end{array} \qquad \left[\begin{array}{rr|r} 1 & 1 & 7 \\ 0 & 1 & 4 \end{array}\right]$$

Subtracting the second equation corresponds to the type 3 elementary row operation $R_1 \to R_1 - R_2$:

$$\begin{array}{rcrcr} x & & & = & 3 \\ & & y & = & 4 \end{array} \qquad \left[\begin{array}{rr|r} 1 & 0 & 3 \\ 0 & 1 & 4 \end{array}\right]$$

Observe that once we converted the augmented matrix into the form

$$\left[\begin{array}{rr|r} 1 & 0 & a \\ 0 & 1 & b \end{array}\right]$$

we could then read off the solution $x = a$, $y = b$.

Now that we are familiar with augmented matrices and elementary row operations, we can try to solve a system solely through matrix manipulation:

Example 1.5: Solving systems with matrices

Solve the following system of equations using elementary row operations:

$$\begin{array}{rcrcrl} 2x & + & 3y & = & 7 & \qquad (E_1) \\ x & - & y & = & 1 & \qquad (E_2) \end{array}$$

Solution. We begin by constructing the augmented matrix representation of the system of equations above:

$$\left[\begin{array}{rr|r} 2 & 3 & 7 \\ 1 & -1 & 1 \end{array}\right] \qquad \begin{array}{l} (R_1) \\ (R_2) \end{array}$$

Can we use elementary row operations to convert this matrix into the form $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \middle| \begin{smallmatrix} a \\ b \end{smallmatrix}\right]$?

We first try to get the first column into the form $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$. We have a few options to get a 1 in the upper left corner. We could multiply the first row by $\frac{1}{2}$ but that will introduce fractions, which we might want to avoid as long as possible. We could subtract the second equation from the first or we could interchange the equations. All these ways are valid. We'll do the third. Once we get the "1" in place, we will do a type 3 operation to get a zero in the second row:

$$\left[\begin{array}{cc|c} 2 & 3 & 7 \\ 1 & -1 & 1 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{cc|c} 1 & -1 & 1 \\ 2 & 3 & 7 \end{array}\right] \xrightarrow{R_2 \to R_2 - 2R_1} \left[\begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 5 & 5 \end{array}\right]$$

Our next goal is to get a one in the lower right corner and then use a type 2 row operation to get a zero above it. Continuing from where we left off:

$$\left[\begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 5 & 5 \end{array}\right] \xrightarrow{R_2 \to \frac{1}{5}R_2} \left[\begin{array}{cc|c} 1 & -1 & 1 \\ 0 & 1 & 1 \end{array}\right] \xrightarrow{R_1 \to R_1 + R_2} \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \end{array}\right]$$

Converting back to equation form, we thus get $x = 2$, $y = 1$. Thus the system has a unique solution $(2,1)$.

♠

Remark 1.6

An option in the example above is to stop applying elementary row operations when we reach the point $\left[\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix} \,\middle|\, \begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$. *You can then decode the augmented matrix to get* $x - y = 1$ *and* $y = 1$. *Substituting* $y = 1$ *into the first of these equations yields* $x = 2$. *This step is called* back substitution.

Example 1.7: Systems with no solution

Solve the following system of equations using elementary row operations.

$$\begin{array}{rcrcr} 2x & + & y & = & 7 \\ -4x & & -2y & = & -13 \end{array}$$

Solution.

We apply elementary row operations to the corresponding augmented matrix:

$$\left[\begin{array}{cc|c} 2 & 1 & 7 \\ -4 & -2 & -13 \end{array}\right] \xrightarrow{R_1 \to \frac{1}{2}R_1} \left[\begin{array}{cc|c} 1 & \frac{1}{2} & \frac{7}{2} \\ -4 & -2 & -13 \end{array}\right] \xrightarrow{R_2 \to R_2 + 4R_1} \left[\begin{array}{cc|c} 1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{array}\right]$$

The second row implies $0x + 0y = 1$, *which clearly cannot be the case. Therefore there are no solutions to the system.*

♠

Summary. In the first two examples, we were able to convert the augmented matrix to the form $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \,\middle|\, \begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and then read off the solution $x = a$, $y = b$. In the third example, however, we instead ended up with an augmented matrix of the form $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix} \,\middle|\, \begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$, enabling us to see that there was no solution.

More generally, we want to convert any augmented matrix arising from a system of linear equations into a nice form that will enable us either to read off the solution or else show that there are no solutions. As the number of equations and/or variables increases, so does the variety of possible "nice" forms. In the next subsection, we introduce the types of augmented matrices that play the role of the "nice" ones.

1.2.2. The goal: row-echelon form or reduced row-echelon form

In this subsection, we first introduce row-echelon form and reduced row-echelon form of augmented matrices and then explain how to read off the solution of a system of linear equations once the augmented matrix is in one of these forms.

1.2.2.1. What are row-echelon form and reduced row-echelon form ?

Definition 1.8: Row-Echelon Form and Reduced Row-Echelon Form

An augmented matrix is in row-echelon form if

1. *All rows that consist only of zeros are at the bottom.*
2. *The first nonzero entry in any nonzero row is a* 1*. We call it a pivot.*
3. *The pivot in any nonzero row is further to the right than that of the row above it.*
4. *In any column containing a pivot, all entries below the pivot are zeros. (The columns containing pivots are called pivot columns.)*

We say the augmented matrix is in reduced row-echelon form if in addition to the conditions above, every entry in a pivot column except for the pivot itself is zero.

The following examples describe matrices in these various forms. As an exercise, take the time to carefully verify that they are in the specified form.

Example 1.9: Not in Row-Echelon Form

The following augmented matrices are not in row-echelon form (and therefore also not in reduced row-echelon form).

$$\left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{cc|r} 1 & 2 & 3 \\ 2 & 4 & -6 \\ 4 & 0 & 7 \end{array}\right], \left[\begin{array}{ccc|c} 0 & 2 & 3 & 3 \\ 1 & 5 & 0 & 2 \\ 7 & 5 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right]$$

Example 1.10: Matrices in Row-Echelon Form

The following augmented matrices are in row-echelon form, but not in reduced row-echelon form.

$$\left[\begin{array}{ccccc|c} 1 & 2 & 6 & 5 & 8 & 2 \\ 0 & 0 & 1 & 2 & 7 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{ccc|c} 1 & 3 & 5 & 4 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{ccc|c} 1 & 0 & 6 & 4 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

In the first of these augmented matrices, the pivot columns are the first, third and fifth columns. In the second one, every column is a pivot column. In the third one, the first 3 columns are pivot columns.

Example 1.11: Matrices in Reduced Row-Echelon Form

The following augmented matrices are in reduced row-echelon form.

$$\left[\begin{array}{ccccc|c} 1 & 2 & 0 & 5 & 0 & 2 \\ 0 & 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right], \left[\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Notice that the difference between these augmented matrices and the ones in the previous example is that all entries above any pivot are zeros.

1.2.2.2. Reading off the solution once the augmented matrix is in row-echelon form or reduced row-echelon form

In the next subsection, we will give an algorithm for applying elementary row operations to reduce the original augmented matrix to row-echelon form and then to reduced row-echelon form. The reduced row-echelon form of the augmented matrix will always be unique, but the row-echelon form is not. Once the augmented matrix first reaches row-echelon form, it will remain in row-echelon form through the remainder of the elementary row operations that take it to reduced row-echelon form.

In this subsection, we see how to determine the solution of the system of linear equations from the reduced row-echelon form or, if you prefer, from the row-echelon form. As you will see, it's faster to read off the solution if you row reduce all the way to reduced row-echelon form. The trade-off is that it takes more elementary row operations to reach reduced row-echelon form.

We first address the question:

Question: How many solutions does a system of linear equations have?

Motivation

Before converting the augmented matrix of a system into row-echelon form, it is usually difficult to tell whether the system is consistent and how many solutions it has. You can think of each equation as placing a constraint on the values of the variables. If there are fewer constraints than variables, you expect infinitely many solutions; if there are the same number of constraints as variables, you expect one solution; and if there are more constraints than variables, you don't expect any solutions. However, these expectations don't always hold, because the different equations may contradict each other (making the system inconsistent even if the number of constraints is small) or some of the equations may be duplicating others, sometimes in a non-obvious way, so that you really have fewer constraints than at first appears.
Once the system is in echelon form, any inconsistencies show up in an obvious way and any duplications disappear into rows of zeros. Thus at that point, you can determine the number of solutions by comparing the number of constraints (non-zero rows) to the number of variables, as indicated in the theorem below.

Theorem 1.12: Determining the number of solutions

From an augmented matrix in row-echelon form, you can determine the number of solutions of the linear system as follows:

(a) If the augmented matrix in row-echelon form contains a row of the form $\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & 1 \end{array}\right]$*, then the system is inconsistent, i.e., it has no solutions.*

(b) If the number of non-zero rows equals the number of variables and if there is no row of the form $\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & 1 \end{array}\right]$*, then the system has exactly one solution.*

(c) If the number of non-zero rows is less than the number of variables and if there is no row of the form $\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & 1 \end{array}\right]$*, then the system has infinitely many solutions.*

These are the only possibilities that can occur.

Case (a). Example 1.7 illustrates Case (a). As you saw there, a row of the form $\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & 1 \end{array}\right]$ gives the contradiction $0 = 1$ when you decode it back into equation form. More generally, if in the process of carrying out elementary row operations, a row appears of the form $\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & c \end{array}\right]$ where c is any non-zero real number, then you immediately know that the system is inconsistent, since this row translates into the equation $0 = c$. Thus you can stop at that point.

We next discuss each of the cases (b) and (c) and see how to find the solutions in these cases.

Case (b). In this case, if the augmented matrix is in reduced row-echelon form, then ignoring the zero rows, it must look like the following:

$$\left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 3 \end{array}\right] \text{ if the number of variables is 2}$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 5 \end{array}\right] \text{ if the number of variables is 3}$$

and similarly if there are more variables. (Of course the constants in the column to the right of the bar can be anything.)

You can then immediately read off the solution; in the two cases above the solutions are $(2,3)$ and $(2,3,5)$.

The following example illustrates how to read off the solution in Case(b) if the augmented matrix is in row-echelon form but not reduced row-echelon form . One uses "back substitution" as in Remark 1.6.

Example 1.13: Back substitution

Suppose that you have a system of linear equations in the variables x, y, z and that the non-zero rows of the augmented matrix in row-echelon form are given by

$$\left[\begin{array}{ccc|c} 1 & 4 & 1 & 2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 5 \end{array}\right].$$

Find the solution of the system of equations.

Solution. Decoding the augmented matrix back to equation form, we have:

$$\begin{aligned} x+4y+z &= 2 \\ y+2z &= 3 \\ z &= 5 \end{aligned}$$

Substitute $z=5$ into the second equation to get $y=-7$. Then substitute $y=-7, z=5$ into the first equation, yielding $x=25$. Thus the solution is $(25,-7,5)$. ♠

Case (c).

In case (c) we will always have more columns to the left of the bar than we do non-zero rows.

Example 1.14: Infinitely many solutions

Suppose you start with a system of linear equations in three variables x, y, z and have reduced it to the following reduced row-echelon form :

$$\left[\begin{array}{ccc|c} 1 & 0 & 7 & 2 \\ 0 & 1 & 5 & 4 \end{array}\right]$$

Find the solution of the system of equations.

Solution. Decoding the augmented matrix back to equation form, we get

$$\begin{aligned} x+7z &= 2 \\ y+5z &= 4 \end{aligned}$$

We can rewrite this as

$$x = 2-7z$$

$$y = 4 - 5z$$

Notice that there are no restrictions on z. We call z a free variable and call x and y determined variables. We can choose z to be *any* real number and will still get a solution. For example, if we set $z = 2$, then we find that $x = -12$ and $y = -6$, so $(-12, -6, 2)$ is one of the solutions. Similarly, if we choose $z = 1$, we get a solution $(-5, -1, 1)$.

One way of writing down all the solutions is to give a name to the arbitrarily chosen value of z, writing, say $z = t$ where t is allowed to be any real number. We then get the solution

$$\boxed{(2 - 7t, 4 - 5t, t).}$$

You can also write this as

$$\boxed{x = 2 - 7t, \quad y = 4 - 5t, \quad z = t \text{ where } t \text{ is any real number}}$$

As we will discuss later, these solutions form a line. We call t a parameter.

♠

Example 1.15

Suppose you begin with a system of three equations in three variables x, y, z and after converting to reduced row-echelon form, you get

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Find the solution of the system of equations.

Solution. We can ignore the row of zeros. The system becomes:

$$x = 2$$

$$z = 4$$

Here x and z are determined variables. There is no restriction on y, so y is a free variable. No matter how y is chosen, we get a solution $(2, y, 4)$. You can write the solution this way, or you can again use a parameter t and write it as $(2, t, 4)$, or say:

$$\boxed{x = 2, \quad y = t, \quad z = 4 \text{ where } t \text{ is any real number}}$$

Aside: These solutions form a line parallel to the y-axis. ♠

Recall that each column to the left of the vertical bar corresponds to one of the variables in the system of linear equations. Notice the pattern in the examples:

- Pivot columns correspond to determined variables.
- Columns that do not contain a pivot correspond to free variables.

Example 1.16

Suppose you begin with a system of two equations in four variables x, y, z, w and after converting to reduced row-echelon form, you have

$$\left[\begin{array}{cccc|c} 1 & 4 & 0 & 2 & 1 \\ 0 & 0 & 1 & 3 & 2 \end{array}\right]$$

Find the solution of the system of equations.

Solution.

The first and third columns are pivot columns so x and z are determined variables, while y and w are free. The system now reads:

$$x + 4y + 2w = 1$$

$$z + 3w = 2$$

Thus

$$x = 1 - 4y - 2w$$

$$z = 2 - 3w$$

Since y and w can be anything, we write $y = s$, $w = t$ and write the solution as:

$$\boxed{x = 1 - 4s - 2t,\ y = s,\ z = 2 - 3t,\ w = t \ \text{ where } s \text{ and } t \text{ are any real numbers}}$$

♠

1.2.3. Gauss-Jordan Elimination

In the previous subsection, we saw how to determine the solution of a system of linear equations when it is in row-echelon form or in reduced row-echelon form. The algorithm below provides a method for using elementary row operations to convert an augmented matrix to row-echelon form and then further to reduced row-echelon form. We begin with the augmented matrix in its original form.

Algorithm 1.17: Gauss-Jordan Elimination

The following algorithm, called Gauss-Jordan Elimination gives a method for applying elementary row operations to convert any augmented matrix first to row-echelon form and then to reduced row-echelon form.

1. *Starting from the left, find the first nonzero column. This is the first pivot column, and the position at the top of this column is the first pivot position. Use elementary row operations to place a non-zero entry in the pivot position. (You can either make this entry a* 1 *at this point, e.g., by multiplying the row by a constant, or wait until step 4.)*
2. *Add multiples of the first row to the other rows in order to make all the entries in the first pivot column below the pivot position equal to zero.*
3. *Ignoring the first row, repeat steps 1 and 2 with the remaining rows. Then repeat the process ignoring the first two rows. Continue this way until there are no more non-zero rows left to modify.*
4. *Divide each nonzero row by a constant if needed so that the entry in the pivot position becomes a* 1*. The matrix will then be in row-echelon form. (See Remark 1.18, item (2) for an alternative.)*

The following step will carry the matrix from row-echelon form to reduced row-echelon form.

5. *Moving from right to left, use the third type of elementary row operation to create zeros in the entries of the pivot columns that are above the pivot positions. The result will be a matrix in reduced row-echelon form.*

Remark 1.18

1. *You have choices in how you carry out step one. Even if there is already a non-zero entry in the pivot position, you may want to switch rows for example in order to get a “friendlier” value there that will make computations easier as you continue to the next step.*
2. *If you are continuing all the way to reduced row-echelon form, it is okay to switch the order of steps four and five if this makes the computations easier.*
3. *You should do the elementary row operations one at a time, indicating what operation you are doing so that it is clear to the reader. The one exception to this is in Steps 2 and 5: In step 2, it’s okay to add appropriate multiples of row 1 to each of the other rows at once in order to get 0’s in all the entries below the pivot as long as you indicate what operations you are using. Similarly in step 5.*

Example 1.19

Solve the following system of linear equations:

$$\begin{array}{c} 2x+3y+11z=13 \\ x+4y+3z=4 \\ 5x+10y+8z=13 \end{array}$$

Solution. The augmented matrix is given by

$$\left[\begin{array}{ccc|c} 2 & 3 & 11 & 13 \\ 1 & 4 & 3 & 4 \\ 5 & 10 & 8 & 13 \end{array}\right]$$

We apply Gauss-Jordan elimination as in the algorithm above.

1. The first pivot column is the first column of the matrix, as this is the first nonzero column from the left. There is already a non-zero entry in that position. We could leave it that way and either multiply the first row by $\frac{1}{2}$ or go on to the next step and wait until step 4 to convert the 2 to a 1. However, by interchanging the first and second rows, we will get a 1 in this pivot position without introducing fractions. This is an attractive option, so let's do it that way. We apply the elementary row operation $R_1 \leftrightarrow R_2$ yielding

$$\left[\begin{array}{ccc|c} 1 & 4 & 3 & 4 \\ 2 & 3 & 11 & 13 \\ 5 & 10 & 8 & 13 \end{array}\right]$$

2. Step two involves creating zeros in the entries below the first pivot position. To do this, we use the operations $R_2 \to R_2 - 2R_1$ and $R_3 \to R_3 - 5R_1$ to get

$$\left[\begin{array}{ccc|c} 1 & 4 & 3 & 4 \\ 0 & -5 & 5 & 5 \\ 0 & -10 & -7 & -7 \end{array}\right]$$

3. Now ignore the top row and repeat steps 1 and 2 to the augmented matrix made up of the remaining rows $\left[\begin{smallmatrix} 0 & -5 & 5 \\ 0 & -10 & -7 \end{smallmatrix} \middle| \begin{smallmatrix} 5 \\ -7 \end{smallmatrix}\right]$ In this smaller augmented matrix, the second column is a pivot column, and -5 is in the first pivot position. To get a one in the pivot position, we multiply the first row by $-\frac{1}{5}$, yielding $\left[\begin{smallmatrix} 0 & 1 & -1 \\ 0 & -10 & -7 \end{smallmatrix} \middle| \begin{smallmatrix} -1 \\ -7 \end{smallmatrix}\right]$. We then apply step 2 to our smaller augmented matrix, adding 10 times row 1 to row 2, yielding $\left[\begin{smallmatrix} 0 & 1 & -1 \\ 0 & 0 & -17 \end{smallmatrix} \middle| \begin{smallmatrix} -1 \\ -17 \end{smallmatrix}\right]$.

   Returning to the augmented matrix we obtained in 2, we have now done the following operations:

$$\left[\begin{array}{ccc|c} 1 & 4 & 3 & 4 \\ 0 & -5 & 5 & 5 \\ 0 & -10 & -7 & -7 \end{array}\right] \xrightarrow{R_2 \to -\frac{1}{5}R_2} \left[\begin{array}{ccc|c} 1 & 4 & 3 & 4 \\ 0 & 1 & -1 & -1 \\ 0 & -10 & -7 & -7 \end{array}\right] \xrightarrow{R_3 \to R_3 + 10R_2} \left[\begin{array}{ccc|c} 1 & 4 & 3 & 4 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & -17 & -17 \end{array}\right]$$

4. We now have non-zero elements in all our pivot positions now and 0's below them. The pivots in the first two rows are already 1's, we just need to fix the last one using the operation $R_3 \to -\frac{1}{17}R_3$ to get
$$\left[\begin{array}{ccc|c} 1 & 4 & 3 & 4 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

5. The augmented matrix is now in row-echelon form . To get to reduced row-echelon form , we begin with the furthest right column that contains a pivot, in this case the third column, and clear out the entries above it:
$$\left[\begin{array}{ccc|c} 1 & 4 & 3 & 4 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 1 \end{array}\right] \xrightarrow[R_2 \to R_2 + R_3]{R_1 \to R_1 - 3R_3} \left[\begin{array}{ccc|c} 1 & 4 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right]$$
Now move to the next pivot column to the left (the second column) and clear out the entry above the pivot:
$$\xrightarrow{R_1 \to R_1 - 4R_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{array}\right]$$
The augmented matrix is now in reduced row-echelon form .

We can read off the solution:
$$\boxed{x = 1, y = 0, z = 1.}$$

♠

As we will see later, each of the three equations in the example above is the equation of a plane. We have now shown that the three planes intersect in one point $(1,0,1)$.

Example 1.20

Find all points of intersection, if any, of the planes given by the equations $3x + y + 4z = 6$ *and* $2x - y + 6z + 5$.

Solution. We need to solve the system of linear equations
$$\begin{aligned} 3x + y + 4z &= 6 \\ 2x - y + 6z &= 5 \end{aligned}$$
with augmented matrix
$$\left[\begin{array}{ccc|c} 3 & 1 & 4 & 6 \\ 2 & -1 & 6 & 5 \end{array}\right]$$

We have a 3 in the first pivot position. An attractive option is to subtract row 2 from row 1 to get a 1 in the pivot position. After that step we will continue following the Gauss-Jordan algorithm.
$$\left[\begin{array}{ccc|c} 3 & 1 & 4 & 6 \\ 2 & -1 & 6 & 5 \end{array}\right] \xrightarrow{R_1 \to R_1 - R_2} \left[\begin{array}{ccc|c} 1 & 2 & -2 & 1 \\ 2 & -1 & 6 & 5 \end{array}\right] \xrightarrow{R_2 \to R_2 - 2R_1} \left[\begin{array}{ccc|c} 1 & 2 & -2 & 1 \\ 0 & -5 & 10 & 3 \end{array}\right]$$

$$\xrightarrow{R_2 \to -\frac{1}{5}R_2} \left[\begin{array}{ccc|c} 1 & 2 & -2 & 1 \\ 0 & 1 & -2 & -\frac{3}{5} \end{array}\right] \xrightarrow{R_1 \to R_1 - 2R_2} \left[\begin{array}{ccc|c} 1 & 0 & 2 & \frac{11}{5} \\ 0 & 1 & -2 & -\frac{3}{5} \end{array}\right]$$

The matrix is now in reduced row-echelon form. We read off:

$$x \ + 2z = \frac{11}{5} \qquad \text{so} \qquad x = \frac{11}{5} - 2z$$

$$y - 2z = -\frac{3}{5}, \qquad \text{so} \qquad y = -\frac{3}{5} + 2z$$

Here z is a free variable and x and y are determined. Writing $z = t$, we get the solution:

$$\boxed{x = \frac{11}{5} - 2t, \quad y = -\frac{3}{5} + 2t, \quad z = t}$$

where t is any real number.

Aside: As we will see in Section 2.3 of the next chapter, as t varies, these solutions trace out a line. This is consistent with the geometric discussion in Section 1.1; any pair of planes that are not parallel will intersect in a line. ♠

1.2.4. Section Summary

- Augmented matrices give a way of encoding all the information in a system of linear equations. Each row of the matrix corresponds to one of the equations. Each column to the left of the bar corresponds to one of the variables. One can go back and forth between augmented matrices and systems of linear equations.
- If we perform an elementary row operation on an augmented matrix, the system of linear equations corresponding to the new augmented matrix has exactly the same solutions as the original system.
- Gauss-Jordan elimination gives an algorithm for converting the augmented matrix to row-echelon form and further to reduced row-echelon form using elementary row operations.
- In row-echelon form, the first non-zero entry of each row is a one and its location is called a pivot position. The columns containing pivot positions are called pivot columns.
- Using row-echelon form and back substitution, one can obtain the solution to the system of linear equations. One can read off the solution more quickly using reduced row-echelon form.
- Every system of linear equations either has exactly one solution, infinitely many solutions, or no solutions.
- If there are no solutions, a row $\left[\begin{array}{cccc|c} 0 & 0 & \ldots & 0 & 1 \end{array}\right]$ will appear when the augmented matrix is converted to row-echelon form.
- If there is a unique solution, the number of non-zero rows in the augmented matrix in row-echelon form will equal the number of variables. In this case, every column to the left of the bar will be a pivot column.

- If there are infinitely many solutions, then the number of non-zero rows in the augmented matrix in row-echelon form will be less than the number of variables. In this case, the variables corresponding to pivot columns can be viewed as determined variables and the other variables are then free variables. Given any choice of values for the free variables, one can then solve for the determined variables to get a solution to the original system. We use parameters to write down all the solutions at once.

Exercises

Exercise 1.2.1 *For each of the following, indicate whether the augmented matrix is in row-echelon form and whether it is in reduced row-echelon form .*

(a) $\left[\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 1 & 7 \end{array}\right]$

(b) $\left[\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 2 & 7 \end{array}\right]$

(c) $\left[\begin{array}{ccc|c} 1 & 5 & 0 & 8 \\ 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$

(d) $\left[\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & 4 & 5 \\ 0 & 0 & 1 & 2 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{array}\right]$

Exercise 1.2.2 *In each of the following, you are given the reduced row-echelon form of the augmented matrix for a system of linear equations in the variables x, y, z. Determine whether the system of equations is consistent and, if so, write down all solutions. (If there are infinitely many solutions, express the solutions using parameters.)*

(a) $\left[\begin{array}{ccc|c} 1 & 2 & 0 & 5 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 1 \end{array}\right]$

(b) $\left[\begin{array}{ccc|c} 1 & 2 & 0 & 5 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right]$

(c) $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 1 \end{array}\right]$

Exercise 1.2.3 *In each of the following, you are given the reduced row-echelon form of the augmented matrix for a system of linear equations in the four variables x,y,z,w. Determine whether the system of equations is consistent and, if so, write down all solutions. (If there are infinitely many solutions, express the solutions using parameters.)*

(a) $\left[\begin{array}{cccc|c} 1 & 0 & 0 & 8 & 5 \\ 0 & 1 & 0 & 3 & 2 \\ 0 & 0 & 1 & 2 & 4 \end{array}\right]$

(b) $\left[\begin{array}{cccc|c} 1 & 2 & 0 & 7 & 5 \\ 0 & 0 & 1 & 2 & 4 \end{array}\right]$

Exercise 1.2.4 *In each of the following, you are given the row-echelon form of the augmented matrix for a system of linear equations in the variables x,y,z. Determine whether the system of equations is consistent and, if so, use back substitution to find the solution. (If there are infinitely many solutions, express the solutions using parameters.)*

(a) $\left[\begin{array}{ccc|c} 1 & 2 & 0 & 5 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 1 & 1 \end{array}\right]$

(b) $\left[\begin{array}{ccc|c} 1 & 2 & 6 & 5 \\ 0 & 1 & 4 & 4 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$

(c) $\left[\begin{array}{ccc|c} 1 & 2 & 6 & 5 \\ 0 & 1 & 4 & 4 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right]$

Exercise 1.2.5 *Row reduce the following matrix to obtain the row-echelon form. Then continue to obtain the reduced row-echelon form.*

$$\left[\begin{array}{ccc|c} 0 & 0 & -1 & -1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & -1 \end{array}\right]$$

Exercise 1.2.6 *Use augmented matrices and Gauss-Jordan elimination to find the intersection, if any, of the lines $x+3y=1$ and $4x-y=3$.*

Exercise 1.2.7 *Use augmented matrices and Gauss-Jordan elimination to find the intersection, if any, of the two lines*
$2x+y=4$ *and* $3x-2y=-1$

Exercise 1.2.8 *Use augmented matrices and Gauss-Jordan elimination to determine whether the three lines, $x+2y=1, 2x-y=1$, and $4x+3y=5$ have a common point of intersection. If so, find the point of intersection.*

Exercise 1.2.9 *Use augmented matrices and Gauss-Jordan elimination to determine whether the three lines,* $x+2y=1, 2x-y=1$, *and* $4x+3y=3$ *have a common point of intersection. If so, find the point of intersection.*

Exercise 1.2.10 *Use augmented matrices and Gauss-Jordan elimination to determine whether the three lines,* $x+2y=4, 2x-y=0$, *and* $x+3y=6$ *have a common point of intersection. If so, find the point of intersection.*

Exercise 1.2.11 *Use augmented matrices and Gauss-Jordan elimination to determine whether the planes,* $x+y-3z=2$ *and* $2x+y+z=1$ *intersect. If so, find the intersection.*

Exercise 1.2.12 *Use augmented matrices and Gauss-Jordan elimination to determine whether the three planes,* $x+y-3z=2$, $2x+y+z=1$, *and* $3x+2y-2z=0$ *have a common point of intersection. If so, find the intersection.*

Exercise 1.2.13 *For each of the following systems of linear equations, use augmented matrices and Gauss-Jordan elimination to determine whether the system of equations is consistent and, if so, to find all solutions. If there are infinitely many solutions, express them using parameters.*

(a) $2x+y=4$ *and* $3x-2y=-1$

(b) $3x-6y-7z=-8$, $x-2y-2z=-2$, *and* $x-2y-3z=-4$.

(c) $3+6y+z=17$ *and* $2x+4y+z=13$

(d) $x+2y=4$ $3x+6y+z=17$ *and* $2x+4y+z=13$

(e) $x+2y=4$ $3x+6y+z=17$ *and* $2x+4y+z=16$

(f) $9x-2y+4z=-17$, $13x-3y+6z=-25$, *and* $-2x-z=3$.

(g) $7x+14y+15z=22$, $2x+4y+3z=5$, *and* $3x+6y+10z=13$.

(h) $3x-y+4z=6$, $y+8z=0$, *and* $-2x+y=-4$.

(i) $9x-2y+4z=-17$, $13x-3y+6z=-25$, *and* $-2x-z=3$.

(j) $8x+2y+3z=-3$, $8x+3y+3z=-1$, *and* $4x+y+3z=-9$. *(Suggestion: See Remark 1.18.)*

(k) $3x-y-2z=3$, $y-4z=0$, *and* $-2x+y=-2$.

(l) $x+2y+3z+4w=5$, $2x+4y+7z+10w=12$, *and* $3x+6y+10z+14w=17$.

(m) $x+2y+3z+4w=5$, $2x+4y+7z+10w=12$, *and* $3x+6y+10z+14w=18$.

Exercise 1.2.14 *Suppose a system of equations has fewer equations than variables. Will such a system necessarily be consistent? If so, explain why and if not, give an example which is not consistent.*

Exercise 1.2.15 *If a system of equations has more equations than variables, can it have a solution? If so, give an example and if not, tell why not.*

2. Vectors: a Linear Viewpoint

Prerequisite 2.1

Before reading this chapter, you need to be familiar with vector addition, including the parallelogram law, and also multiplication of vectors by scalars.

Notational Conventions 2.2

- *Boldface letters such as* $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x}$ *will always stand for vectors and non-boldface letters such as* a, b, c, t, x, y, z *will stand for scalars (real numbers) or variables whose values are real numbers.*
- $\mathbb{R}^2$ *denotes 2-dimensional space, what you normally think of as the* xy*-plane.* $\mathbb{R}^3$ *denotes 3-dimensional space.*
- *When working in* $\mathbb{R}^3$, *we write* $\mathbf{i} = \langle 1,0,0 \rangle$, $\mathbf{j} = \langle 0,1,0 \rangle$ *and* $\mathbf{k} = \langle 0,0,1 \rangle$ *as in Stewart.*
- *When working in* $\mathbb{R}^2$ *rather than* $\mathbb{R}^3$, *we will write* $\mathbf{i} = \langle 1,0 \rangle$ *and* $\mathbf{j} = \langle 0,1 \rangle$.

2.1 Linear Combinations and Spans

Motivation

If there were no obstacles such as buildings, trees or ponds in your way, you could get anywhere in Hanover by first walking directly east or west and then walking north or south. You can convince yourself that you could also get wherever you wanted by walking due northeast/southwest and then walking due east/west. Directions can be specified by vectors. If you are constrained to move only in certain specified directions, what places can you reach? This question underlies the notions of linear combinations and spans that we will introduce in this section.

Outcomes

- *Understand linear combinations and their geometric interpretation.*
- *Be able to check whether a given vector is a linear combination of other specified vectors.*
- *Understand the concepts of span and vector span of a collection of vectors.*
- *Determine whether the span of a collection of a vectors is a line, a plane, or* $\mathbb{R}^3$.

2.1.1. Linear combinations of vectors

Question: Given a finite collection of vectors, what points can you reach if you start at the origin and are constrained to move only in directions specified by these vectors?

As a warm-up, let's first consider a single non-zero vector $\mathbf{v}$. If you start from the origin and walk arbitrarily far forwards or backwards in the direction of the vector $\mathbf{v}$, you will trace out a line parallel to $\mathbf{v}$. For example, the vector $\mathbf{v} = \langle 2, 1 \rangle$ and the line ℓ through the origin parallel to $\mathbf{v}$ are illustrated in Figure 2.1. Starting from the origin and moving parallel to $\mathbf{v}$ we can reach any point on this line but can never leave the line. Observe that the position vector of every point on ℓ is a scalar multiple $t\mathbf{v}$ of $\mathbf{v}$. Similarly, if $\mathbf{w}$ is any non-zero vector in $\mathbb{R}^3$, then walking arbitrarily far forwards or backwards from the origin in the direction $\mathbf{w}$, you will trace out a line in $\mathbb{R}^3$. Again the line consists precisely of those points in $\mathbb{R}^3$ whose position vectors are multiples $t\mathbf{w}$ of $\mathbf{w}$.

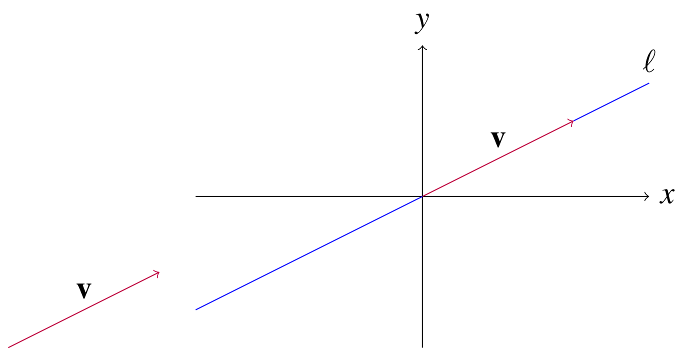


Figure 2.1

Next let's consider a pair of vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^3$ (or $\mathbb{R}^2$). Starting from the origin, suppose we are allowed only to walk along paths made up of line segments parallel to $\mathbf{v}$ and $\mathbf{w}$. In Figure 2.2, we have drawn a pair of vectors $\mathbf{v}$ and $\mathbf{w}$ and illustrated two such paths from the origin. We have labelled the edges of the path by their displacement vectors. E.g., in the first path, the displacement vector $\mathbf{P_1P}$ is the vector $-\mathbf{w}$. Observe that the position vector of our final point P in the first path is $\mathbf{OP} = 2\mathbf{v} - \mathbf{w}$, while the position vector of the final point Q on the second path is given by $\mathbf{OQ} = -3\mathbf{w} + 2\mathbf{v} + 2\mathbf{w} + \mathbf{v}$, which simplifies to $3\mathbf{v} - \mathbf{w}$.

The examples above illustrate the following proposition:

Proposition 2.3

- *Starting from the origin and following paths made up of line segments parallel to two given vectors $\mathbf{v}$ and $\mathbf{w}$, we can reach precisely those points P whose position vectors satisfy $\mathbf{OP} = s\mathbf{v} + t\mathbf{w}$, where s and t are scalars.*

- *Similarly, starting from the origin and following paths made up of line segments parallel to three given vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$, we can reach precisely those points P whose position vectors satisfy $\mathbf{OP} = a\mathbf{u} + b\mathbf{v} + c\mathbf{w}$, where a, b and c are scalars.*

One can make similar statements using any number of vectors. Proposition 2.3 motivates the following definition:

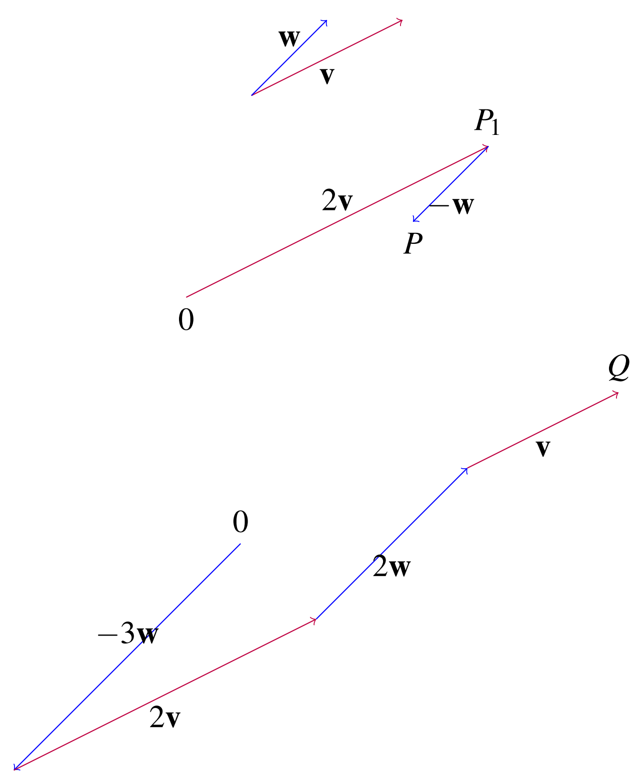


Figure 2.2

Definition 2.4: Linear Combinations of Vectors

(i) Let $\mathbf{v}$ *and* $\mathbf{w}$ *be vectors. Any vector of the form* $s\mathbf{v}+t\mathbf{w}$*, where* s *and* t *are scalars, is called a linear combination of* $\mathbf{v}$ *and* $\mathbf{w}$*. The scalars* s *and* t *are called the coefficients of* $\mathbf{v}$ *and* $\mathbf{w}$ *in this linear combination.*

(ii) We can similarly define linear combinations of any finite collection of vectors. E.g., if $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *are vectors and* a, b *and* c *are scalars, then* $a\mathbf{u}+b\mathbf{v}+c\mathbf{w}$ *is a linear combination of* $\mathbf{u}$, $\mathbf{v}$ *and* $\mathbf{w}$*. Again the scalars* a,b,c *are called the coefficients.*

(iii) We even say that $t\mathbf{v}$ *is a linear combination of the vector* $\mathbf{v}$*. (This sounds strange but it's convenient to be able to use the same language for any number of vectors, including a single vector.)*

We emphasize that the coefficients can be any scalars, including 0. Thus the zero vector is always a linear combination of any given collection of vectors. Similarly, $\mathbf{v}$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$ since $\mathbf{v} = 1\mathbf{v} + 0\mathbf{w}$.

Example 2.5

When we write $\langle 2,-1,4\rangle = 2\mathbf{i}-\mathbf{j}+4\mathbf{k}$*, we are expressing the vector* $\langle 2,-1,4\rangle$ *as a linear combination of the standard basis vectors* $\mathbf{i}, \mathbf{j}$ *and* $\mathbf{k}$*. More generally, any vector* $\langle x,y,z\rangle$ *is a linear combination of* $\mathbf{i}$, $\mathbf{j}$ *and* $\mathbf{k}$ *since* $\langle x,y,z\rangle = x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$.

Example 2.6

Let $\mathbf{v} = \langle 2,1 \rangle$ *and* $\mathbf{w} = \langle 1,3 \rangle$. *Is* $\langle 10,15 \rangle$ *a linear combination of* $\mathbf{v}$ *and* $\mathbf{w}$*?*

Solution. We are asking whether there exist scalars s and t such that

$$\langle 10,15 \rangle = s\langle 2,1 \rangle + t\langle 1,3 \rangle$$

Matching components, we obtain two equations:

$$2s + t = 10$$

$$s + 3t = 15$$

Solving this system of equations, we get $s = 3, t = 4$. So

$$\langle 10,15 \rangle = 3\langle 2,1 \rangle + 4\langle 1,3 \rangle$$

and the answer is yes!

♠

Example 2.7

Determine whether the following vectors are linear combinations of $\mathbf{v} = \langle 1,0,1 \rangle$ *and* $\mathbf{w} = \langle 1,2,3 \rangle$.
(i) $\langle 3,4,8 \rangle$
(ii) $\langle 3,4,7 \rangle$

Solution.
(i) We are asking whether there exist scalars s,t such that

$$\langle 3,4,8 \rangle = s\langle 1,0,1 \rangle + t\langle 1,2,3 \rangle$$

has a solution. Matching components, this gives a system of three equations in the two variables s,t:

$$s + t = 3$$

$$0s + 2t = 4$$

$$s + 3t = 8$$

Method 1: Solve the first two of these equations simultaneously and check whether the solution also satisfies the third equation. The second equation says that $t = 2$. Substituting into the first equation we must have $s = 1$. But if we put $s = 1$ and $t = 2$ into the third equation, we get $7 = 8$, which is clearly a contradiction! So $\langle 3,4,8 \rangle$ is *not* a linear combination of $\mathbf{v}$ and $\mathbf{w}$.

Method 2: We can check whether the system of equations has a solution by using row reduction as in chapter I. The augmented matrix for this system is given by

$$\left[\begin{array}{cc|c} 1 & 1 & 3 \\ 0 & 2 & 4 \\ 1 & 3 & 8 \end{array}\right] \tag{2.1}$$

After row reducing to reduced echelon form, we obtain:

$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{array}\right] \tag{2.2}$$

The last row shows us that there are no solutions.

Conclusion: $\langle 3,4,8\rangle$ is not a linear combination of $\mathbf{v}$ and $\mathbf{w}$.

(ii) If we try the same process for the vector $\langle 3,4,7\rangle$, the only difference is that the third equation above becomes $s+3t=7$, while the first two equations remain the same. Using method 1, we see that the solution $s=1, t=2$ to the first two equations does satisfy the third equation as well. Thus we get a solution: $s=1, t=2$. So $\langle 3,4,7\rangle$ is the linear combination

$$\langle 3,4,7\rangle = \mathbf{v}+2\mathbf{w}$$

of $\mathbf{v}$ and $\mathbf{w}$. If you use method 2 instead, you will find after row reduction that you get the matrix

$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right] \tag{2.3}$$

♠

which again gives you the solution $s=1, t=2$.

Remark 2.8: Solving systems of 3 equations in 2 variables.

Example 2.7 illustrates two ways to determine whether a system of three linear equations in two variables has a solution: The first way is to solve two of the equations simultaneously and then test whether your solution – if there is one – satisfies the third equation.
A second way is to use row reduction as in the previous chapter.

2.1.2. Span of a set of vectors

We define the span of a set of vectors geometrically as the collection of all points that you can reach from the origin if you are constrained to walk only in the directions of the specified vectors. More precisely:

Definition 2.9: Span of a set of a vectors

Let $A=\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$ be a finite collection of vectors in $\mathbb{R}^3$ (or $\mathbb{R}^2$). The span of A, denoted $\mathrm{Span}(A)$ or $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k)$, is defined as follows:

- *If $A=\{0\}$, then $\mathrm{Span}(A)=\{0\}$. (We will not discuss this uninteresting case further.)*
- *Otherwise, $\mathrm{Span}(A)$ is the set of all points in $\mathbb{R}^3$ (respectively, $\mathbb{R}^2$) that can be connected to the origin by a path made up of line segments each of which is parallel to a non-zero vector in A.*

Proposition 2.3 tells us that we can reach precisely those points whose position vectors are linear combinations of the vectors in A. This motivates the definition of vector span:

Definition 2.10: Vector span of a set of vectors

Let $A = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a finite collection of vectors. The vector span of A, which we will denote by $V\,\mathrm{Span}(A)$ or $V\,\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$, is the collection of all vectors that are linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_k$.

As a corollary of Proposition 2.3, we have:

Corollary 2.11

A point P is in $\mathrm{Span}(A)$ if and only if its position vector $\mathbf{OP}$ is in $V\,\mathrm{Span}(A)$.

Example 2.12

Let $A = \{\mathbf{v}, \mathbf{w}\}$ where $\mathbf{v} = \langle 1, 0, 1\rangle$ and $\mathbf{w} = \langle 1, 2, 3\rangle$. In Example 2.7, we saw that the vector $\langle 3, 4, 7\rangle$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$ but that $\langle 3, 4, 8\rangle$ is not. Thus $\langle 3, 4, 7\rangle$ belongs to $V\,\mathrm{Span}(A)$ but $\langle 3, 4, 8\rangle$ does not.

It then follows from Corollary 2.11 that the point $(3, 4, 7)$ belongs to $\mathrm{Span}(A)$ but the point $(3, 4, 8)$ does not.

Question: What types of geometric objects might we get by taking the spans of finite sets of vectors?

As we have seen, if we can move only in one direction, then the points we can access form a line. Thus:

Example 2.13: The span of a single vector

Let A consist of a single vector non-zero vector $\mathbf{v}$. Then $V\,\mathrm{Span}(\mathbf{v})$ consists of all vectors of the form $t\mathbf{v}$ with t any scalar, and

$$\mathrm{Span}(\mathbf{v}) \text{ is the line through the origin parallel to } \mathbf{v}.$$

For a specific example, if $\mathbf{v} = \langle 2, 1\rangle$, then $\mathrm{Span}(\mathbf{v})$ is the line in Figure 2.1. From the vector $\mathbf{v}$, we can see that the slope of this line is $\frac{1}{2}$ and thus the Cartesian equation of the line is given by $y = \frac{x}{2}$.

Example 2.14

- *Let* $A = \{\mathbf{i},\mathbf{j},\mathbf{k}\}$*.Since every vector in* $\mathbb{R}^3$ *is a linear combination* $x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$ *of the standard basis vectors and thus lies in* $V\,\mathrm{Span}(\mathbf{i},\mathbf{j},\mathbf{k})$, *Corollary 2.11 tells us that*

$$\mathrm{Span}(\mathbf{i},\mathbf{j},\mathbf{k}) = \mathbb{R}^3.$$

- *Let* $A = \{\mathbf{i},\mathbf{k}\}$*. Then* $V\,\mathrm{Span}(\mathbf{i},\mathbf{k})$ *consists of all vectors* $s\mathbf{i}+t\mathbf{k} = \langle s,0,t\rangle$ *with* s,t *arbitrary scalars and*

$$\mathrm{Span}(\mathbf{i},\mathbf{k}) \text{ is the } xz\text{-plane.}$$

- *Working in* $\mathbb{R}^2$ *rather than* $\mathbb{R}^3$*, we similarly see that* $\mathrm{Span}(\mathbf{i},\mathbf{j}) = \mathbb{R}^2$ *where* $\mathbf{i}$ *and* $\mathbf{j}$ *are the standard basis vectors of* $\mathbb{R}^2$

Thus far, we've seen examples of spans that are lines, planes, and $\mathbb{R}^3$. As we will see below, these are all the possibilities (assuming $A \neq \{0\}$).

Let's consider the span of a pair of non-parallel vectors in $\mathbb{R}^2$. As a warm-up, Figure 2.3 is a picture of $\mathbb{R}^2$ with dots at all points (m,n) where m and n are integers. We have labeled these points by their position vectors $m\mathbf{i}+n\mathbf{j}$. The dots form the corners of squares that tile the plane and give us families of horizontal and vertical lines. As usual, once you have all the integer points marked, it's then easy to find the approximate location of any other point. Figure 2.3 looks like the road map of a city in which all the streets are perfectly straight and each goes either north/south or east/west. The fact that $\mathrm{Span}(\mathbf{i},\mathbf{j}) = \mathbb{R}^2$ is of course the familiar fact that you can reach any point by first going east/west and then north/south.

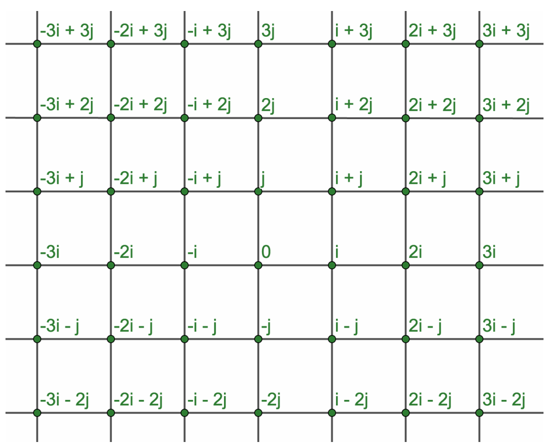


Figure 2.3

Now let's consider the span of another pair of vectors, say

$$\mathbf{v} = \langle 2,1\rangle \text{ and } \mathbf{w} = \langle 1,3\rangle.$$

In Figure 2.4, we have drawn a picture analogous to that in Figure 2.3 but using the vectors $\mathbf{v}$ and $\mathbf{w}$ in place of $\mathbf{i}$ and $\mathbf{j}$. The dots now are at all the points with position vectors $m\mathbf{v}+n\mathbf{w}$ where m and n are integers. The parallelogram law for addition of vectors enables us to tile the plane with parallelograms whose corners are these dots as shown in the picture.

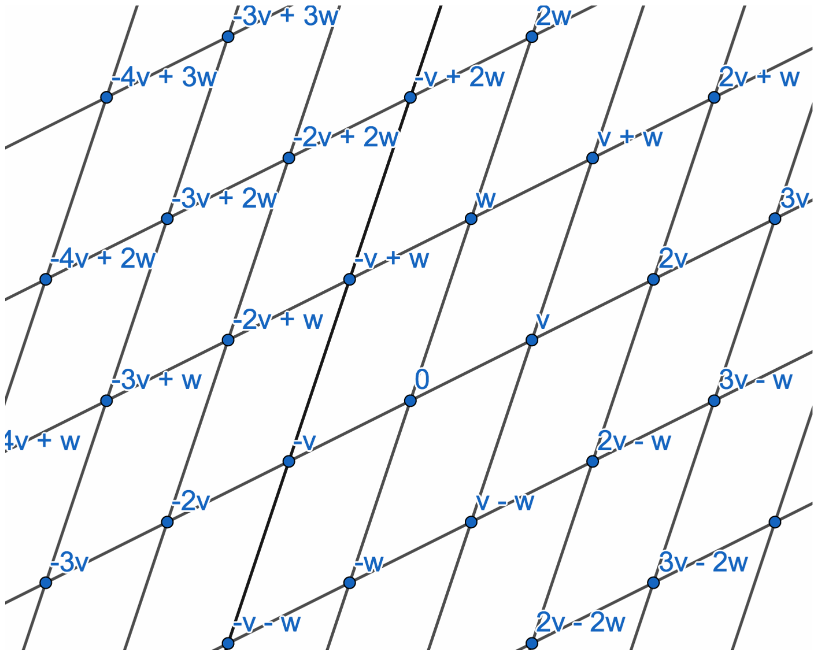


Figure 2.4

We again have two families of parallel lines. You might think of Figure 2.4 as a grid of streets in a city, where now the streets no longer intersect at right angles. Let's temporarily refer to the lines in the grid parallel to **v** and **w** as **v** streets and **w** streets.

In Figure 2.5, we have randomly chosen a point P in the plane $\mathbb{R}^2$.

Question: Can we get to P from the origin 0 by walking only parallel to the "city streets"?

Answer: Yes! We have shown such a path in the drawing. We first walk "downhill" (i.e. towards the left) along the **v** street from 0 until we find ourselves at a point Q such that **QP** is parallel to **w**. The point Q has position vector $-2.7\mathbf{v}$. (The value -2.7 is an estimate here.) We now head up to P along the segment QP, thus achieving our goal. In vector language, the displacement vector vector **QP** is given by $1.6\mathbf{w}$. (Again the 1.6 is an estimate based on the picture.) By the parallelogram law, we see that the position vector **OP** of P satisfies

$$\mathbf{OP} = -2.7\mathbf{v} + 1.6\mathbf{w}.$$

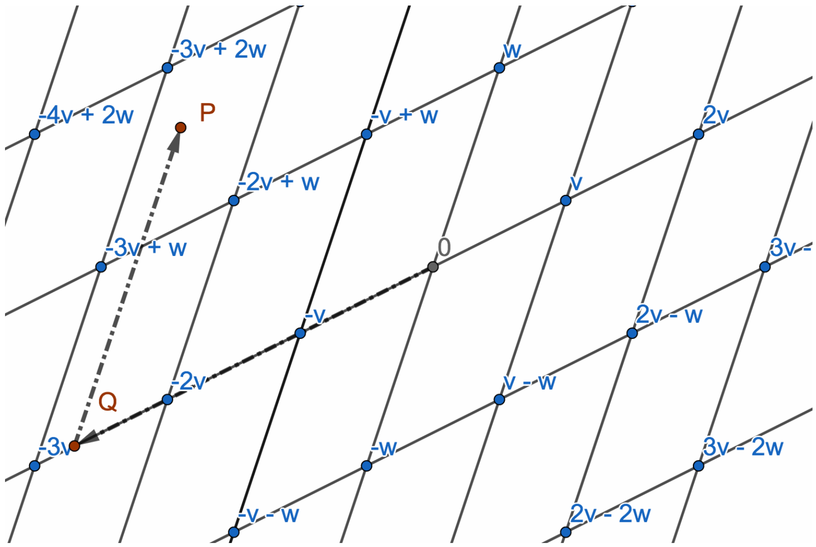


Figure 2.5

Try to convince yourself now that you can get to *every* point in the plane by walking only parallel to the "city streets", i.e., parallel to $\mathbf{v}$ and $\mathbf{w}$. Thus $\mathrm{Span}(\mathbf{v}, \mathbf{w}) = \mathbb{R}^2$.

More generally, we have:

Theorem 2.15: Span of any collection of vectors in $\mathbb{R}^2$

Let A be a finite set of vectors in $\mathbb{R}^2$.

1. *If A consists only of the zero vector, then* $\mathrm{Span}(A) = \{0\}$.
2. *If all the vectors in A are parallel to each other (and at least one of them is non-zero), then* $\mathrm{Span}(A)$ *is the line through the origin parallel to the vectors in A.*
3. *If A contains at least two vectors that aren't parallel to each other, then* $\mathrm{Span}(A) = \mathbb{R}^2$.

Thus the span of any finite collection of vectors in $\mathbb{R}^2$ is either the origin, a line through the origin, or all of $\mathbb{R}^2$.

To see (2), note that if all the vectors in A are parallel and you are constrained to move only parallel to vectors in A, then you have only one direction in which to move.

The most interesting part of the theorem is item (3), which tells us that the directions given by any pair of non-parallel vectors are enough to enable us to reach every point in the plane $\mathbb{R}^2$.

Definition 2.16

- *The origin, lines through the origin, and $\mathbb{R}^2$ are called subspaces of $\mathbb{R}^2$.*
- *Similarly, the origin, lines through the origin, planes through the origin, and $\mathbb{R}^3$ are called subspaces of $\mathbb{R}^3$.*
- *The dimension of a subspace is 0 for the origin, 1 for a line, 2 for a plane and 3 for $\mathbb{R}^3$.*

Theorem 2.15 says that the span of any finite collection of vectors in $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$. We next give an analogous theorem for $\mathbb{R}^3$. We first need a definition.

Definition 2.17

A collection of vectors in $\mathbb{R}^3$ is said to be coplanar if they all lie in a single plane.

Note: When we say that the vectors all "lie" in a single plane, we really mean that they are all parallel to a single plane. Recall that vectors have length and direction but not position.

Note. Any set consisting of only 2 vectors must be coplanar, since you can always find a plane parallel to two vectors.

Theorem 2.18: Span of any collection of vectors in $\mathbb{R}^3$

Let A be any finite set of vectors in $\mathbb{R}^3$. Then $\operatorname{Span}(A)$ is a subspace of $\mathbb{R}^3$ given as follows:

1. *If A consists only of the zero vector, then $\operatorname{Span}(A) = \{0\}$.*
2. *If all the vectors in A are parallel to each other (and at least one of them is non-zero), then $\operatorname{Span}(A)$ is the line through the origin parallel to the vectors in A.*
3. *If the vectors in A are coplanar and not all of them are parallel, then $\operatorname{Span}(A)$ is the plane through the origin parallel to the vectors in A.*
4. *If the vectors in A are not coplanar, then $\operatorname{Span}(A) = \mathbb{R}^3$.*

Remark 2.19

You may have learned in high school geometry that given two lines ℓ_1 and ℓ_2 that intersect in a point, there is a unique plane $\mathscr{P}$ containing ℓ_1 and ℓ_2. Now suppose instead that you are given two non-parallel vectors $\mathbf{v}$ and $\mathbf{w}$. Then $\operatorname{Span}(\mathbf{v})$ and $\operatorname{Span}(\mathbf{w})$ are lines through the origin and thus there is a unique plane through the origin containing these lines. This is precisely the plane $\operatorname{Span}(\mathbf{v},\mathbf{w})$ and is the unique plane through the origin parallel to $\mathbf{v}$ and $\mathbf{w}$.

Example 2.20

Let $A = \{\mathbf{u},\mathbf{k}\}$ where $\mathbf{u} = \frac{1}{2}\mathbf{i} + \frac{1}{2}\mathbf{j} = \langle \frac{1}{2}, \frac{1}{2}, 0\rangle$ and $\mathbf{k} = \langle 0,0,1\rangle$. The vectors are not parallel but must be coplanar (since there are only two of them). Thus $\operatorname{Span}(A)$ is the plane $\mathscr{P}$ through the origin parallel to these vectors. This plane is illustrated in Figure 2.6. As we did in Figures 2.3 and 2.4, we have put dots at the points that correspond to all linear combinations $m\mathbf{u} + n\mathbf{k}$ where m and n are integers. Convince yourself that the position vector of every point on this plane is a linear combination $s\mathbf{u} + t\mathbf{k}$ (as guaranteed by the theorem).

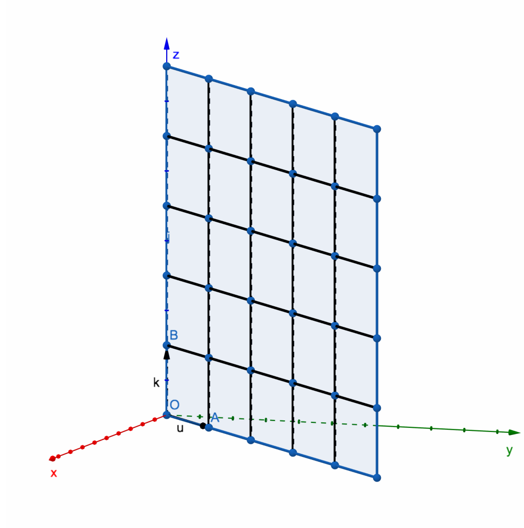


Figure 2.6

Item (4) in Theorem 2.18 tells us that we can reach any point in $\mathbb{R}^3$ by paths consisting of segments parallel to just 3 vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$, provided that the 3 given vectors are not coplanar. The three vectors form the edges of a parallelopiped in $\mathbb{R}^3$ as in Figure 2.7. We denote this parallelopiped by $\mathrm{Par}(\mathbf{u}, \mathbf{v}, \mathbf{w})$. In analogy with what we did in Figures 2.3 and 2.4 in dimension 2, we can tile $\mathbb{R}^3$ by copies of this parallelopiped. The corners of the parallelopipeds in the tiling have position vectors $m\mathbf{u} + n\mathbf{v} + p\mathbf{w}$, where m, n and p are integers. The position vector of every point in the parallelopiped shown in Figure 2.7 are of the form $a\mathbf{u} + b\mathbf{v} + c\mathbf{w}$, with $0 \le a, b, c \le 1$.

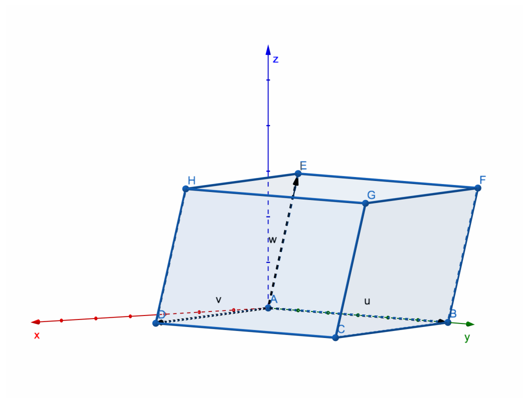


Figure 2.7

Proposition 2.21: Checking whether vectors are coplanar

Let A be a set of vectors containing at least two vectors that are not parallel to each other. Pick out two such vectors $\mathbf{v}$ and $\mathbf{w}$. The set A is coplanar if and only if all the remaining vectors are linear combinations of $\mathbf{v}$ and $\mathbf{w}$.

Proof. Since $\mathbf{v}$ and $\mathbf{w}$ are not parallel but are coplanar, we have that $\mathrm{Span}(\mathbf{v}, \mathbf{w})$ is a plane $\mathscr{P}$ through the origin. To see whether A is coplanar, we need to check whether the remaining vectors are parallel to $\mathscr{P}$.

Now note that any vector $\mathbf{u}$ parallel to $\mathscr{P}$ is the position vector of a point on $\mathscr{P}$. (To see this, position $\mathbf{u}$ at the origin. Since the origin is in $\mathscr{P}$ and $\mathbf{u}$ is parallel to $\mathscr{P}$, its tip lies in $\mathscr{P}$ as well.) By Corollary 2.11, it follows that a vector is parallel to $\mathscr{P}$ if and only if it is a linear combination of $\mathbf{v}$ and $\mathbf{w}$. The proposition follows. ♠

Determining whether $\text{Span}(A)$ *is a line, a plane, or* $\mathbb{R}^3$

(We are assuming here that $A \neq \{0\}$.)

1. Check by inspection whether all the vectors in A are parallel. If so, you know $\text{Span}(A)$ is a line through the origin. If not, continue to next step.

2. Use Proposition 2.21 to determine whether A is coplanar. If so, then $\text{Span}(A)$ is a plane through the origin. If not, $\text{Span}(A) = \mathbb{R}^3$.

Example 2.22: Determining the type of span

- *Let* $A = \{\langle 3,4,7\rangle, \langle 1,0,1\rangle, \langle 1,2,3\rangle\}$. *The vectors in* A *are not all parallel, so* $\text{Span}(A)$ *is at least 2-dimensional. To check whether the vectors are coplanar, let's check whether the first vector is a linear combination of the last two. We already did this computation in Example 2.7 and found that the answer is yes. Thus the vectors are coplanar and the span is a plane through the origin.*

- *Let* $A = \{\langle 3,4,8\rangle, \langle 1,0,1\rangle, \langle 1,2,3\rangle\}$. *Again the vectors are not parallel. In Example 2.7, we saw that the first vector is* not *a linear combination of the other two vectors. Thus the vectors are not coplanar and* $\text{Span}(A) = \mathbb{R}^3$.

2.1.3. Section Summary

Let A be a finite set of vectors.

- You can check whether a vector is a linear combination of the vectors in A by solving a system of linear equations.

- $\text{Span}(A)$ consists of all *points* that can be reached from the origin by moving only in the directions specified by the vectors in A.

- $V\,\text{Span}(A)$ consists of all *vectors* that are linear combinations of the vectors in A.

- A point P is in $\text{Span}(A)$ if and only if the position vector $\mathbf{OP}$ is in $V\,\text{Span}(A)$.

- $\text{Span}(A)$ is a line through the origin if all the vectors in A are parallel (and not all zero).

- $\text{Span}(A)$ is a plane through the origin if the vectors in A are coplanar but not all parallel.

- $\text{Span}(A) = \mathbb{R}^3$ if A is not coplanar.

- To check whether A is coplanar, pick out two vectors in A that aren't parallel and check whether the other vectors are linear combinations of these two.

Exercises

Exercise 2.1.1 *Let* $\mathbf{u} = \langle 1,2,4\rangle$, $\mathbf{v} = \langle 3,1,5\rangle$ *and* $\mathbf{w} = \langle 2,4,0\rangle$. *Evaluate the linear combination* $2\mathbf{u} - \mathbf{v} + 4\mathbf{w}$.

Exercise 2.1.2 *Draw pictures analogous to those in Figure 2.2 to illustrate the following linear combinations:*

(a) $2\langle 1,1\rangle - \langle 3,5\rangle$

(b) $-2\langle 1,4\rangle + 3\langle 1,5\rangle$

Exercise 2.1.3 *For each of the following, you are given a pair of vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *in* $\mathbb{R}^2$ *and a third vector* $\mathbf{u}$. *Either express* $\mathbf{u}$ *explicitly as a linear combination* $s\mathbf{v} + t\mathbf{w}$ *by finding the coefficients* s *and* t *or else show that* $\mathbf{u}$ *is not a linear combination of* $\mathbf{v}$ *and* $\mathbf{w}$.

(a) $\mathbf{v} = \langle 1,1\rangle$, $\mathbf{w} = \langle 3,5\rangle$, $\mathbf{u} = \langle -1,-3\rangle$

(b) $\mathbf{v} = \langle 1,4\rangle$, $\mathbf{w} = \langle -2,-8\rangle$, $\mathbf{u} = \langle -1,-3\rangle$

(c) $\mathbf{v} = \langle 1,4\rangle$, $\mathbf{w} = \langle 1,5\rangle$, $\mathbf{u} = \langle 2.5,12\rangle$

(d) $\mathbf{v} = \langle 1,1,1\rangle$, $\mathbf{w} = \langle 2,1,4\rangle$, $\mathbf{u} = \langle 5,4,7\rangle$

(e) $\mathbf{v} = \langle 1,1,1\rangle$, $\mathbf{w} = \langle 2,1,4\rangle$, $\mathbf{u} = \langle 0,1,3\rangle$

Exercise 2.1.4 *For each of the following, you are given a pair of vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *in* $\mathbb{R}^3$ *and a third vector* $\mathbf{u}$. *Determine whether* $\mathbf{u}$ *lies in* $V\,\mathrm{Span}(\mathbf{v},\mathbf{w})$.

(a) $\mathbf{v} = \langle 1,2,4\rangle$, $\mathbf{w} = \langle 2,0,3\rangle$, $\mathbf{u} = \langle 4,4,8\rangle$

(b) $\mathbf{v} = \langle 1,2,4\rangle$, $\mathbf{w} = \langle 2,0,3\rangle$, $\mathbf{u} = \langle 7,6,18\rangle$

(c) $\mathbf{v} = \langle 1,4,1\rangle$, $\mathbf{w} = \langle -2,-8,0\rangle$, $\mathbf{u} = \langle 5,20,3\rangle$

(d) $\mathbf{v} = \langle 1,4,1\rangle$, $\mathbf{w} = \langle -2,-8,0\rangle$, $\mathbf{u} = \langle -3,-12,-3\rangle$

Exercise 2.1.5

a) Check whether $\langle 5,6,10\rangle$ *lies in* $V\,\mathrm{Span}(\langle 1,1,2\rangle, \langle 2,3,1\rangle)$.

b) Find all values of z *so that the vector* $\langle 5,6,z\rangle$ *lies in* $V\,\mathrm{Span}(\langle 1,1,2\rangle, \langle 2,3,1\rangle)$. *(Before computing, look at your work in part (a) so that you don't repeat computations you've already done.)*

Exercise 2.1.6 *For each of the following, find the Cartesian equation of the line* $\mathrm{Span}(\mathbf{v})$. *(See Example 2.13.)*

(a) $\mathbf{v} = \langle 5,6\rangle$

(b) $\mathbf{v} = \langle 0,4 \rangle$

(c) $\mathbf{v} = \langle 4,0 \rangle$

(d) $\mathbf{v} = \langle 1,-7 \rangle$

Exercise 2.1.7 *For each of the following pairs of vectors, draw a parallelogram tiling analogous to that in Figure 2.4. Please include all the points $m\mathbf{v}+n\mathbf{w}$ where m and n are integers with $-2 \leq m,n \leq 3$. Label at least ten of these points, including $0, \mathbf{v}, \mathbf{w}$. Your labelled points should include ones with both positive and negative values for each of m and n. (You don't have to label every point, as that becomes tedious!)*

(a) $\mathbf{v} = \langle 1,1 \rangle$, $\mathbf{w} = \langle -1,2 \rangle$

(b) $\mathbf{v} = \langle 1,-1 \rangle$, $\mathbf{w} = \langle -1,2 \rangle$

(c) $\mathbf{v} = \langle -1,-1 \rangle$, $\mathbf{w} = \langle 1,-2 \rangle$

Exercise 2.1.8 *For each of the pairs $\mathbf{v}, \mathbf{w}$ in Exercise 2.1.7, express the vector $\langle 0,4 \rangle$ as a linear combination of $\mathbf{v}$ and $\mathbf{w}$. Illustrate your answer by drawing a path in your tiling analogous to that in Figure 2.5. (You can use the picture from Exercise 2.1.7; you do not have to redraw the tiling.)*

Exercise 2.1.9 *Determine whether each of the following triples of vectors in $\mathbb{R}^3$ is coplanar.*

(a) $\langle 1,1,2 \rangle$, $\langle 2,0,1 \rangle$, $\langle -4,2,1 \rangle$

(b) $\langle 3,1,2 \rangle$, $\langle -1,2,1 \rangle$, $\langle 1,5,4 \rangle$

(c) $\langle 3,1,2 \rangle$, $\langle -1,2,1 \rangle$, $\langle 2,-4,-2 \rangle$

(d) $\langle 3,1,2 \rangle$, $\langle -1,2,1 \rangle$, $\langle 1,4,4 \rangle$

Exercise 2.1.10 *For each of the following, you are given a set A of vectors in $\mathbb{R}^3$. Determine whether* $\operatorname{Span}(A)$ *is a line through the origin, a plane through the origin, or $\mathbb{R}^3$.*

(a) $A = \{\langle 1,1,2 \rangle, \langle 2,0,1 \rangle, \langle -4,-4,-8 \rangle\}$

(b) $A = \{\langle 3,1,2 \rangle, \langle 0,0,0 \rangle, \langle 1,5,4 \rangle\}$

(c) $A = \{\langle 1,1,1 \rangle, \langle 2,0,1 \rangle, \langle 1,-1,1 \rangle\}$

(d) $A = \{\langle 1,1,1 \rangle, \langle 2,0,1 \rangle, \langle 1,-1,0 \rangle\}$

(e) $A = \{\langle 1,1,1 \rangle, \langle 2,0,1 \rangle, \langle 2,2,2 \rangle \langle -6,0,-3 \rangle\}$

Exercise 2.1.11 *True or false:* $\operatorname{Span}(\langle 1,2,1 \rangle) = \operatorname{Span}(\langle 1,2,1 \rangle, \langle 3,6,3 \rangle)$*? Explain*

Exercise 2.1.12 *True or false:* $\operatorname{Span}(\langle 1,2 \rangle, \langle 1,1 \rangle) = \operatorname{Span}(\langle 17,2 \rangle, \langle 0,1 \rangle)$*? Explain.*

Exercise 2.1.13 *Figure 2.8 illustrates the point P whose position vector is the sum $\langle 2,1 \rangle + \langle 0,1 \rangle$. Note that we haven't drawn the actual position vector* **OP** *but rather illustrated a path to P by adding the two vectors.*

(a) *Draw a copy of Figure 2.8. Then on the same picture, draw the analogous path to the point Q whose position vector is* $\langle 2,1\rangle + \langle \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\rangle$. *(Of course, the first segment of the path is already drawn. Just fill in the second segment and indicate the point Q.)*

(b) *In the same picture, draw analogous paths to the points whose position vectors are* $\langle 2,1\rangle + \langle \cos(\theta), \sin(\theta)\rangle$ *for each of the following choices of* θ: $\frac{\pi}{3}$, $\frac{2\pi}{3}$, $\frac{7\pi}{6}$.

(c) *As* θ *varies over the entire interval* $[0,2\pi]$, *the resulting points fill out a familiar curve. Draw this curve and describe it in words. (If you find it helpful, you may want to go back to part (b) and make some additional choices of* θ *as well.)*

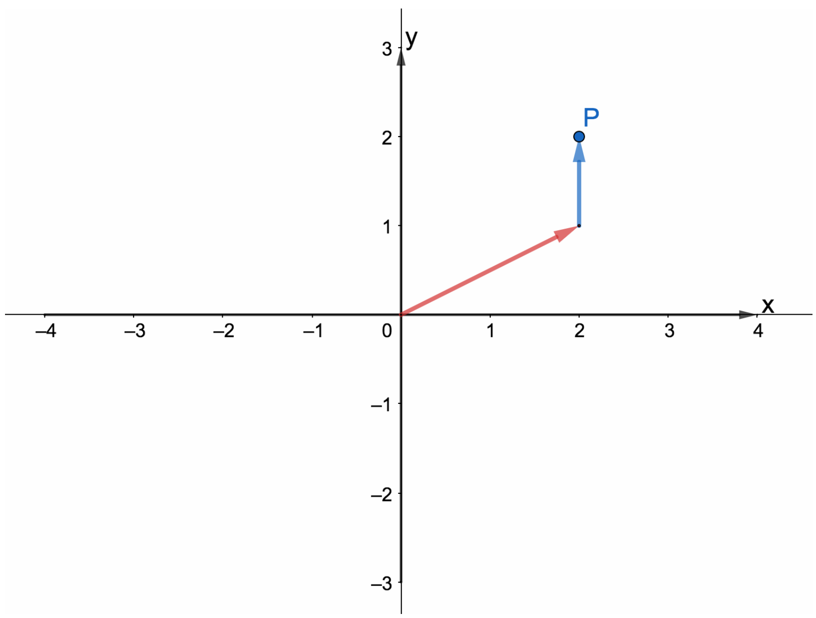


Figure 2.8

Exercise 2.1.14 *Describe the curve traced out by the points with position vectors* $\langle 2,5\rangle + \langle 3\cos(\theta), 3\sin(\theta)\rangle$ *as* θ *varies over the interval* $[0,2\pi]$. *Then write down the Cartesian equation of this curve.*

Exercise 2.1.15 *Let* $\mathbf{v} = \langle 1,1\rangle$, *let P be the point* $(2,3)$ *and let* **OP** *be the position vector of the point P.*

(a) *Using a single coordinate chart, draw the points with position vectors* $\mathbf{OP}$, $\mathbf{OP} \pm \mathbf{v}$, $\mathbf{OP} \pm 2\mathbf{v}$, *and* $\mathbf{OP} \pm \frac{1}{2}\mathbf{v}$. *(Don't draw the actual position vectors but rather draw the path to each point given by the indicated linear combination, just as you did in Exercise 4.4.) What do you notice?*

(b) *As t varies over all values in* $\mathbb{R}$, *the points with position vectors* $\mathbf{OP} + t\mathbf{v}$ *fill out a familiar curve. You sketched 7 points on this curve in part (a). In the same picture, draw the rest of the curve. (By "familiar curve" here, we mean things like circles, straight lines, parabolas, ...)*

(c) *Find the Cartesian equation (i.e., an equation involving x and y, no vectors) of the curve in part (b).*

Exercise 2.1.16 *Follow the directions of Exercise 2.1.15 but with* $\mathbf{v} = \langle 2,-1\rangle$ *and* $P = (0,1)$.

2.2 Using a basis to provide a map of a subspace

Motivation

You are used to labeling points in $\mathbb{R}^2$ by their (x,y)-coordinates. How do we label points on arbitrary lines and planes? In this section, we will consider lines and planes through the origin.

Outcomes

- *Understand the concepts of "spanning set", "linear independence", and "basis".*
- *Given a spanning set, find a basis.*
- *Use a basis to label points on lines and planes.*
- *Understand the concept of "parameters".*
- *Determine whether a point lies on a plane. If so, find its parameters with respect to a basis for the plane.*

Recall that lines through the origin, planes through the origin and $\mathbb{R}^3$ are all the subspaces of $\mathbb{R}^3$ of dimension at least one. (We are saying dimension at least one to avoid the trivial subspace $\{0\}$.)

We saw in the previous section that:

- To span a line ℓ through the origin, we need just a single non-zero vector parallel to ℓ.
- To span a plane $\mathscr{P}$ through the origin, we need just two vectors parallel to $\mathscr{P}$ but not parallel to each other.
- To span $\mathbb{R}^3$ we just need three vectors that aren't coplanar.

We emphasize that the dimension of $\mathrm{Span}(A)$ can *never* be greater than the number of vectors in A.

Definition 2.23

Let S be a subspace of $\mathbb{R}^3$ of dimension at least one.

- *We will say that a finite set A of vectors is a spanning set for S if $\mathrm{Span}(A) = S$.*
- *We say that A is a basis for S (also called a minimal spanning set) if $\mathrm{Span}(A) = S$ and the number of vectors in A equals the dimension of S. (Note that the second condition says that A is linearly independent.)*

Thus we have:

- A basis for a line through the origin consists of a single vector parallel to the line.
- A basis for a plane through the origin consists of 2 vectors that are both parallel to the plane but not parallel to each other.

- A basis for $\mathbb{R}^3$ consists of 3 vectors that are not coplanar.

Proposition 2.24

Every spanning set for a subspace S contains a basis for S. In other words, if the spanning set A is linearly dependent, then we can remove some vectors from A to obtain a basis.

Of course, we need to be careful which vectors we remove!

Example 2.25

For each of the following, you are given 3 vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$. Let $A = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Determine whether $\operatorname{Span}(A)$ is a line, a plane or $\mathbb{R}^3$. Then indicate whether A is linearly independent and find a basis for $\operatorname{Span}(A)$.

1. $\mathbf{v}_1 = \langle 1,0,1\rangle$, $\mathbf{v}_2 = \langle 2,0,2\rangle$, $\mathbf{v}_3 = \langle -4,0,-4\rangle$
2. $\mathbf{v}_1 = \langle 1,0,1\rangle$, $\mathbf{v}_2 = \langle 1,2,3\rangle$, $\mathbf{v}_3 = \langle 3,4,7\rangle$
3. $\mathbf{v}_1 = \langle 1,0,1\rangle$, $\mathbf{v}_2 = \langle 1,2,3\rangle$, $\mathbf{v}_3 = \langle 3,4,8\rangle$
4. $\mathbf{v}_1 = \langle 1,0,1\rangle$, $\mathbf{v}_2 = \langle 1,2,3\rangle$, $\mathbf{v}_3 = \langle 2,4,6\rangle$.

s

Solution.

1. The 3 vectors are parallel so $\operatorname{Span}(A)$ is a line. A is linearly dependent. Any one of the vectors by itself forms a basis for this line.
2. In Example 2.22, we saw that these 3 vectors are coplanar, so $\operatorname{Span}(A)$ is a plane $\mathscr{P}$ and A is linearly dependent. If we remove any one of the vectors from A, we get a pair of vectors that are parallel to $\mathscr{P}$ and not to each other. Thus any two of these vectors forms a basis for $\mathscr{P}$.
3. We saw in Example 2.22 that these 3 vectors are *not* coplanar. Thus $\operatorname{Span}(A) = \mathbb{R}^3$, so A is already a basis and A is linearly independent.. (If we remove any vector, we get a smaller span.)
4. Here $\mathbf{v}_2$ and $\mathbf{v}_3$ are parallel and thus A is coplanar. $\operatorname{Span}(A)$ is a plane and A is a linearly dependent set. We can remove either of $\mathbf{v}_2$ or $\mathbf{v}_3$ to obtain a basis. (On the other hand, we can't remove $\mathbf{v}_1$ as the span would then shrink to a line.)

♠

Let S be a subspace of dimension at least one. Corollary 2.11 tells us that if A is *any* spanning set for S, then the position vector of any point in S is a linear combination of the the vectors in A. E.g., suppose $\mathscr{P}$ is a plane through and $A = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a spanning set for $\mathscr{P}$. Then if P is a point in $\mathscr{P}$, we know that the equation

$$\mathbf{OP} = a\mathbf{v}_1 + b\mathbf{v}_2 + c\mathbf{v}_3 \tag{2.4}$$

has a solution. But in fact, when the spanning set A is not minimal as in this case (A has 3 vectors and $\mathscr{P}$ is only 2-dimensional), Equation (2.4) actually has infinitely many solutions, i.e., there are infinitely many

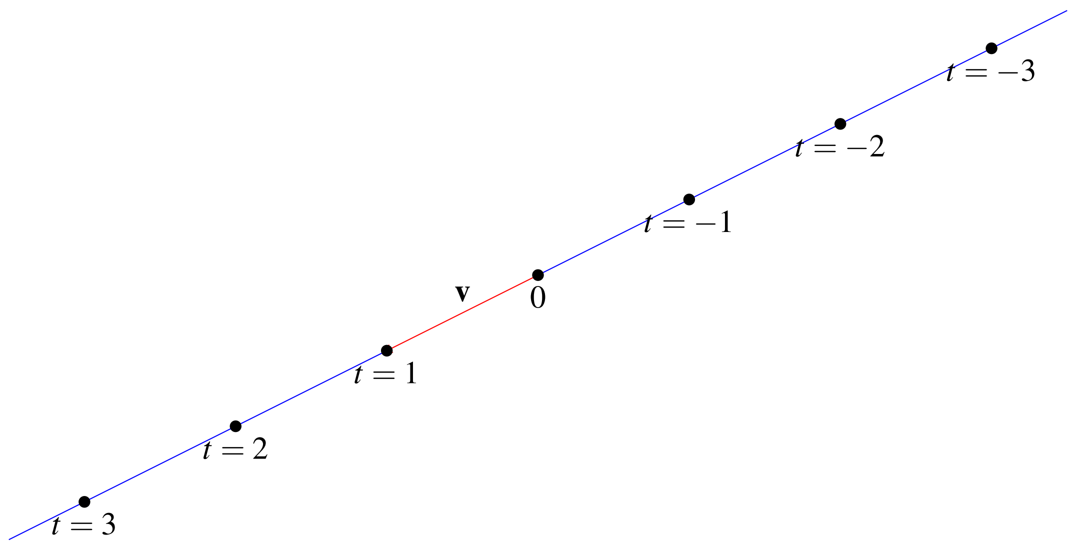


Figure 2.9

different ways of expressing $\mathbf{OP}$ as a linear combination of the vectors in A. In contrast, if we take a basis $\{\mathbf{v},\mathbf{w}\}$ for the plane, then the equation $\mathbf{OP} = s\mathbf{v}+t\mathbf{w}$ has a unique solution s,t. To summarize:

Theorem 2.26

If A is a basis for the subspace S, then the position vector for each point in S can be written in exactly one way as a linear combination of the vectors in A. In other words, the coefficients in the linear combination are unique determined.

You are used to labeling points in $\mathbb{R}$ by their x coordinate, and points in $\mathbb{R}^2$ by coordinates (x,y). But how do we label points on other lines and planes? Theorem 2.26 gives us a way:

(a) *Labeling points on a line ℓ through the origin.*

Choose a basis $A = \{\mathbf{v}\}$ for ℓ by choosing a non-zero vector parallel $\mathbf{v}$ to ℓ. Then the position vector of each point P on ℓ is given by $t\mathbf{v}$, and we have a one-to-one correspondence between points on ℓ and values of t. Figure 2.9 shows a line ℓ, a choice of $\mathbf{v}$ and the corresponding labeling of the points on the line. You can now identify points on the line by their label t. One way to view the labels is the way we view mile markers on a highway. Another point of view is to imagine that you are walking along the line and the label t tells you the point that you pass through at time t. Of course the labeling depends on the initial choice of $\mathbf{v}$, just as labeling points on a highway in terms of miles from, say, a state border differs from labeling them in terms of kilometers from the border.

(b) *Labeling points on a plane $\mathscr{P}$ through the origin.* Choose a basis $A = \{\mathbf{v},\mathbf{w}\}$ for $\mathscr{P}$ by choosing two vectors that are parallel to $\mathscr{P}$ but not parallel to each other. Theorem 2.26 tells us that for each point P on $\mathscr{P}$, there is exactly one pair of scalars s,t such that $\mathbf{OP} = s\mathbf{v}+t\mathbf{w}$. Thus we can label each point by the corresponding scalars s,t. (Again, if you choose a different basis, you will get a different label so you need to specify the basis that you are using.)

Notation 2.27

In linear algebra, the scalar t in (a) and the ordered pair of scalars s,t in (b) are called the coordinates of P with respect to the basis A. In preparation for the next section, we will instead refer to them as the parameters for the point P with respect to the basis A. (If you have made clear what basis you are using, you can just say 'parameters' for P".)

Example 2.28

- *Let $\mathscr{P}$ by the xy-plane and let A be the standard basis $\{\mathbf{i},\mathbf{j}\}$. Let P be the point whose x,y coordinates are (x_0,y_0). We have $\mathbf{OP} = x_0\mathbf{i} + y_0\mathbf{j}$. Thus x_0,y_0 are the parameters for P with respect to the standard basis. (This is the reason that the language "coordinates for P with respect to the basis" is used in linear algebra.)*

- *Let A be the basis for $\mathbb{R}^2$ given by $A = \{\langle 2,1\rangle, \langle 1,3\rangle\}$. Figure 2.4 in the previous section illustrates how this basis gives a "road map". If we imagine that your house is located at a point P, the parameters s,t for P can be viewed as your home address; they uniquely identify the location of your home. For example the point P in Figure 2.5 in the previous section has parameters $-2.7, 1.6$ since $\mathbf{OP} = -2.7\mathbf{v} + 1.6\mathbf{w}$.*

Example 2.29: Determining whether a point lies on a plane

Let $\mathscr{P}$ be the plane through the origin parallel to the vectors $\mathbf{v} = \langle 1,0,1\rangle$ and $\mathbf{w} = \langle 1,1,2\rangle$. Observe that $A = \{\mathbf{v},\mathbf{w}\}$ is a basis for $\mathscr{P}$.
Let $P = (5,3,8)$. Determine whether P lies on the plane $\mathscr{P}$. If it does, find its parameters with respect to the basis A.

Solution. Since $\mathbf{OP} = \langle 5,3,8\rangle$, the point P lies on $\mathscr{P}$ if and only if the equation

$$\langle 5,3,8\rangle = s\langle 1,0,1\rangle + t\langle 1,1,2\rangle$$

has a solution. Since A is a basis, we know that the solution s,t (if it exists) is unique and these will be the parameters for P.

We need to solve the system:

$$s + t = 5$$

$$0s + t = 3$$

$$s + 2t = 8$$

We leave it to the reader to solve the system (using either method in Example 2.7). It does have a solution $s = 2, t = 3$. Thus P lies on the plane and its parameters with respect to the basis A are $2,3$. ♠

As the example above illustrates, you can check whether a point P lies on a given plane $\mathscr{P}$ through $\mathbf{0}$ by choosing a basis for $\mathscr{P}$ and checking whether the position vector $\mathbf{OP}$ is a linear combination of the basis vectors. (Note: You could instead use a spanning set that's not a basis, but this will result in a system of more equations than if you use a basis. Moreover if P lies on the plane, you will then get infinitely many solutions instead of just one so you won't be able to use the solution to label the point.)

Definition 2.30: Linear independence

We say that a finite set A of vectors is linearly independent if the number of vectors in A is equal to the dimension of $\operatorname{Span}(A)$*. Otherwise we say that the set A of vectors is linearly dependent.*

Thus:

- A set consisting of a single non-zero vector is linearly independent, since $\operatorname{Span}(A)$ is a line and thus 1-dimensional.
- A set A containing two vectors is linearly independent if $\operatorname{Span}(A)$ is a plane, equivalently if the vectors are not parallel. It is linearly dependent if the two vectors are parallel.
- A set A consisting of 3 vectors is linearly independent if $\operatorname{Span}(A) = \mathbb{R}^3$, equivalently, the vectors are not coplanar. It is linearly dependent if the vectors are coplanar.

2.2.1. Section Summary

- A set A of vectors is linearly independent if the number of vectors in A is equal to the dimension of $\operatorname{Span}(A)$.
- A basis for a subspace S is a linearly independent set of vectors that spans S.
- A basis for a line through 0 consists of a single vector parallel to the line. A basis for a plane through 0 consists of two vectors parallel to the plane but not parallel to each other. A basis for $\mathbb{R}^3$ consists of 3 vectors that aren't coplanar.
- Every spanning set for a subspace S contains a basis for S.
- If A is a basis for a subspace, the position vector of every point in the subspace can be expressed in exactly one way as a linear combination of the basis elements. Thus if, say, $A = \{\mathbf{v}, \mathbf{w}\}$ is a basis for a plane $\mathscr{P}$, then every point P in $\mathscr{P}$ has position vector $\mathbf{OP} = s\mathbf{v} + t\mathbf{w}$ and we can use the pair s, t as a label for the point P. We call s, t parameters.
- To check whether a point lies on a given plane $\mathscr{P}$ through 0, choose a basis and check whether the position vector $\mathbf{OP}$ is a linear combination of the basis vectors.

Exercises

Exercise 2.2.1 *Check whether each of the following sets of vectors is linearly independent or dependent:*

(a) $A = \{\langle 3,1\rangle, \langle -6,-2\rangle\}$.

(b) $A = \{\langle 3,1\rangle, \langle -6,2\rangle\}$.

(c) $A = \{\langle 3,1,4\rangle, \langle 3,1,5\rangle\}$.

(d) $A = \{\langle 3,1,1\rangle, \langle 1,2,1\rangle, \langle 9,8,5\rangle\}$.

(e) $A = \{\langle 3,1,1\rangle, \langle 1,2,1\rangle, \langle 9,8,0\rangle\}$.

(f) $A = \{\langle 3,1,1\rangle, \langle -6,-2,-2\rangle, \langle 9,8,0\rangle\}$.

Exercise 2.2.2 *For each of the sets A in Exercise 2.2.1, find a basis for* $\mathrm{Span}(A)$ *consisting of vectors from A. (In some cases, there may not be a unique answer. You just need to give one choice of basis.)*

Exercise 2.2.3 *For each of the following points P in* $\mathbb{R}^2$ *or* $\mathbb{R}^3$*, find the parameters for the point P with respect to the given basis B of* $\mathbb{R}^2$*, respectively* $\mathbb{R}^3$.

(a) $P = (1,1)$, $B = \{\langle 4,5\rangle, \langle 2,3\rangle\}$

(b) $P = (0,0,-2)$, $B = \{\langle 2,0,1\rangle, \langle 1,2,-1\rangle, \langle 3,2,2\rangle\}$

Exercise 2.2.4 *True or false: Let* $\mathbf{v}$ *and* $\mathbf{w}$ *be any two non-zero vectors in* $\mathbb{R}^3$ *that aren't parallel. Then* $\mathrm{Span}(\mathbf{v},\mathbf{w}) = \mathrm{Span}(\mathbf{v}+2\mathbf{w},\mathbf{w})$. *Explain your answer.*

Exercise 2.2.5 *Answer the same question as in Exercise 2.2.4, but now assume that* $\mathbf{v}$ *and* $\mathbf{w}$ *are parallel.*

2.3 Vector Equations of Lines and Planes

Motivation

In the previous section, we saw how to label points on lines and planes through the origin using a basis.

In this section, we will explore how to represent arbitrary lines and planes, not just those that pass through the origin, via vectors. In the process we will see our first examples of vector-valued functions.

Outcomes

- *Understand the concepts of vector and parametric equations of lines and planes.*
- *Be able to write down the parametric equations of lines and planes when enough information is given to determine the line or plane.*
- *Given lines in $\mathbb{R}^3$, check whether they intersect in a single point, are parallel or skew.*

2.3.1. Lines in $\mathbb{R}^2$ or $\mathbb{R}^3$

Given a point P and a direction specified by a vector $\mathbf{v}$, there is a unique line through P parallel to $\mathbf{v}$. The vector $\mathbf{v}$ is called a direction vector for the line. (Aside: if the line ℓ passes through the orgin, then as discussed in the previous section, a basis for ℓ consists of a direction vector for ℓ. The word "basis", however, is not used for lines that do not pass through the origin.)

As a first example, consider the line ℓ through the point $P = (1,3)$ parallel to the vector $\mathbf{v} = \langle 2,1\rangle$. Figure 2.10 below is a copy of Figure 2.4 that appeared in Section 2.1. The vector $\mathbf{w} = \langle 1,3\rangle$ is the position vector of the point P. As you move from $(1,3)$ (the point labelled $\mathbf{w}$ in Figure 2.4) in the directions $\pm\mathbf{v}$, you follow the line in Figure 2.10 passing through the points labelled with their position vectors $\mathbf{w}\pm\mathbf{v}$, $\mathbf{w}\pm 2\mathbf{v}$, etc. Observe that the direction vector of every point Q on the line ℓ is of the form

$$\mathbf{OQ} = \mathbf{w} + t\mathbf{v} = \langle 1,3\rangle + t\langle 2,1\rangle. \tag{2.5}$$

Just as in the previous section, we can think of each value of t as a label for the corresponding point Q on the line. The line ℓ is redrawn in Figure 2.11. A few points corresponding to integer values of t are labelled. The front of the car (driving uphill) is approximately at the point corresponding to $t = \frac{3}{2}$, i.e., the point $(1+\frac{3}{2}(2), 3+\frac{3}{2}(1)) = (4, \frac{9}{2})$.

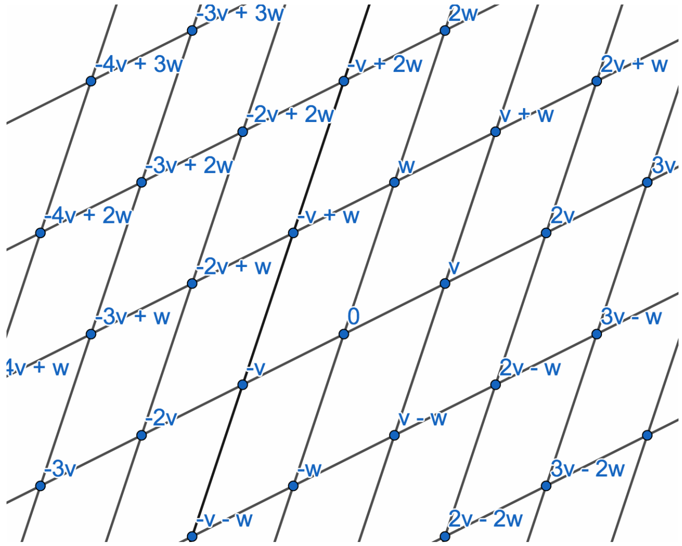

-3v + 3w
2w
-4v + 3w
-v + 2w
2v + w
-2v + 2w
v + w
-3v + 2w
w
3v
-4v + 2w
-v + w
2v
-2v + w
v
-3v + w
0
3v - w
4v + w
-v
2v - w
-2v
v - w
-3v
-w
3v - 2w
-v - w
2v - 2w

Figure 2.10

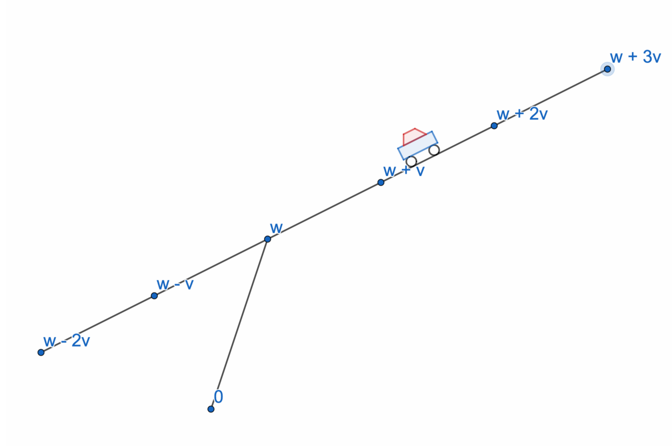


Figure 2.11

There are a couple ways to give an equation for this line. Since we are working in $\mathbb{R}^2$, we can express it with a Cartesian equation. From the direction vector $\mathbf{v} = \langle 2, 1\rangle$, we can read off that the slope (rise/run) is $\frac{1}{2}$. So the Cartesian equation is $y-3 = \frac{1}{2}(x-1)$ or $x-2y = -5$. In contrast, however, lines in $\mathbb{R}^3$ cannot be expressed by a single Cartesian equation.

A second way to give an equation for this line – a method that will work in 3 dimensions as well – is to take the point of view of the driver of the car. Imagining that t denotes time and that the driver is at the point corresponding to $\langle 1,3\rangle + t\langle 2,1\rangle$ at time t, we can write down a vector-valued function that gives the position vector of the car at time t:

$$\mathbf{r}(t) = \langle 1,3\rangle + t\langle 2,1\rangle. \tag{2.6}$$

This function gives an equation for the line in the sense that as t varies, $\mathbf{r}(t)$ traces out the line. The function inputs a "label" t and outputs the point on the line with that label. An equation of this form is called a vector equation for the line. The variable t is often called a parameter.

There are lots of different ways to give a vector equation for the same line. The driver of the car in the picture was at point (1,3) at time 0. A different driver or pedestrian might start at a different point on the line at time 0 and might go faster or slower than the car depicted or might go downhill rather than uphill. So (1,3) can be replaced by the position vector of any point on the line and the choice of direction vector $\langle 2,1\rangle$ can be replaced by any non-zero scalar multiple of $\langle 2,1\rangle$.

A final way to represent this line (just a slight variation on the vector equation method) is by what are called parametric equations. We know from the Equation (2.6) that a point (x,y) is on the line if and only if its position vector satisfies $\langle x,y\rangle = \langle 1,3\rangle + t\langle 2,1\rangle$ for some t. So we can write

$$x = 1+2t, \;\; y = 3+t \tag{2.7}$$

The pair of equations (2.7) are called parametric equations of the line. Again, as t varies, the parametric equations yield all points (x,y) on the line.

The method we just used to represent the line in Figure 2.11 works equally well for any line in $\mathbb{R}^2$ or $\mathbb{R}^3$. We summarize by describing lines in $\mathbb{R}^3$: (Of course, you can replace $\mathbb{R}^3$ by $\mathbb{R}^2$.)

2.31: Vector and parametric equations of lines in $\mathbb{R}^3$

Given a point $P = (x_0,y_0,z_0)$ in $\mathbb{R}^3$ and a non-zero vector $\mathbf{v} = \langle a,b,c\rangle$, the vector equation of the line in $\mathbb{R}^3$ through P and parallel to $\mathbf{v}$ is given by

$$\mathbf{r}(t) = \mathbf{OP} + t\mathbf{v} = \langle x_0,y_0,z_0\rangle + t\langle a,b,c\rangle.$$

Simplifying the right hand side, we get $\mathbf{r}(t) = \langle x_0+at, y_0+bt, z_0+ct\rangle$. The line can also be written in parametric equations as

$$x = x_0+at, \;\; y = y_0+bt, \;\; z = z_0+ct$$

Example 2.32: Line through two points

One of the axioms of Euclidean geometry is that there is a unique line through any two distinct points. Find the vector and parametric equations of the line through the points $P:(1,2,3)$ and $Q:(2,1,5)$.

Solution. *The vector $\mathbf{PQ} = \langle 1,-1,2\rangle$ from P to Q is parallel to the line, so we can use it for our direction vector. Thus we get the vector equation:*

$$\mathbf{r}(t) = \langle 1,2,3\rangle + t\langle 1,-1,2\rangle$$

and parametric equations

$$x = 1+t, \;\; y = 2-t, \;\; z = 3+2t$$

♠

In $\mathbb{R}^2$, any pair of lines that are not parallel must intersect in a point. In $\mathbb{R}^3$, there are more possibilities: a pair of lines ℓ_1 and ℓ_2 could:

- be parallel
- intersect in a single point
- be skew

Skew means that the lines are not coplanar. Skew lines do not intersect but they are also not parallel. For example, let ℓ_1 be the x-axis and let ℓ_2 be any line lying in the plane $z = 1$ that isn't parallel to the x-axis. Then ℓ_1 and ℓ_2 are skew. See Figure 2.12 for a drawing of a pair of skew lines. Two randomly chosen lines are most likely to be skew.

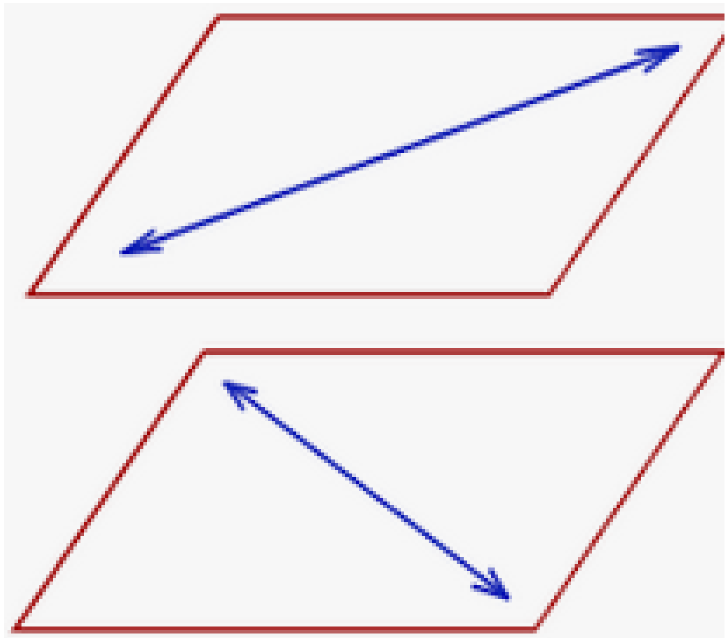


Figure 2.12

To compare two lines, you'll probably want to begin by looking at their direction vectors to see if they are parallel. If you find that they are parallel, it's important to check whether the two lines really are different lines or whether they coincide. Remember, there are many different vector/parametric equations for the same line!

Example 2.33: Compare two lines

Compare the following two lines:

$$\ell_1 : \mathbf{r}(t) = \langle 1,2,3\rangle + t\langle 1,-1,2\rangle$$

$$\ell_2 : \mathbf{q}(t) = \langle 3,1,4\rangle + t\langle -3,3,-6\rangle$$

Solution. *Since the direction vectors $\langle 1,-1,2\rangle$ and $\langle -3,3,-6\rangle$ are parallel, the two lines are parallel. We give two methods for checking whether the lines coincide:*
Method 1. *Pick a point on ℓ_2, say (3,1,4) and see whether it lies on ℓ_1. If it does, then the two lines will be the same since parallel lines that intersect must coincide. If it doesn't, then the lines must be different.*
The point (3,1,4) lies on the first line if and only if there is some value of t such that $\mathbf{r}(t) = \langle 3,1,4\rangle$. In other words, $3 = 1+t,\ 1 = 2-t,$ and $4 = 3+2t$. The first equation says $t = 2$, which contradicts the other equations. So the two lines are parallel but not the same line.
Method 2. *The point $P_1 : (1,2,3)$ lies on ℓ_1 and the point $P_2 : (3,1,4)$ lies on ℓ_2. If ℓ_1 is actually equal to ℓ_2, then both P_1 and P_2 lie on ℓ_1, so the displacement vector $\mathbf{P}_1\mathbf{P}_2$ must be parallel to this line. But $\mathbf{P}_1\mathbf{P}_2 = \langle 2,-1,1\rangle$ is not parallel to the direction vector $\langle 1,-1,2\rangle$. So $\ell_1 \neq \ell_2$.* ♠

Next suppose that you are comparing two lines ℓ_1 and ℓ_2 expressed by vector-valued functions $\mathbf{r}(t)$ and $\mathbf{q}(t)$, respectively (or expressed in parametric form), and you have already established that they are not parallel.

Caution! If you think of t as time, then the vector equations for the two lines each give the position of, say, a pedestrian walking along the line at time t. If you walk down Main Street passing by the intersection with Wheelock Street at 3:00, and your friend walks along Wheelock passing this intersection at 5:00, you're not going to see each other! Thus, when we ask whether the lines intersect, we are *not* asking whether there is some time t such that $\mathbf{r}(t)$ and $\mathbf{q}(t)$ are equal, but rather whether there is some t_1 and some possibly different t_2 such that $\mathbf{r}(t_1) = \mathbf{q}(t_2)$. To check, you first need to change the name of one of the parameters as in the following example.

Example 2.34: Intersect or skew?

Determine whether the following two lines are parallel, whether they intersect in a point (if so, find the point) or are skew:

$$\ell_1 : \mathbf{r}(t) = \langle 1,5,3\rangle + t\langle 1,-1,2\rangle$$
$$\ell_2 : \mathbf{q}(t) = \langle 3,1,4\rangle + t\langle 1,1,5\rangle$$

Solution. *First check that the direction vectors are not parallel (left to the reader).*
Following the "caution" above, we rename the second parameter s:

$$\ell_2 : \mathbf{q}(s) = \langle 3,1,4\rangle + s\langle 1,1,5\rangle$$

Now, we set $\mathbf{r}(t) = \mathbf{q}(s)$ and see if there is a solution. This gives us 3 equations in 2 variables:

$$1+t = 3+s,\ 5-t = 1+s,\ 3+2t = 4+5s$$

or simplifying:

$$s-t = -2,\ s+t = 4,\ 5s-2t = -1$$

The unique solution of the first two equations is $s = 1$, $t = 3$. Since this satisfies the third equation as well, we have $\mathbf{r}(3) = \mathbf{q}(1)$, so the two lines do intersect. To find the point of intersection, plug $t = 3$ into the equation for ℓ_1 to get the point $(4,2,9)$. (It's a good idea to check your work by verifying that $\mathbf{q}(1)$ also gives $(4,2,9)$.) ♠

2.3.2. Vector equations of planes

In Section 2.2 we saw that any plane through the orgin can be expressed as $\mathrm{Span}(\mathbf{v},\mathbf{w})$ where $\mathbf{v},\mathbf{w}$ are linearly independent vectors that are both parallel to the plane and we labelled the point with position vector $s\mathbf{v}+t\mathbf{w}$ by the parameters s,t. Analogous to the vector equation for a line, we obtain a vector-valued function, this time with two independent variables s and t for this plane:

$$\mathbf{r}(s,t) = s\mathbf{v}+t\mathbf{w}.$$

Again this vector equation inputs a label (the pair of parameters) and outputs the corresponding point on the plane. As s and t vary, the points with position vectors $\mathbf{r}(s,t)$ fill up the plane.

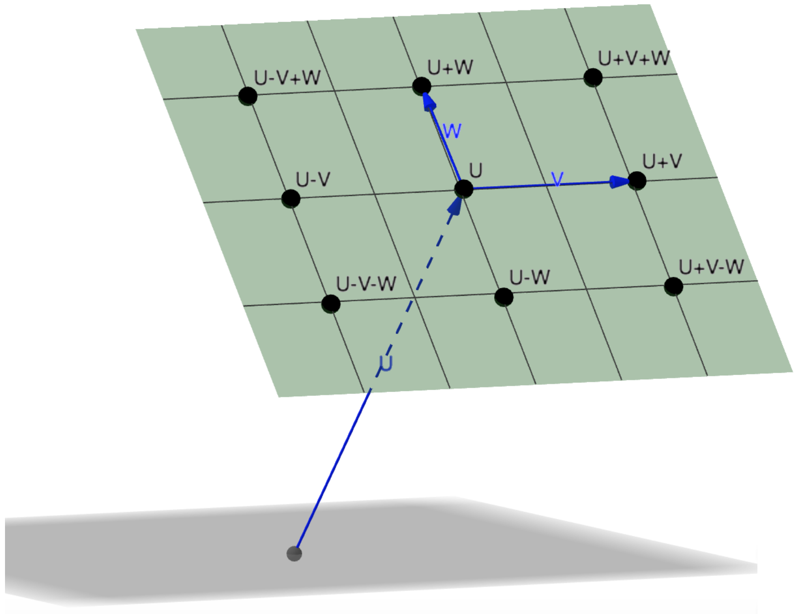


Figure 2.13

More generally, we can specify any plane – not necessarily passing through the origin – by specifying (i) a linearly independent pair of vectors $\mathbf{v}, \mathbf{w}$ parallel to the plane and (ii) a point P on the plane. The vectors $\mathbf{v}, \mathbf{w}$ tell us that the plane is parallel to $\text{Span}(\mathbf{v}, \mathbf{w})$; the point P situates the plane in space. We can reach any point in the plane by starting at P and then moving parallel to $\text{Span}(\mathbf{v}, \mathbf{w})$. Thus, as illustrated in Figure 2.13, the position vectors of all points on the plane can be expressed as $\mathbf{OP} + s\mathbf{v} + t\mathbf{w}$. We obtain a vector equation of the plane:

$$\mathbf{r}(s,t) = \mathbf{OP} + s\mathbf{v} + t\mathbf{w} \tag{2.8}$$

If $P = (x_0, y_0, z_0)$, $\mathbf{v} = \langle a_1, b_1, c_1 \rangle$ and $\mathbf{w} = \langle a_2, b_2, c_2 \rangle$, we get

$$\mathbf{r}(s,t) = \langle x_0, y_0, z_0 \rangle + s\langle a_1, b_1, c_1 \rangle + t\langle a_2, b_2, c_2 \rangle.$$

Writing $\langle x, y, z \rangle = \mathbf{r}(s,t)$, we obtain parametric equations for the plane:

$$x = x_0 + a_1 s + a_2 t, \ \ y = y_0 + b_1 s + b_2 t, \ \ z = z_0 + c_1 s + c_2 t \tag{2.9}$$

As with lines, there are infinitely many different vector equations of the same plane, since there are infinitely many ways to choose P on the plane and to choose the vectors $\mathbf{v}, \mathbf{w}$.

Example 2.35: Plane through three points

Any three distinct non-collinear points determine a unique plane. ("Non-collinear" means that they don't all lie on a single line.)
Find vector and parametric equations of the plane containing the points $P_1 : (1,2,0)$, $P_2 : (3,1,5)$ and $P_3 : (2,0,1)$.

Solution. *Since the points P_1, P_2, P_3 all lie on the plane, the displacement vectors $\mathbf{P_1P_2}$ and $\mathbf{P_1P_3}$ are parallel to the plane. Let $\mathbf{v} = \mathbf{P_1P_2} = \langle 2,-1,5\rangle$ and $\mathbf{w} = \mathbf{P_1P_3} = \langle 1,-2,1\rangle$. Using P_1 as our initial point on the plane, we then obtain the vector equation*

$$\mathbf{r}(s,t) = \langle 1,2,0\rangle + s\langle 2,-1,5\rangle + t\langle 1,-2,1\rangle$$

and parametric equations

$$x = 1+2s+t,\ y = 2-s-2t,\ z = 5s+t$$

♠

Preview of Things to Come 2.36

The vector equations of lines and planes through the origin are examples of what are called linear transformations, the topic of a later chapter of these notes.

Exercises

Exercise 2.3.1 *Write down a vector equation for each of the following lines in $\mathbb{R}^3$:*

(a) *the line through the point $(1,5,6)$ and parallel to the vector $\langle 4,7,8\rangle$*

(b) *the line through the point $(1,4,7)$ and parallel to the y-axis*

(c) *the line containing the points $(1,4,5)$ and $(2,1,6)$*

(d) *the line containing the points $(3,1,2)$ and $(1,7,5)$*

(e) *the line through the point $(1,2,3)$ and parallel to the line whose parametric equations are $x = 5+2t$, $y = 4+3t$, $z = 8-t$*

Exercise 2.3.2 *Write down parametric equations of each of the lines in Exercise 2.3.1*

Exercise 2.3.3 *Find the Cartesian equation of the line in $\mathbb{R}^2$ represented by the given vector equation:*

(a) $\mathbf{r}(t) = t\langle 3,8\rangle$

(b) $\mathbf{r}(t) = \langle 4,1\rangle + t\langle 3,5\rangle$

(c) $\mathbf{r}(t) = \langle 1+7t, 2-t \rangle$

Exercise 2.3.4 *Give a vector equation representing each of the following lines in* $\mathbb{R}^2$*:*

(a) $3x+4y=8$

(b) $5x-y=10$

(c) $y=5$

(d) $x=2$

Exercise 2.3.5 *Let t measure time in minutes and let units in the (x,y)-plane be measured in feet (so the vectors* $\mathbf{i}$ *and* $\mathbf{j}$ *are each one foot long). Suppose that a sluggish turtle is at the point $(1,1)$ at time 0 and crawls along a straight line ℓ parallel to the vector $\langle 3,4 \rangle$ at the constant speed of one foot per minute. Write down a vector equation for this line, where* $\mathbf{r}(t)$ *gives the turtle's position at time t. (Thus you must make a careful choice of direction vector.)*

Exercise 2.3.6 *Suppose that a rabbit hops along the same line ℓ as the turtle in Exercise 2.3.5, starting from the same point but going at the constant speed of 100 feet per minute. Write down a vector equation representing the rabbit's position at time t.*

Exercise 2.3.7 *Suppose that, instead of starting at the same point as the turtle, the rabbit in Exercise 2.3.6 is at some point on ℓ behind the turtle at time 0 and, hopping at the constant speed of 100 feet per minute, overtakes the turtle one minute later. Write down a vector equation representing the rabbit's position at time t.*

Exercise 2.3.8 *Give a vector equation for each of the following planes:*

(a) The plane through the point $(1,2,1)$ and parallel to both $\langle 5,1,6 \rangle$ and $\langle 3,3,2 \rangle$

(b) The plane containing the points $(1,2,1)$, $(0,5,2)$ and $(3,2,4)$

(c) The plane containing the points $(1,0,1)$, $(4,5,8)$ and $(3,2,4)$

(d) The plane through the point $(1,2,1)$ and parallel to the xy-plane

(e) The plane through the point $(1,2,1)$ and parallel to the xz-plane

Exercise 2.3.9 *Write down parametric equations for each of the planes in Exercise 2.3.8.*

3. Introduction to Matrices and Matrix Arithmetic

Motivation

Matrices give us a way of organizing and representing information in an array. Matrix operations such as addition and multiplication are defined in a way that makes matrices powerful tools in business, economics, and in virtually every area of science. In mathematics, matrices as well as vectors are the foundations for linear algebra and will enable us to define derivatives later in the course.

3.1 Matrices

A matrix is a rectangular array of numbers such as

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{bmatrix} \tag{3.1}$$

The size of a matrix is defined as $m \times n$ where m is the number of rows and n is the number of columns. Thus the matrix above is a 3×4 matrix. When specifying the size of a matrix, you always list the number of rows before the number of columns.

There are several matrix sizes with special names:

- A matrix with the same number of rows as columns is called a square matrix, e..g, $\begin{bmatrix} 2 & 4 \\ 1 & 5 \end{bmatrix}$
- A $1 \times n$ matrix (so only one row) such as $\begin{bmatrix} 2 & 4 & 1 \end{bmatrix}$ is called a row matrix. Often it is useful to identify row matrices with vectors. Thus it is also referred to as a row vector.
- Similarly, an $m \times 1$ matrix (so only one column) is called a column matrix or column vector.

Notation 3.1: Entries, Rows, Columns

- *The entry in row i, column j of a matrix A is referred to as the (i, j)-entry of A and will usually be denoted A_{ij}. We may deviate from this double-subscript notation when there is only a single row or column.*
- *We will denote the ith row of A by $\mathrm{Row}_i(A)$ and the jth column by $\mathrm{Col}_j(A)$.*

Example 3.2

Let

$$A = \begin{bmatrix} 1 & 4 & 3 \\ 5 & 2 & 8 \end{bmatrix}$$

Then A is a 2×3 matrix. Examples of its entries are $A_{12} = 4$, $A_{21} = 5$, etc. An example of a row is

$$\mathrm{Row}_2(A) = \begin{bmatrix} 5 & 2 & 8 \end{bmatrix}$$

and an example of a column is

$$\mathrm{Col}_3(A) = \begin{bmatrix} 3 \\ 8 \end{bmatrix}$$

Definition 3.3: Equality of Matrices

Let A and B be two $m \times n$ matrices. Then $A = B$ means that the matrices have the same size and $A_{ij} = B_{ij}$ for all i, j.

Thus

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

because their corresponding entries are not identical.

3.2 Matrix Operations

3.2.1. Addition of Matrices

We can add two matrices A and B if and only if they are the same size. We then add the corresponding entries, i.e.,

$$(A+B)_{ij} = A_{ij} + B_{ij}.$$

For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 4 \end{bmatrix} + \begin{bmatrix} 5 & 2 & 3 \\ -6 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1+5 & 2+2 & 3+3 \\ 1+-6 & 0+2 & 4+1 \end{bmatrix} = \begin{bmatrix} 6 & 4 & 6 \\ -5 & 2 & 5 \end{bmatrix}$$

Addition of matrices obeys very much the same properties as normal addition with numbers.

Proposition 3.4: Properties of Matrix Addition

Let A,B and C be $m\times n$ matrices. Then, the following properties hold.

- *Commutative Law of Addition*
$$A+B=B+A \tag{3.2}$$

- *Associative Law of Addition*
$$(A+B)+C=A+(B+C) \tag{3.3}$$

- *Existence of an Additive Identity*
$$\begin{array}{c}\text{There exists an } m\times n \text{ zero matrix } 0 \text{ such that}\\ A+0=A\end{array} \tag{3.4}$$
(The zero matrix has all entries equal to zero.)

- *Existence of an Additive Inverse*
$$\begin{array}{c}\text{There exists a matrix } -A \text{ such that}\\ A+(-A)=0\end{array} \tag{3.5}$$

We may refer to the $m\times n$ zero matrix just as the zero matrix if the size is understood. You can easily check that the additive inverse $-A$ is obtained by simply changing the signs of all the entries. E.g.,

$$\text{if } A=\begin{bmatrix}5 & 2 & -8\\ 1 & -4 & 2\end{bmatrix} \text{ then } -A=\begin{bmatrix}-5 & -2 & 8\\ -1 & 4 & -2\end{bmatrix}$$

3.2.2. Scalar Multiplication of Matrices

To multiply a matrix by a scalar, we multiply each entry by the scalar. For example,

$$3\begin{bmatrix}1 & 2 & 3 & 4\\ 5 & 2 & 8 & 7\\ 6 & -9 & 1 & 2\end{bmatrix}=\begin{bmatrix}3 & 6 & 9 & 12\\ 15 & 6 & 24 & 21\\ 18 & -27 & 3 & 6\end{bmatrix}$$

Proposition 3.5: Properties of Scalar Multiplication

Let A, B be matrices, and k, p be scalars. Then, the following properties hold.

- *Distributive Law over Matrix Addition*
$$k(A+B) = kA + kB$$
- *Distributive Law over Scalar Addition*
$$(k+p)A = kA + pA$$
- *Associative Law for Scalar Multiplication*
$$k(pA) = (kp)A$$
- *Rule for Multiplication by* 1
$$1A = A$$

The proof of this proposition is left to the reader.

3.2.3. Multiplication of Matrices

While the definitions of addition and scalar multiplication of matrices probably appeared very natural, multiplication is not defined in the way one would first expect, i.e., we don't simply multiply corresponding entries. The reason we don't do this is that it doesn't seem to be useful for anything. It's easy to see why the notions defined in the previous subsections of addition and multiplication by scalars are important. In business applications, for example, you might want to chart daily information about different products, where the entries in the various rows and columns correspond to different data such as production costs, etc. When adding two such matrices, you get the combined costs, etc. over a two-day period.

The initially strange looking notion of multiplication defined below has powerful applications.

First, what size matrices can we multiply?

In order to form the product AB, the number of columns of A must equal the number of rows of B. The resulting matrix AB that we will define below will have the same number of rows as A and the same number of columns as B.

$$(m \times \overbrace{n)\,(n}^{\text{these must match!}} \times p) = m \times p$$

Given an $m \times n$ matrix A and an $n \times p$ matrix B, how do we define AB?

We first address the case that B is a column matrix.

3.2.3.1. Multiplying a matrix times a column vector

Before writing down the formal definition, we illustrate with an example:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} = 7 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + 8 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + 9 \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 50 \\ 122 \end{bmatrix} \tag{3.6}$$

Observe that the right hand side is a linear combination of the columns of A. The entries of B become the coefficients in the linear combination.

Definition 3.6: Product of a matrix and a column vector

If A is an $m \times n$ matrix and $B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ is an $n \times 1$ column matrix, then the product AB is defined to be the linear combination of the columns of A with coefficients specified by the entries of B. In other words,

$$AB = b_1 \operatorname{Col}_1(A) + b_2 \operatorname{Col}_2(A) + \cdots + b_n \operatorname{Col}_n(A) = \sum_{j=1}^{n} b_j \operatorname{Col}_j(A).$$

Example 3.7

Let

$$A = \begin{bmatrix} 2 & 1 \\ 4 & 5 \\ 8 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

Express AB as a linear combination of the columns of A and then compute it.

Solution.

$$AB = 3 \begin{bmatrix} 2 \\ 4 \\ 8 \end{bmatrix} - \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 22 \end{bmatrix}$$

♠

If we view the columns of A as vectors, then Definition 3.6 says that the column vector AB lies in the span of the columns of A. This fact will play an important role when we study linear transformations.

Observation 3.8

Looking carefully at Definition 3.6 or the examples above, you'll see that the entry in the ith row of the product AB is the dot product of the ith row of A with the column matrix B (viewing both as vectors). E.g., in Equation 3.6, the first entry 50 of the product is $(1)(7) + 2(8) + 3(9)$, which is precisely $\langle 1,2,3 \rangle \cdot \langle 7,8,9 \rangle$, the dot product of the row $\operatorname{Row}_1(A)$ with the column vector B. Similarly $\langle 4,5,6 \rangle \cdot \langle 7,8,9 \rangle = 122$, the second entry.

When B has more that one column, we multiply A by each column of B to get the columns of the product:

Definition 3.9: Multiplication of Two Matrices

Let A be an $m \times n$ matrix and B an $n \times p$ matrix. The product AB is defined to be the $m \times p$ matrix whose jth column is given by $A\,\mathrm{Col}_j(B)$. (We know how to compute $A\,\mathrm{Col}_j(B)$ by Definition 3.6.)

Putting together Observation 3.8 and Definition 3.9, we obtain what you will probably find to be an easier way to compute matrix products:

Proposition 3.10: The ij-entry of the product AB

Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Then AB is the $m \times p$ matrix whose ij-entry is the dot product of the ith row of A with the jth column of B:

$$(AB)_{ij} = \mathrm{Row}_i(A) \cdot \mathrm{Col}_j(B) = \sum_{k=1}^{n} A_{ik}B_{kj}.$$

Example 3.11

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 5 & 0 \end{bmatrix} \begin{bmatrix} 1 & 5 \\ 6 & 2 \end{bmatrix} = \begin{bmatrix} \langle 2,1\rangle \cdot \langle 1,6\rangle & \langle 2,1\rangle \cdot \langle 5,2\rangle \\ \langle 3,4\rangle \cdot \langle 1,6\rangle & \langle 3,4\rangle \cdot \langle 5,2\rangle \\ \langle 5,0\rangle \cdot \langle 1,6\rangle & \langle 5,0\rangle \cdot \langle 5,2\rangle \end{bmatrix} = \begin{bmatrix} 8 & 12 \\ 27 & 23 \\ 5 & 25 \end{bmatrix}$$

Example 3.12: The Identity Matrix

For each n, there is an $n \times n$ square matrix I_n (usually just denoted I if n is understood) that plays a similar role in matrix multiplication as the number one does in scalar multiplication. This matrix has 1s going down what is called the main diagonal (the diagonal from the upper left corner to the bottom right corner) and 0's everywhere else. E.g.,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We recommend that you do a few examples to convince yourself that for every $m \times n$ matrix A and every $n \times p$ matrix B, we have $AI_n = A$ and $I_nB = B$. The matrix I_n is called the $n \times n$ identity matrix.

Definition 3.9 has a consequence that will play an important role when we discuss linear transformations:

Theorem 3.13: Columns of the product

Let A and B be any two matrices of the correct size so that AB is defined. Then each column of AB is a linear combination of the columns of A.

We give an example of an application of matrix multiplication:

Example 3.14: Systems of linear equations

Consider the system of equations:

$$2x+3y+5z=10$$
$$4x+y+7z=6$$

Letting

$$A=\begin{bmatrix}2&3&5\\4&1&7\end{bmatrix},\ X=\begin{bmatrix}x\\y\\z\end{bmatrix}\ \text{and } B=\begin{bmatrix}10\\6\end{bmatrix}$$

we get

$$AX=x\begin{bmatrix}2\\4\end{bmatrix}+y\begin{bmatrix}3\\1\end{bmatrix}+z\begin{bmatrix}5\\7\end{bmatrix}=\begin{bmatrix}2x+3y+5z\\4x+y+7z\end{bmatrix}$$

so the system of equations above can be expressed as

$$AX=B.$$

3.2.4. Properties of Matrix Multiplication

While matrix multiplication has a number of nice properties, it is not commutative. Depending on the size of the matrices, one of AB or BA might be defined and the other not. Moreover, even if both AB and BA are defined, they may not be equal.

Example 3.15: Matrix Multiplication is Not Commutative

Let $A=\begin{bmatrix}1&2\\3&4\end{bmatrix}$ *and* $B=\begin{bmatrix}0&1\\1&0\end{bmatrix}$. *Then*

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

and

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

Therefore, $AB\neq BA$.

While most matrices don't commute with each other, there are some pairs of matrices that do happen to commute. In particular, every $n\times n$ matrix A commutes with the identity matrix I_n since $AI_n=A=I_nA$ as in Example 3.12.

On the other hand, although it's not immediately obvious, matrix multiplication is associative (the fourth property in the proposition below). This property will be very important for us when we study linear transformations.

Proposition 3.16: Properties of Matrix Multiplication

The following properties hold for matrices A, B, and C and for scalars k: (By this we mean that if the matrices are the correct size so that one side of the equation is defined, then the other side makes sense as well and the two sides are equal.)

$$A(kB) = k(AB) = (kA)B \tag{3.7}$$

$$(A+B)C = AC+BC \tag{3.8}$$

$$A(B+C) = AB+AC \tag{3.9}$$

$$A(BC) = (AB)C \tag{3.10}$$

3.2.5. The Transpose

One final concept that we will use occasionally:

Definition 3.17: Transpose

The transpose of an $m \times n$ matrix A, denoted A^T, is the $n \times m$ matrix obtained by switching the rows with the columns.

For example,

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

3.2.6. Section Summary

- You can add matrices of the same size by adding their entries. The familiar properties of addition hold.
- To multiply a matrix by a scalar, multiply each entry by the scalar.
- To form the product AB, the number of columns of A must equal the number of rows of B.
- If B is a column matrix and A is a matrix of the correct size so that AB is defined, then the product AB is a column matrix formed by taking a linear combination of the columns of A. The entries of B are the coefficients of the linear combination.
- Matrix multiplication is not commutative.
- Properties of matrix multiplication that will be especially important for us later are associativity $(AB)C = A(BC)$, the distributive law $A(B+C) = AB+AC$ and the fact that scalars pull out: $A(kB) = kAB$.

- The transpose A^T is the matrix obtained by switching the rows of A with the columns.

Exercises

Exercise 3.2.1 *For the following pairs of matrices, determine if the sum $A+B$ is defined. If so, find the sum.*

(a) $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

(b) $A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 0 \end{bmatrix}, B = \begin{bmatrix} -1 & 0 & 3 \\ 0 & 1 & 4 \end{bmatrix}$

(c) $A = \begin{bmatrix} 1 & 0 \\ -2 & 3 \\ 4 & 2 \end{bmatrix}, B = \begin{bmatrix} 2 & 7 & -1 \\ 0 & 3 & 4 \end{bmatrix}$

Exercise 3.2.2 *For each matrix A, find the matrix $-A$ such that $A+(-A)=0$.*

(a) $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

(b) $A = \begin{bmatrix} -2 & 3 \\ 0 & 2 \end{bmatrix}$

(c) $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & -1 & 3 \\ 4 & 2 & 0 \end{bmatrix}$

Exercise 3.2.3 *For each matrix A, find $-2A, 0A$, and $3A$.*

(a) $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

(b) $A = \begin{bmatrix} -2 & 3 \\ 0 & 2 \end{bmatrix}$

(c) $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & -1 & 3 \\ 4 & 2 & 0 \end{bmatrix}$

Exercise 3.2.4 *Using only the properties given in Proposition 3.4 and Proposition 3.5 show $0A = 0$. Here the 0 on the left is the scalar 0 and the 0 on the right is the zero matrix of appropriate size.*

Exercise 3.2.5 *Using only the properties given in Proposition 3.4 and Proposition 3.5, as well as previous problems show $(-1)A = -A$.*

Exercise 3.2.6 *For each of the following, express the product AB as a linear combination of the columns of A and then compute it. (See Example 3.7.)*

(a) $A = \begin{bmatrix} 3 & 1 & 5 \\ 2 & 1 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$

(b) $A = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$

(c) $A = \begin{bmatrix} 3 & 1 \\ 4 & 2 \\ 7 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$

Exercise 3.2.7 *Consider the matrices* $A = \left[\begin{array}{rrr} 1 & 2 & 3 \\ 2 & 1 & 7 \end{array}\right]$, $B = \left[\begin{array}{rrr} 3 & -1 & 2 \\ -3 & 2 & 1 \end{array}\right]$, $C = \left[\begin{array}{rr} 1 & 2 \\ 3 & 1 \end{array}\right]$, $D = \left[\begin{array}{rr} -1 & 2 \\ 2 & -3 \end{array}\right]$, $E = \left[\begin{array}{r} 2 \\ 3 \end{array}\right]$.

Find the following if possible. If it is not possible explain why.

(a) $-3A$

(b) $3B - A$

(c) AC

(d) CB

(e) AE

(f) EA

Exercise 3.2.8 *Consider the matrices* $A = \left[\begin{array}{rr} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{array}\right]$, $B = \left[\begin{array}{rrr} 2 & -5 & 2 \\ -3 & 2 & 1 \end{array}\right]$, $C = \left[\begin{array}{rr} 1 & 2 \\ 5 & 0 \end{array}\right]$, $D = \left[\begin{array}{rr} -1 & 1 \\ 4 & -3 \end{array}\right]$, $E = \left[\begin{array}{r} 1 \\ 3 \end{array}\right]$

Find the following if possible. If it is not possible explain why.

(a) $-3A$

(b) $3B - A$

(c) AC

(d) CA

(e) AE

(f) EA

(g) BE

(h) DE

Exercise 3.2.9 *Let A, B and C be the matrices in Exercise 3.2.8.*

(a) *Compute* $(BA)C$ *and* $B(AC)$*, and check that the associative law is satisfied.*

(b) *Compute* $A(5C)$ *and* $5(AC)$ *and check that they are equal as asserted in Equation 3.7.*

Exercise 3.2.10 *Let* $A = \begin{bmatrix} 1 & 1 \\ -2 & -1 \\ 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 1 & -1 & -2 \\ 2 & 1 & -2 \end{bmatrix}$, *and* $C = \begin{bmatrix} 1 & 1 & -3 \\ -1 & 2 & 0 \\ -3 & -1 & 0 \end{bmatrix}$. *Find the following if possible.*

(a) AB

(b) BA

(c) AC

(d) CA

(e) CB

(f) BC

Exercise 3.2.11 *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *be a* 2×2 *matrix. Compute* $\det(A)$ *and* $\det(A^T)$*. Are they the same?*

Exercise 3.2.12 *Write the system*

$$\begin{aligned} x - 2y + 3z &= 5 \\ 2x + y + 7z &= 10 \\ x + 4y &= 12 \end{aligned}$$

in the form $AX = B$ *where A and B are appropriate matrices and* $X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$.

Exercise 3.2.13 *Let*

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 5 \end{bmatrix}$$

(a) *Let*

$$B = \begin{bmatrix} 4 \\ 5 \\ 11 \end{bmatrix}$$

Does there exist a column matrix $X = \begin{bmatrix} x \\ y \end{bmatrix}$ *such that* $AX = B$*? If so, find* X*. (You will need to solve a system of linear equations.)*

(b) *Express $\langle 4,5,11\rangle$ as a linear combination of $\langle 1,2,3\rangle$ and $\langle 2,1,5\rangle$. (You don't need to do a computation. Look at what you did in part (a)!)*

(c) *Let*

$$B = \begin{bmatrix} 4 \\ 5 \\ 12 \end{bmatrix}$$

Does there exist $X = \begin{bmatrix} x \\ y \end{bmatrix}$ *such that $AX = B$? Is $\langle 4,5,12\rangle$ in* $\mathrm{Span}(\langle 1,2,3\rangle, \langle 2,1,5\rangle$*?*

Exercise 3.2.14 *Let*

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

(a) *Compute* $A\begin{bmatrix} 3 \\ 2 \end{bmatrix}$.

(b) *Viewing the columns of A as vectors, are they linearly independent or linearly dependent? What is* $\mathrm{Span}(\langle 1,2,3\rangle, \langle 2,4,6\rangle)$*?*

(c) *Find all column matrices B for which the equation* $A\begin{bmatrix} x \\ y \end{bmatrix} = B$ *has a solution* $\begin{bmatrix} x \\ y \end{bmatrix}$.

Exercise 3.2.15 *Let* $A = \begin{bmatrix} -1 & -1 \\ 3 & 3 \end{bmatrix}$. *Find all 2×2 matrices, B such that $AB = 0$.*

Exercise 3.2.16 *Let* $X = \begin{bmatrix} -1 & -1 & 1 \end{bmatrix}$ *and* $Y = \begin{bmatrix} 0 & 1 & 2 \end{bmatrix}$. *Find $X^T Y$ and XY^T if possible.*

Exercise 3.2.17 *Let* $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 3 & k \end{bmatrix}$. *Is it possible to choose k such that $AB = BA$? If so, what should k equal?*

Exercise 3.2.18 *Let* $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 1 & 2 \\ 1 & k \end{bmatrix}$. *Is it possible to choose k such that $AB = BA$? If so, what should k equal?*

Exercise 3.2.19 *Find 2×2 matrices, A, B, and C such that $A \neq 0, C \neq B$, but $AC = AB$.*

Exercise 3.2.20 *Give an example of matrices (of any size), A, B, C such that $B \neq C$, $A \neq 0$, and yet $AB = AC$.*

Exercise 3.2.21 *Find 2×2 matrices A and B such that $A \neq 0$ and $B \neq 0$ but $AB = 0$.*

Exercise 3.2.22 *Give an example of matrices (of any size), A, B such that $A \neq 0$ and $B \neq 0$ but $AB = 0$.*

Exercise 3.2.23 *Find 2×2 matrices A and B such that $A \neq 0$ and $B \neq 0$ with $AB \neq BA$.*

Exercise 3.2.24 *A matrix A is called* idempotent *if* $A^2 = A$. *Let*

$$A = \begin{bmatrix} 2 & 0 & 2 \\ 1 & 1 & 2 \\ -1 & 0 & -1 \end{bmatrix}$$

and show that A is idempotent .

Exercise 3.2.25 *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *and* $B = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$.

(a) *Compute the product* $\mathrm{Det}(A)\,\mathrm{Det}(B)$

(b) *Compute* $\mathrm{Det}(AB)$.

(c) *Compare your answers to (a) and (b).*

4. Directional Derivatives and Differentiability

4.1 Directional Derivatives of Real-valued Functions

Motivation

The partial derivatives f_x and f_y of $f(x,y)$ at (x_0,y_0) tell us the rate of change of f at (x_0,y_0) as you move parallel to the x or the y axis, respectively. How do you determine the rate of change of f as you move in a different direction. For example, what is the rate of change of f as you move along a line such as $y=x$?

Outcomes

- *Understand the definition of directional derivatives and be able to compute directional derivatives using this definition.*
- *Be able to find the tangent line to a curve given by the intersection of a vertical plane with the graph $z=f(x,y)$ of a function.*

4.1.1. Definition and examples

Suppose that the temperature at a point (x,y) on a flat metal plate is given by

$$f(x,y)=3x^2y$$

where x and y are measured in centimeters and the temperature is in degrees celsius. At, say, the point $(x_0,y_0)=(1,2)$, we can compute that $f_x(1,2)=12$ and $f_y(1,2)=3$. This tells us that as we move in the x, respectively y, directions from $(1,2)$, the instantaneous rate of change of temperature is 12 deg/cm, respectively 3 deg/cm.

How do we measure the rate of change of the temperature as we move in other directions? We can specify the direction we are interested in by a unit vector, e.g., $\mathbf{u}=\langle \frac{3}{5},\frac{4}{5}\rangle$. Imagine moving a temperature probe at unit speed (one cm per second) along a straight line ℓ in the direction $\mathbf{u}$ passing through $(1,2)$ at time zero. This line is given by

$$\ell: \quad x=1+\tfrac{3}{5}t,\ \ y=2+\tfrac{4}{5}t.$$

Letting $g(t)$ be the temperature measured by the probe at time t, we have

$$g(t)=f(1+\tfrac{3}{5}t,2+\tfrac{4}{5}t)=3(1+\tfrac{3}{5}t)^2(2+\tfrac{4}{5}t).$$

The rate of change of temperature at time $t=0$ as measured by the probe is $g'(0)$. We leave it to the reader to compute that

$$g'(0) = \tfrac{48}{5}.$$

Thus the temperature probe is measuring a rate of change of $\frac{48}{5}$ deg/sec. Since the probe is moving at unit speed (i.e., the probe is t centimeters from the point $(1,2)$ at time t seconds), we conclude that:
The instantaneous rate of change of temperature at the point $(1,2)$ in the direction $\mathbf{u} = \langle \frac{3}{5}, \frac{4}{5} \rangle$ is $\frac{48}{5}$ deg/cm.

We emphasize the importance of using a unit vector $\mathbf{u}$, so that the probe is moving at unit speed. If the probe moved faster, the rate of change of temperature with respect to time as measured by the probe would be greater that the rate of change of temperature with respect to distance on ℓ; the latter rate of change is the one we want.

The example above motivates the following definition:

Definition 4.1: Directional Derivatives

Let $f : D \to \mathbb{R}$ be a real-valued function of two variables. For (x_0, y_0) a point in the domain D and $\mathbf{u} = \langle u_1, u_2 \rangle$ any unit vector, we define the directional derivative *of f at (x_0, y_0) in the direction $\mathbf{u}$ by*

$$D_{\mathbf{u}} f(x_0, y_0) = \frac{d}{dt}_{|t=0} f(x_0 + tu_1, y_0 + tu_2).$$

We can also write this as $D_{\mathbf{u}} f(x_0, y_0) = g'(0)$ where $g(t) = f(x_0 + tu_1, y_0 + tu_2)$.

Example 4.2

Let $f(x,y) = e^{xy}$. Find the directional derivative of f at the point $P(2,0)$ in the direction from P towards $Q(5,-4)$.

Solution. We first need the unit vector $\mathbf{u}$ in the direction of the displacement vector $\mathbf{PQ} = \langle 3, -4 \rangle$. Since the displacement vector has length 5, we multiply by 1/5 to obtain $\mathbf{u} = \langle \frac{3}{5}, -\frac{4}{5} \rangle$. Thus:

$$D_{\mathbf{u}} f(2,0) = \tfrac{d}{dt}_{|t=0} f(2 + \tfrac{3}{5}t, 0 - \tfrac{4}{5}t) = \frac{d}{dt}_{|t=0} e^{(2+\frac{3t}{5})(-\frac{4t}{5})} = -\frac{8}{5}$$

(We leave it to the reader to check the last equality.)

♠

4.1.2. Comparing directional derivatives and partial derivatives

You have learned that the partial derivative $f_x(x_0, y_0)$ tells you the rate of change of f at (x_0, y_0) as you head in the x-direction, i.e., in the direction of the vector $\langle 1, 0 \rangle$. But that's exactly the information conveyed by the directional derivative $D_{\langle 1,0 \rangle} f(x_0, y_0)$. In fact we have:

Proposition 4.3

$$f_x(x_0, y_0) = D_{\langle 1,0 \rangle} f(x_0, y_0) \;\text{ and }\; f_y(x_0, y_0) = D_{\langle 0,1 \rangle} f(x_0, y_0)$$

Proof. Let $g(x) = f(x, y_0)$. Then

$$f_x(x_0, y_0) = g'(x_0).$$

Next $D_{\langle 1,0\rangle} f(x_0, y_0) = \frac{d}{dt}_{|t=0} f(x_0 + t, y_0) = \frac{d}{dt}|_{t=0}\, g(x_0 + t)$. Thus by the chain rule,

$$D_{\langle 1,0\rangle} f(x_0, y_0) = g'(x_0 + 0)\frac{d}{dt}|_{t=0}\,(x_0 + t) = g'(x_0)(1) = g'(x_0).$$

Comparing the two displayed formulas, we obtain the first statement of the proposition. The second one is similar. ♠

4.1.3. Tangent lines to mountain paths

Let $f : D \to \mathbb{R}$ be a real-valued function of two variables and consider the graph $z = f(x, y)$. View the graph as the surface of a mountain (or valley) by thinking of the positive x-axis as pointing east, the positive y axis as pointing north and thinking of z as elevation. If you walk due east or due north on the mountain starting from the point $(x_0, y_0, f(x_0, y_0))$, then the partial derivative $f_x(x_0, y_0)$ tells you how steep your path is on the mountain since f_x is the rate of change of $f(x, y)$ (elevation) with respect to x (your east-west position). Similarly:

- $D_{\mathbf{u}} f(x_0, y_0)$ tells you the slope of your mountain path at the point $(x_0, y_0, f(x_0, y_0))$ if you walk in the direction specified by the unit vector $\mathbf{u}$.

Figure 4.1 shows part of the graph of a function $z = f(x, y)$, a point $(x_0, y_0) = (a, b)$ in the xy plane and a unit vector $\mathbf{u}$ positioned at P. (P is expressed as $(a, b, 0)$, since the xy plane is viewed as sitting in $\mathbb{R}^3$ in the illustration.) The tan-colored plane is the vertical plane that contains the line in the xy plane through P with direction vector $\mathbf{u}$. This plane cuts the graph f in a curve through the point $Q = (a, b, f(a, b))$. This curve is the mountain path referred to above. As you walk in the direction specified by $\mathbf{u}$, it appears that you are going downhill. Thus $D_{\mathbf{u}} f(a, b) < 0$ and the value of $D_{\mathbf{u}} f(a, b)$ tells you how steeply you are descending. Also shown in the picture is the tangent line to your mountain path through Q. We will see shortly how to find this tangent line.

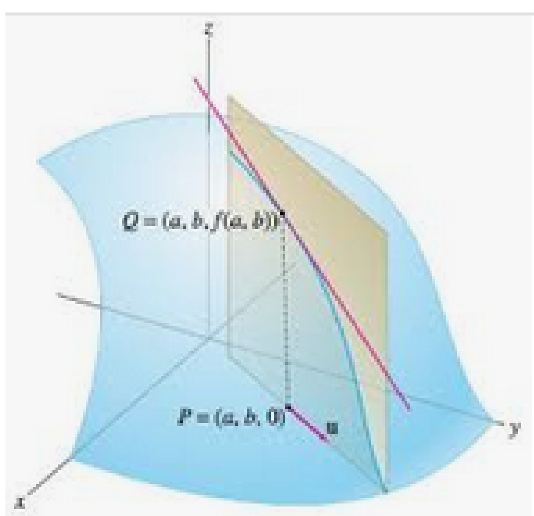


Figure 4.1

Example 4.4

Let $f(x,y) = x^3 e^{5y}$. Viewing the positive x-axis as pointing east, the positive y axis as pointing north and thinking of z as elevation, suppose you walk northwest on the "mountain" (the graph of f), starting from the point $(1,0,f(1,0)) = (1,0,1)$. Are you ascending or descending? At what rate?

Solution. We first need the unit vector $\mathbf{u}$ in $\mathbb{R}^2$ that points northwest. As in Figure 4.2, we see that

$$\mathbf{u} = \langle -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \rangle.$$

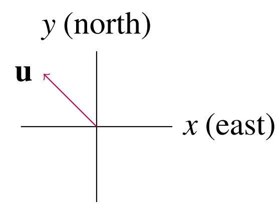


Figure 4.2

We can now compute the slope of this mountain path:

$$D_{\langle -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \rangle} f(1,0) = \frac{d}{dt}_{|t=0} (1 - \tfrac{1}{\sqrt{2}} t)^3 e^{5t/\sqrt{2}} = \sqrt{2}.$$

(We leave it to the reader to check the last equality.) Thus you are going uphill at a slope of $\sqrt{2}$.

♠

Returning to the general situation in Figure 4.1, let's find the vector equation

$$\mathbf{r}(t) = \langle x(t), y(t), z(t) \rangle \tag{4.1}$$

for the path C that is given as the intersection of the graph $z = f(x,y)$ with the tan-colored plane.

Write the vector $\mathbf{u}$ as $\mathbf{u} = \langle u_1, u_2 \rangle$. Since the path C lies directly above the line in the xy-plane through (x_0, y_0) in the direction $\mathbf{u}$, we have

$$x(t) = x_0 + u_1 t, \;\; y(t) = y_0 + u_2 t. \tag{4.2}$$

(Again in the picture, x_0 is denoted a and y_0 is denoted b.) Since the curve lies on the graph of f, we must have

$$z(t) = f(x(t), y(t)) = f(x_0 + u_1 t, \; y_0 + u_2 t) \tag{4.3}$$

Thus the path is given by

$$\boxed{\mathbf{r}(t) = \langle x_0 + u_1 t, y_0 + u_2 t, f(x_0 + u_1 t, y_0 + u_2 t) \rangle.}$$

We next find the tangent line to this path at the point $(x_0, y_0, f(x_0, y_0))$. The tangent line is drawn in red in Figure 4.1. The first two components of the tangent vector $\mathbf{r}'(0)$ are easy to compute. For the third, note that the z-component $z(t) = f(x_0 + u_1 t,\ y_0 + u_2 t)$ of $\mathbf{r}(t)$ is precisely the function that we differentiate to obtain $D_{\mathbf{u}} f(x_0, y_0)$. (This function was called $g(t)$ in the Definition 4.1 of directional derivative.) Thus

$$\boxed{\mathbf{r}'(0) = \langle u_1, u_2, D_{\mathbf{u}} f(x_0, y_0)\rangle.}$$

This gives us the tangent vector. The point of tangency is $(x_0, y_0, f(x_0, y_0))$. The reader can now write down the tangent line.

Example 4.5

Let C be the mountain path in Example 4.4. Let's find the tangent line to our northwest mountain path at the point $(1,0,f(1,0)) = (1,0,1)$. We have $\mathbf{u} = \langle -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\rangle$, and we found that $D_{\mathbf{u}} f(1,0) = \sqrt{2}$, so the tangent vector to our path at the point $(1,0,1)$ is given by

$$\langle -\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, \sqrt{2}\rangle.$$

The tangent line is given by

$$\mathbf{T}(t) = \langle 1,0,1\rangle + t\langle -\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, \sqrt{2}\rangle.$$

(To get a geometric sense of this equation, note that as t increases by one unit, the horizontal change (meaning the change in x, y) on the tangent line is given by $\langle -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\rangle$, a distance of one unit in the northwest direction. Meanwhile, the elevation z changes by $\sqrt{2}$. Thus the (change in elevation)/(change in horizontal position) is $\frac{\sqrt{2}}{1} = \sqrt{2}$. This is exactly what we mean when we say that the slope heading northwest from the point $(1,0,1)$ is $\sqrt{2}$.

4.1.4. An intriguing example

Example 4.6

Let

$$f(x,y) = x^{1/3} y^{2/3}.$$

Compute $D_{\mathbf{u}} f(0,0)$ where $\mathbf{u} = \langle 1,0\rangle$.

Solution. $D_{\mathbf{u}} f(0,0) = g'(0)$ where $g(t) = f(0+t, 0) = t^{1/3}(0) = 0$. Since $g(t) = 0$ for all t, we have $g'(0) = 0$, so $D_{\mathbf{u}} f(0,0) = 0$.

♠

What's strange about this example? Recall Proposition 4.3: We know that $D_{\langle 1,0\rangle} f(0,0) = f_x(0,0)$. Thus we've shown that $f_x(0,0) = 0$. You might at first think that this is a mistake, since you may want to compute that

$$f_x(x,y) = \tfrac{1}{3} x^{-2/3} y^{2/3} \tag{4.4}$$

and then be concerned by the fact that $x^{-2/3} y^{2/3}$ is of the form $\infty \cdot 0$ when $(x,y) = (0,0)$, which is an undefined quantity. The mystery goes away, however, if you remember that $f_x(0,0) = h'(0)$ where $h(x) = f(x,0)$. (See Equation (1) on page 953 of Stewart.) Since $f(x,0) \equiv 0$, we again get $f_x(0,0) = 0$.

What's going on here is that the expression $f_x(x,y) = \frac{1}{3}x^{-2/3}y^{2/3}$, which is valid when $x \neq 0$, does not have a limit as $x \to 0$. The fact that the limit doesn't exist says that f_x isn't continuous at the origin $(0,0)$. Nevertheless, $f_x(0,0)$ does exist, we computed it! We just had to be sure to plug in $y_0 = 0$ (and to note that $f(x,0) \equiv 0$) before differentiating with respect to x rather than after. For nice functions with continuous partials, it doesn't matter whether you plug in the value y_0 for y before or after differentiating with respect to x, but it does matter when the partials aren't continuous.

Aside: When $y_0 \neq 0$, we *do* run into a problem with $f_x(0,y_0)$. For example, if $y_0 = 1$, then to compute $f_x(0,1)$, we set $h(x) = f(x,1) = x^{1/3}$. Since $h'(0)$ doesn't exist, the partial $f_x(0,1)$ doesn't exist. The only point on the y axis where f_x does exist is the origin, as computed above.

Figure 4.3 shows the graph of the function in Example 4.6. The red dotted line is the x-axis and lies on the graph of the function. When you head east from the origin, you are simply walking on the x-axis, going neither up nor down and so your slope $f_x(0,0)$ is zero. The y axis is the horizontal green line that separates the light green region from the steep gray region. If you move slightly away from the x-axis to (x,y_0), then (no matter how small y_0 is), you find yourself on a cliff when you pass through $(0,y_0)$.

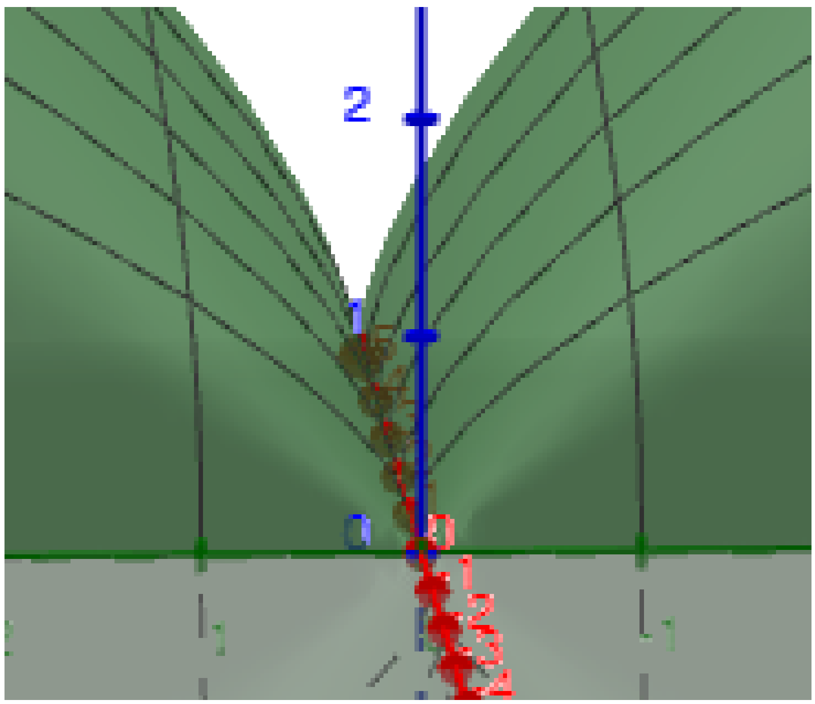


Figure 4.3

4.1.5. Directional derivatives in higher dimensions

For real-valued functions of three variables, directional derivatives are defined analogously to Definition 4.1.

Example 4.7

Let $f(x,y,z) = xyz^2$. Compute $D_{\mathbf{u}} f(2,1,1)$ where $\mathbf{u}$ is the unit vector $\langle \frac{1}{9}, -\frac{2}{9}, \frac{2}{9} \rangle$.

Solution.

$$D_{\mathbf{u}} f(2,1,1) = \frac{d}{dt}_{|t=0} f\left(2+\frac{t}{9}, 1-\frac{2t}{9}, 1+\frac{2t}{9}\right) = \frac{d}{dt}_{|t=0} \left(2+\frac{t}{9}\right)\left(1-\frac{2t}{9}\right)\left(1+\frac{2t}{9}\right)^2 = \frac{4}{9}$$

♠

4.1.6. Section summary

- The directional derivative $D_{\mathbf{u}} f(x_0, y_0)$ tells you the rate of change of f in the direction of the unit vector $\mathbf{u}$.
- Definition 4.1 tells you how to compute directional derivatives.
- It is essential that $\mathbf{u}$ be a unit vector. If you want to find the directional derivative in a direction specified, say, by a displacement vector $\mathbf{PQ}$, first find a unit vector in the specified direction.
- The partial derivatives f_x and f_y are the directional derivatives $D_{\mathbf{u}} f$ corresponding to $\mathbf{u} = \langle 1, 0\rangle$ and $\mathbf{u} = \langle 0, 1\rangle$, respectively.
- The tangent vector at $(x_0, y_0, f(x_0, y_0))$ to the curve in the graph $z = f(x,y)$ that lies above the line $\langle x, y\rangle = \langle x_0, y_0\rangle + t\mathbf{u}$ in the xy-plane is given by $\langle u_1, u_2, D_{\mathbf{u}} f(x_0, y_0)\rangle$ where $\mathbf{u} = \langle u_1, u_2\rangle$.

Exercises

Exercise 4.1.1 *For each of the following, compute the directional derivative $D_{\mathbf{u}} f(x_0, y_0)$.*

(a) $f(x,y) = x^2 \ln(y)$, $(x_0, y_0) = (2,1)$, $\mathbf{u} = \langle -\frac{4}{5}, \frac{3}{5}\rangle$.

(b) $f(x,y) = e^{2x-y}$, $(x_0, y_0) = (1,2)$, $\mathbf{u} = \langle \frac{4}{5}, -\frac{3}{5}\rangle$.

(c) $f(x,y) = x^2 \ln(y)$, $(x_0, y_0) = (2,1)$, $\mathbf{u} = \langle -\frac{1}{2}, \frac{\sqrt{3}}{2}\rangle$.

(d) $f(x,y) = x\sqrt{x+y}$, $(x_0, y_0) = (1,3)$, $\mathbf{u} = \langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\rangle$.

Exercise 4.1.2 *For each of the following, compute the directional derivative of f at the point P in the direction from P to Q.*

(a) $f(x,y) = \sin(\pi xy)$, $P = (2,0)$, $Q = (4,2)$

(b) $f(x,y) = (x^2 - y)^3$, $P = (1,2)$, $Q = (3,5)$.

(c) $f(x,y) = \ln(x^2 + y^2)$, $P = (1,1)$, $Q = (3,1)$.

Exercise 4.1.3 *For each of the functions in Exercise 4.1.1, find the tangent vector at* $(x_0, y_0, f(x_0, y_0))$ *to the curve in the graph* $z = f(x,y)$ *that lies above the line* $\langle x, y\rangle = \langle x_0, y_0\rangle + t\mathbf{u}$.

Exercise 4.1.4 *For each of the following, compute* $D_{\mathbf{u}} f(x_0, y_0, z_0)$.

(a) $f(x,y,z) = x^2 e^{yz}$, $(x_0, y_0, z_0) = (2,2,0)$, $\mathbf{u} = \langle \frac{2}{3}, \frac{2}{3}, \frac{1}{3}\rangle$.

(b) $f(x,y,z) = \tan(xy - z)$, $(x_0, y_0, z_0) = (2,1,2)$, $\mathbf{u} = \langle \frac{3}{5}, 0, \frac{4}{5}\rangle$.

Exercise 4.1.5 *(This exercise illustrates Proposition 4.3.)*

Let $f(x,y) = y\sqrt{x^2+y^2}$. *Compute* $f_x(3,4)$ *and* $D_{\langle 1,0\rangle} f(3,4)$ *and check that they agree. Similarly compare* $f_y(3,4)$ *and* $D_{\langle 0,1\rangle} f(3,4)$.

Exercise 4.1.6 *Let the x-axis point east, let the y axis point north, and let z denote elevation. Let* $f(x,y) = x^2y + xy^3$, *and view the graph* $z = f(x,y)$ *as a mountain. Suppose that you are hiking northeast on this mountain passing through the point* $(2,1,6)$. *Find the tangent line to your path at this point.*

Exercise 4.1.7 *Answer the same question as in Exercise 4.1.6, where now* $f(x,y) = x\sqrt{y}$, *you are hiking due south, and you are passing through the point* $(2,1,2)$.

Exercise 4.1.8 *Let* f *be the function in Example 4.6. Compute* $D_{\mathbf{u}} f(0,0)$ *for each of the following unit vectors* $\mathbf{u}$.

(Note: All these directional derivatives do exist. After you write down the expression that you need to differentiate, be sure to simplify it before taking the derivative. Otherwise you will not be able to carry out the computation.)

(a) $\mathbf{u} = \langle \frac{3}{5}, \frac{4}{5}\rangle$

(b) $\mathbf{u} = \langle -\frac{3}{5}, -\frac{4}{5}\rangle$

(c) $\mathbf{u} = \langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\rangle$

4.2 Differentiable functions and tangent planes

Motivation

In single variable calculus, you learned that a function $f : \mathbb{R} \to \mathbb{R}$ *is differentiable at a point* x_0 *if the limit*

$$\lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h}$$

exists. The derivative $f'(x_0)$ *was then defined to be this limit.*
Geometrically, you then learned that differentiability of f *at* x_0 *means that the graph of* f *has a tangent line at* $(x_0, f(x_0))$ *of slope* $f'(x_0)$. *The tangent line is given by*

$$y - f(x_0) = f'(x_0)(x - x_0).$$

In this section, we will give an informal geometric approach to defining the concept of differentiability of a real-valued function $f(x,y)$ *of two variables by introducing the notion of tangent planes.*

Outcomes

- *Gain a geometric understanding of what it means for a plane to be tangent to a surface.*
- *Understand the concept of differentiability.*
- *Learn a sufficient condition for testing whether a function is differentiable.*
- *Be able to find the tangent plane of a differentiable function.*
- *Be able to use the tangent approximation of a differentiable function to estimate values of the function, and be able to carry this out in applications.*

4.2.1. Tangent Planes to Surfaces

Before talking about tangent planes, let's briefly review what we know about tangent lines to curves. You first saw tangent lines in single variable calculus when you looked at tangent lines to graphs $y = f(x)$ of differentiable functions. The derivative of the function gave you the slope. The tangent lines give you a way of approximating complicated functions near a given point.

We've also seen tangent lines to smooth parametrized curves in $\mathbb{R}^2$ and $\mathbb{R}^3$. In Figure 4.4, we illustrate the tangent line to the circle $\mathbf{r}(t) = \langle \cos(t), \sin(t) \rangle$ at the point (1,0). This particular tangent line is a vertical line, i.e., it is parallel to the y-axis. In contrast, as you learned in calculus, tangent lines to graphs of differentiable functions $f(x)$ are never vertical lines since the slope $f'(x_0)$ can't be infinite.

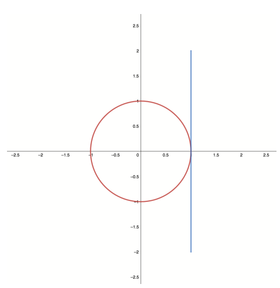


Figure 4.4

The function $f(x) = |x|$, whose graph is shown in Figure 4.5 is a familiar example of a function that does not have a tangent line at the point $x_0 = 0$. There are two rays competing to be tangent at the origin: the ray $y = x$ for $x \geq 0$ and the ray $y = -x$ for $x \leq 0$, but they don't match up.

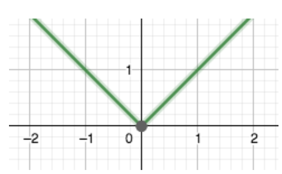


Figure 4.5

When we want to approximate values on a surface rather than a curve, it is natural to ask for a tangent plane. In figure 4.6 we show a picture of the sphere $x^2 + y^2 + z^2 = 1$ and its tangent plane at the north pole

$(0,0,1)$. This plane is horizontal; its equation is $z = 1$. If you imagined walking on the sphere along any smooth path $\mathbf{r}(t) = \langle x(t), y(t), z(t)\rangle$ that passes through the north pole, you would reach your maximum elevation at the moment t_0 when you reached the north pole. Thus $z'(t_0)$ would be zero and so your tangent line would be of the form $\langle x,y,z\rangle = \langle 0,0,1\rangle + t\langle x'(t_0), y'(t_0), 0\rangle$. Since the last coordinate of the tangent vector is 0, the tangent line would be parallel to the xy-plane. Thus every such tangent line lies on the tangent plane $z = 1$. All the different tangent lines fit together to form the tangent plane.

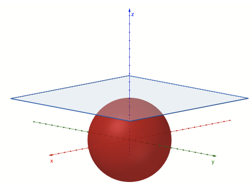


Figure 4.6

The sphere example suggests the following definition: [1]

Definition 4.8: Informal Definition of Tangent Plane to a Surface

A plane $\mathscr{P}$ is tangent to a surface S at a point p if the tangent line at p to **every** *smooth curve in S through p lies in the plane $\mathscr{P}$.*

The condition in Definition 4.8 is extremely demanding. Given a point p on a surface S, there will be infinitely many curves on S passing through p. We are asking that the tangent lines to *all* these curves be coplanar. We know that any two such lines that aren't parallel uniquely determine a plane through p. For all the rest of them to lie on the same plane seems almost too much to ask!

Figure 4.7 shows the double cone S whose equation is $z^2 = x^2 + y^2$. For our point p, let's take the origin $(0,0,0)$, which is the vertex of the cone. The entire surface of the cone is filled up by straight lines passing through the origin. We illustrate 3 of these lines in Figure 4.7. Each of these lines is tangent to a smooth curve (namely the line itself!) in S through the origin. They are not all coplanar, i.e., there's no plane through the origin containing all of them. By Definition 4.8, the surface S does not have a tangent plane at the origin.

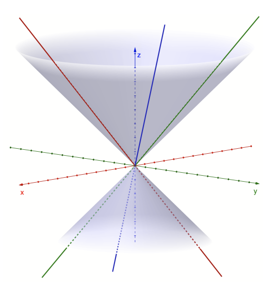


Figure 4.7

[1] To fully define the notion of tangent plane to a surface, one needs a condition to guarantee that there are plenty of smooth curves on the surface to begin with. That's why we are calling Definition 4.8 an "informal" definition.

4.2.2. Tangent Planes to Graphs of Functions

For real-valued functions f of one variable that are differentiable at a point x_0, we know that the graph $y = f(x)$ has a tangent line at $(x_0, f(x_0))$, and the tangent line is not a vertical line, meaning that it is not parallel to the y-axis. The slope of the tangent line is then the derivative. We will use an analogous property to define differentiability of real-valued functions of more than one variable.

Let $z = f(x,y)$ be the graph of a real-valued function of two variables. The function f may be very difficult to evaluate so we would like to be able to get a good approximation near any point (x_0, y_0) in its domain. The graph $z = f(x,y)$ is a surface, so it makes sense to ask whether it has a tangent plane at $(x_0, y_0, f(x_0, y_0))$ and, if so, to try to find an explicit equation for the tangent plane.

Definition 4.9: Definition of differentiability

We say that f is differentiable at (x_0, y_0) if the graph $z = f(x,y)$ has a tangent plane at $\mathscr{P}$ at $(x_0, y_0, f(x_0, y_0))$ and the tangent plane is not vertical, i.e., it is not parallel to the z-axis.

Recall that every plane has an equation of the form $ax + by + cz = d$. The vertical planes – the ones we're avoiding– are those for which $c = 0$. (To see this, note that if a plane $ax + by + cz = d$ is parallel to the z-axis, then the vector $\langle 0, 0, 1 \rangle$ is parallel to the plane and thus orthogonal to the normal vector $\langle a, b, c \rangle$. We then have $\langle a, b, c \rangle \cdot \langle 0, 0, 1 \rangle = 0$, so $c = 0$.)

How does one check whether the graph of a function has a tangent plane and thus that the function is differentiable? Happily, the following result, which is proven in more advanced courses, makes it easy to check that many functions have tangent planes:

Theorem 4.10

If the first order partial derivatives of f exist and are continuous at (x_0, y_0), then f is differentiable at (x_0, y_0), i.e., the graph $z = f(x,y)$ has a (non-vertical) tangent plane at the point $(x_0, y_0, f(x_0, y_0))$.

(Aside: The theorem doesn't give us any information about functions whose partial derivatives aren't continuous at a point (x_0, y_0). In some, but not all such cases, the function actually is differentiable but it can be hard to check.)

As an example of Theorem 4.10, the function $f(x,y) = 4 - x^2 - y^2$ has continuous partial derivatives and thus is differentiable at every point. Figure 4.8 illustrates the tangent plane at the highlighted point on its graph. You can picture what the tangent planes would look like at other points.

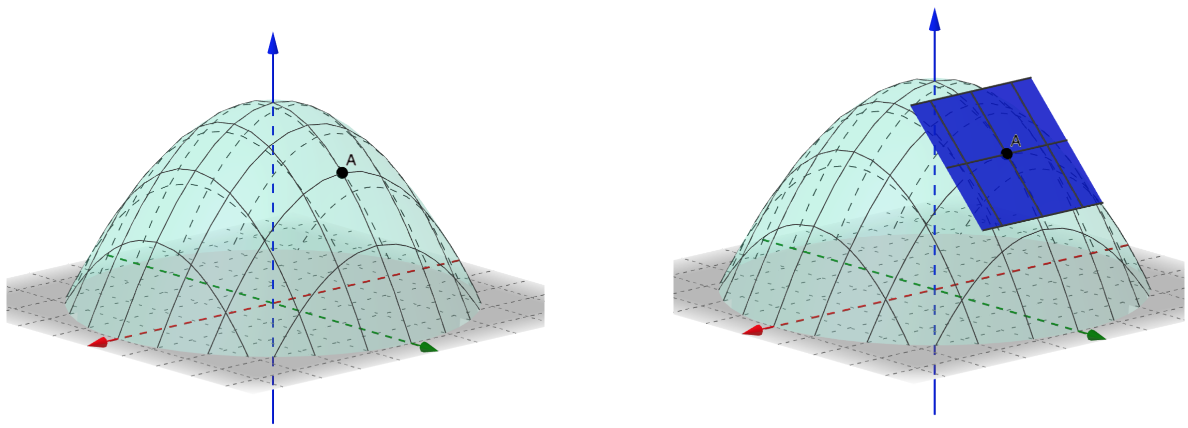


Figure 4.8

Once we know a tangent plane exists, how do we find an explicit equation for the plane? Theorem 4.10 tells us that the tangent plane must contain the tangent line to every smooth curve in the graph through $(x_0, y_0, f(x_0, y_0))$. We know lots of curves in the graph! Here's some:

Let $\mathbf{u} = \langle u_1, u_2 \rangle$ be any unit vector. Then, as in Subsection 4.1.3, the curve

$$\mathbf{r}(t) = \langle x_0 + tu_1, y_0 + tu_2, f(x_0 + tu_1, y_0 + tu_2) \rangle$$

passes through the point $(x_0, y_0, f(x_0, y_0))$ at time $t = 0$ and has tangent line

$$\langle x, y, z \rangle = (x_0, y_0, f(x_0, y_0)) + t\langle u_1, u_2, D_{\mathbf{u}} f(x_0, y_0) \rangle.$$

By Definitions 4.8 and 4.9, this line must lie on the tangent plane through $(x_0, y_0, f(x_0, y_0))$. Thus:

Proposition 4.11

For every unit vector $\mathbf{u} = \langle u_1, u_2 \rangle$, the vector

$$\langle u_1, u_2, D_{\mathbf{u}} f(x_0, y_0) \rangle$$

is parallel to the tangent plane to the graph $z = f(x, y)$ through $(x_0, y_0, f(x_0, y_0))$.

We only need two of these vectors (as long as we choose ones that aren't parallel) to obtain the equation of the plane. We get:

Theorem 4.12

If the graph $z = f(x, y)$ has a tangent plane at $(x_0, y_0, f(x_0, y_0))$, then the tangent plane is given by

$$z - z_0 = a(x - x_0) + b(y - y_0)$$

where $z_0 = f(x_0, y_0)$ and

$$a = f_x(x_0, y_0) \ \text{ and } \ b = f_y(x_0, y_0).$$

Proof. To find a normal vector to the plane, we need two linearly independent vectors that are parallel to $\mathscr{P}$. Proposition 4.11 gives us many choices. Let's use the ones corresponding to the two partials:

When $\mathbf{u} = \langle 1,0 \rangle$, the directional derivative $D_{\mathbf{u}}$ equals the partial derivative f_x. (See Proposition 4.3.) Thus Proposiiton 4.11 tells us that $\langle 1,0,f_x(x_0,y_0)\rangle$ is parallel to the tangent plane. Similarly, taking $\mathbf{u} = \langle 0,1 \rangle$, we see that $\langle 0,1,f_y(x_0,y_0)\rangle$ is also parallel to the tangent plane. Taking the cross product of these two vectors we obtain the normal vector

$$\mathbf{n} = \langle -f_x(x_0,y_0),\, -f_y(x_0,y_0),\, 1\rangle.$$

(The reader should verify the computation of the cross product.) Since we know that the tangent plane contains the point $(x_0,y_0,f(x_0,y_0))$, we now have both a point on the plane and the normal vector so we can write down the equation of the tangent plane:

$$-f_x(x_0,y_0)(x-x_0) - f_y(x_0,y_0)(y-y_0) + (z-z_0) = 0$$

or

$$\boxed{z - z_0 = f_x(x_0,y_0)(x-x_0) \,+\, f_y(x_0,y_0)(y-y_0).}$$

♠

Example 4.13

Let

$$f(x,y) = x^2y^3, \;\; (x_0,y_0) = (2,1).$$

We have $\frac{\partial f}{\partial x}(x_0,y_0) = 4$, $\frac{\partial f}{\partial y}(x_0,y_0) = 12$ *and* $f(x_0,y_0) = 4$. *The partials are polynomials, hence are continuous, so we know by Theorem 4.10 that* f *is differentiable and thus has a tangent plane. By Theorem 4.12, the tangent plane to the graph of* f *at (2,1,4) is given by*

$$z - 4 = 4(x-2) + 12(y-1).$$

We can also express the tangent plane in Example 4.13 as $z = L(x,y)$ where

$$L(x,y) = 4 + 4(x-2) + 12(y-1). \tag{4.5}$$

Definition 4.14: Tangent Approximation to the Graph

If f *is differentiable at* (x_0,y_0), *we define the tangent approximation to* f *near* (x_0,y_0) *to be the function*

$$L(x,y) = f(x_0,y_0) + f_x(x_0,y_0)\,(x-x_0) \,+\, f_y(x_0,y_0)\,(y-y_0).$$

Observe that the tangent plane to the graph of f *at* $(x_0,y_0,f(x_0,y_0))$ *can then be written as*

$$z = L(x,y).$$

Example 4.15

Use the tangent approximation of an appropriate function in order to estimate the value of $\sqrt{2.9^2+4.1^2}$.

Solution. The value should be close to $\sqrt{3^2+4^2}=5$. The natural function to consider here is

$$f(x,y)=\sqrt{x^2+y^2}.$$

Since $(2.9,4.1)$ is close to $(3,4)$, the tangent approximation to f near $(3,4)$ should give us a reasonable estimate of $f(2.9,4.1)$. A computation gives:

$$f_x=\frac{x}{\sqrt{x^2+y^2}} \text{ and } f_y=\frac{y}{\sqrt{x^2+y^2}}.$$

Both partials are continuous at $(3,4)$ so the tangent plane and thus the tangent approximation exist (see Theorem 4.10). Since $f_x(3,4)=\frac{3}{5}=0.6$ and $f_y(3,4)=\frac{4}{5}=0.8$, we obtain the tangent plane

$$z-5=0.6(x-3)+0.8(y-4)$$

and the tangent approximation to f:

$$L(x,y)=5+0.6(x-3)+0.8(y-4).$$

Thus

$$f(2.9,4.1)\sim L(2.9,4.1)=5+0.6(-0.1)+0.8(0.1)=5.02.$$

(Aside: The tangent approximation gave us two digits to the right of the decimal place, but it is unlikely that the approximation is actually valid to that many digits. There are methods available to determine the number of digits of accuracy. However, we will not address this question in this course.)

♠

Notation 4.16

Assume that $f(x,y)$ is differentiable at (x_0,y_0) and let $z_0=f(x_0,y_0)$. Let (x,y) be close to (x_0,y_0) and write

$$\Delta x=x-x_0 \text{ and } \Delta y=y-y_0.$$

Thus Δx and Δy denote small changes in x and y. The resulting change in $z=f(x,y)$ is denoted Δz and is given by

$$\Delta z=f(x,y)-f(x_0,y_0).$$

We emphasize that you have control over Δx and Δy, but the resulting Δz is determined by the function f.

The tangent plane approximation tells us exactly how the value of z on the tangent plane changes when x and y change by Δx and Δy, namely

$$z-z_0=f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0)=f_x(x_0,y_0)\Delta x+f_y(x_0,y_0)\Delta y.$$

When Δx and Δy are small, the tangent approximation is a good approximation of the function. Thus the change Δz of the function is approximated by:

$$\boxed{\Delta z\sim f_x(x_0,y_0)\Delta x+f_y(x_0,y_0)\Delta y.} \tag{4.6}$$

(The notation $\sim$ means "approximately equal to".)

Example 4.17

A closed conical storage tank has radius four feet and height 10 feet. Estimate how much storage volume would be lost if insulation is added to the inside of the tank, resulting in a decrease of the radius by 3 inches and of the height by 6 inches. (Recall that the volume of a cone of radius r and height h is $\frac{1}{3}\pi r^2 h$.)

Solution. We have $r_0 = 4$ feet, $h_0 = 10$ feet, $\Delta r = -\frac{1}{4}$ foot and $\Delta h = -\frac{1}{2}$ foot. Since our independent variables are r and h and our dependent variable is V, Equation (4.6) says that

$$\Delta V \sim \frac{\partial V}{\partial r}(4,10)\Delta r + \frac{\partial V}{\partial h}(4,10)\Delta h.$$

At the point $(4,10)$ the partials of V are given by

$$\frac{\partial V}{\partial r} = \frac{2}{3}\pi r h = \frac{80\pi}{3} \quad \text{and} \quad \frac{\partial V}{\partial h} = \frac{1}{3}\pi r^2 = \frac{16\pi}{3}.$$

We thus have

$$\Delta V \sim \left(\frac{80\pi}{3}\right)\left(-\frac{1}{4}\right) + \left(\frac{16\pi}{3}\right)\left(-\frac{1}{2}\right) = -\frac{28\pi}{3}\text{cubic feet.}$$

♠

4.2.3. Section summary

- For a plane $\mathscr{P}$ to be the tangent plane to a surface at a point P, it must contain the tangent line to *every* smooth curve in the surface that passes through P.
- We say a real-valued function f of two variables is differentiable at a point (x_0, y_0) if the graph of f has a tangent plane at $(x_0, y_0, f(x_0, y_0))$.
- One way to test whether a function f is differentiable at (x_0, y_0) is to check whether its first order partials f_x and f_y are continuous at (x_0, y_0). If so, then the function is differentiable. If not, then the test doesn't give enough information to decide.
- If f is differentiable at (x_0, y_0), then its tangent plane at $(x_0, y_0, f(x_0, y_0))$ is given by $z - f(x_0, y_0) = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$.
- We can then approximate f near (x_0, y_0) by its tangent approximation $f(x,y) \sim L(x,y) = f((x_0, y_0)) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$.
- We can also use the language of differentials, writing $\Delta z \sim f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y$ where $\Delta x = x - x_0$ and $\Delta y = y - y_0$, the change in x and y that we control. The expression Δz then approximates the resulting change in $z = f(x,y)$.

Concluding remarks. While we have discussed what it means for a function to be differentiable, we haven't yet defined the derivative of a differentiable function. We've only defined partial and directional derivatives. The notion of derivative will come later, after we introduce linear transformations.

5. Linear Transformations

5.1 The language of functions

Outcomes

- *Understand the following concepts:*
  - *Functions whose domain and range may both have dimension greater than one*
  - *Components of a function*
  - *Independent and dependent variables*
  - *The image of a subset of the domain of a function.*
- *Understand what information is conveyed when we introduce a function by saying, for example, that $F : \mathbb{R}^2 \to \mathbb{R}^3$.*

You are already familiar with several types of functions:

1. *Real-valued functions of one variable.* For example, consider the function $f(x) = 1 + \sqrt{x}$. This function is defined only for $x \geq 0$, so the domain of the function (the input) is $[0, \infty)$. The range of f (the ouput) is $[1, \infty)$ since $1 + \sqrt{x} \geq 1$ always and takes on every value in $[1, \infty)$. You usually visualize this function by drawing its graph $y = f(x)$. (When we refer to $y = f(x)$ as the *graph* of f, we mean that the graph of f is the curve consisting of all points (x, y) such that $y = f(x)$.)

   It is common to alert the reader to the fact that you are about to define a real-valued function whose domain is $[0, \infty)$ by saying something like: Let $f : [0, \infty) \to \mathbb{R}$ be given by $f(x) = 1 + \sqrt{x}$. The set before the arrow tells the reader the domain. The $\mathbb{R}$ after the arrow says that the function is real-valued, i.e., the range is contained in $\mathbb{R}$.

2. *Real-valued functions of two or three variables.* For example, consider the function $f(x, y) = x^2 + y^2$. The domain of this function is $\mathbb{R}^2$. Since $f(x, y)$ takes on every non-negative real value, the range is $[0, \infty)$. Again, you can visualize this function by drawing its graph $z = f(x, y)$.

   Again it is common to alert the reader to the fact that you are about to define a real-valued function whose domain is $\mathbb{R}^2$ by writing: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x) = x^2 + y^2$.

3. *Vector-valued functions of one variable.* For example, consider the function $F : \mathbb{R} \to \mathbb{R}^2$ given by

$$F(t) = \langle \cos(t), \sin(t) \rangle$$

   The range of F is the circle $x^2 + y^2 = 1$. (More precisely, the range consists of the position vectors of all the points on this circle but we normally just think of it as the circle itself.) You are used to viewing this equation as the vector equation for the circle.

4. *Vector-valued functions of more than one variable.* You encountered functions in which both the domain and the range have dimension greater than one in Section 2.3 when we discussed vector equations of planes. For example, consider the function $F : \mathbb{R}^2 \to \mathbb{R}^3$ given by

$$F(s,t) = \langle 1,1,1\rangle + s\langle 2,1,4\rangle + t\langle 1,2,3\rangle$$

or after simplifying:

$$F(s,t) = \langle 1+2s+t, 1+s+2t, 1+4s+3t\rangle$$

This is an example of a vector-valued function of two variables. The range is the plane in $\mathbb{R}^3$ through the point $(1,1,1)$ parallel to the vectors $\langle 2,1,4\rangle$ and $\langle 1,2,3\rangle$. (Again, we are identifying points with their position vectors as in the previous example.)

The word transformation is a synonym for "function". One views the function as "transforming" the domain to the range.

Notational Conventions 5.1

- *For vector-valued functions of one or more variables, we will usually express the values of the function as column matrices. Thus for example, the function F in the 4th example above may be written as*

$$F(s,t) = \begin{bmatrix} 1+2s+t \\ 1+s+2t \\ 1+4s+3t \end{bmatrix} \tag{5.1}$$

  We will think of the column matrix both as a point in $\mathbb{R}^3$ and also as a vector. The advantage of thinking of it as a point is that we can describe the range geometrically. (E.g., in this example the range is a plane.) The advantage of also allowing ourselves to view it as a vector is that vectors can be added and can be multiplied by scalars; these operations don't make sense for points.

- *We will sometimes denote variables in $\mathbb{R}^2$ or $\mathbb{R}^3$ as boldface letters $\mathbf{x}$. E.g., we may write $\mathbf{x} = (x,y)$ when working in $\mathbb{R}^2$ or $\mathbf{x} = (x,y,z)$ when working in $\mathbb{R}^3$. We will then write $[\mathbf{x}]$ for the column vector $[\mathbf{x}] = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ in $\mathbb{R}^2$ and similarly in $\mathbb{R}^3$.*

A function whose range is contained in $\mathbb{R}^3$ (or in $\mathbb{R}^2$) is made up of three (respectively two) real-valued functions called the components or component functions of F. For example, the components of the function $F : \mathbb{R}^2 \to \mathbb{R}^3$ in Equation (5.1) are $F_1(s,t) = 1+2s+t$, $F_2(s,t) = 1+s+2t$ and $F_3(s,t) = 1+4s+3t$.

Example 5.2

- *Define $F : \mathbb{R}^3 \to \mathbb{R}^2$ by*

$$F(x,y,z) = \begin{bmatrix} x^2yz \\ e^{xy+z} \end{bmatrix}.$$

(Again we said $F : \mathbb{R}^3 \to \mathbb{R}^2$ just to alert the reader that we are about to introduce a function with domain $\mathbb{R}^3$ and with range contained in $\mathbb{R}^2$.) The component functions of F are $F_1(x,y,z) = x^2yz$ and $F_2(x,y,z) = e^{xy+z}$.

- *Just as you are used to writing, say $y = f(x)$ when working with real-valued functions of one variable, we often write expressions like*

$$\begin{bmatrix} u \\ v \end{bmatrix} = F(x,y,z)$$

and refer to x,y,z as the independent variables and u,v as the dependent variables. Thus for the function in this example, we might write

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x^2yz \\ e^{xy+z} \end{bmatrix}.$$

Example 5.3

Define a function $F : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$F(x,y) = \begin{bmatrix} 1 & 2 \\ 4 & 1 \\ 7 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

or in the shorthand notation introduced in 5.1 above,

$$F(\mathbf{x}) = \begin{bmatrix} 1 & 2 \\ 4 & 1 \\ 7 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{x} \end{bmatrix}.$$

Find the component functions of F.

Solution. By multiplying matrices, we obtain

$$F(x,y) = \begin{bmatrix} x+2y \\ 4x+y \\ 7x+2y \end{bmatrix}$$

Thus the component functions are $F_1(x,y) = x+2y$, $F_2(x,y) = 4x+y$, and $F_3(x,y) = 7x+2y$ ♠

Definition 5.4: The image of a subset of the domain

- *Given a function F and a point $\mathbf{x}$ in the domain of F, the value $F(\mathbf{x})$ is sometimes referred to as the image of $\mathbf{x}$ under F (or just as the image of $\mathbf{x}$ if it is clear that we are referring to the function F).*
- *If A is a subset of the domain, the image of A (under F) is the subset of the range of F consisting of the images of all the points in A.*

Example 5.5: Image of a subset

Let $F : \mathbb{R} \to \mathbb{R}^2$ be given by

$$F(t) = \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix}.$$

(Note that this is the same function as in 3 above, now written in column form, and is the vector equation of the circle $^2 + y^2 = 1$.)

Find the image of the interval $[0, \pi]$ under F.

Solution. As t goes from 0 to π, the values $F(t)$ trace out the semicircle $x^2 + y^2 = 1$, $y \geq 0$. Thus the image of $[0, \pi]$ is this semicircle. ♠

Notation 5.6

- *When we want to make a statement about all functions whose domain is any of $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$ and whose range is contained in any of $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$, we will write $F : \mathbb{R}^n \to \mathbb{R}^m$ without specifying n and m. Similarly, if we want to allow the domain to be just a subset of $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$, we might write $F : D \to \mathbb{R}^m$ with $D \subset \mathbb{R}^n$.*

Exercises

Exercise 5.1.1 *For each of the following functions, write an expression of the form $F : \mathbb{R}^n \to \mathbb{R}^m$ to indicate the domain and the target space.*

Example: $F(x,y) = \begin{bmatrix} x^3y \\ y\sin(x) \end{bmatrix}$. *Answer:* $F : \mathbb{R}^2 \to \mathbb{R}^2$.

(a) $F(x) = x^4 + x^2$

(b) $F(x) = \begin{bmatrix} x^2 \\ x^3 + 1 \end{bmatrix}$

(c) $F(x,y,z) = \begin{bmatrix} x - y \\ y - z \end{bmatrix}$

(d) $F(s,t) = \begin{bmatrix} e^s \\ s^2t \\ s+2t \end{bmatrix}$

Exercise 5.1.2 *For each function F in Exercise 5.1.1(b)–(d), write down the component functions of F.*

Exercise 5.1.3 *In each of the following, you are given a function and you are given a subset A of the domain of that function. (i) Find the range of the function, and (ii) find the image of the subset A.*

(As an example, for the function and subset in Example 5.5, you could say that the range of the function is the circle $x^2+y^2=1$ *or that the range is the circle centered at the origin of radius one, and that the image of A is the part of the circle with* $y \geq 0$*.)*

(a) $F(x) = x^2$; $A = [1,2]$

(b) $F(t) = \begin{bmatrix} 1+2t \\ 3t \end{bmatrix}$; $A = [1,2]$

(c) $F(s,t) = \begin{bmatrix} s+2t \\ s-t \\ 3s+t \end{bmatrix}$; *A is the t-axis (i.e., A is the line in* $\mathbb{R}^2$ *consisting of all points of the form* $(0,t)$*.)*

Hint: In the more familiar vector form, you can write $F(s,t)$ *as* $s\langle 1,1,3\rangle + t\langle 2,-1,1\rangle$.

Exercise 5.1.4 *For each of the following functions, (i) find the component functions and (ii) write an expression of the form* $f : \mathbb{R}^n \to \mathbb{R}^m$ *(see Exercise 5.1.1).*

(a) $F(x,y,z) = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

(b) $F(x) = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} \begin{bmatrix} x \end{bmatrix}$

(c) $F(x,y,z) = \begin{bmatrix} 1 & 2 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

5.2 Linear Transformations and Their Representing Matrices

Outcomes

A. *Understand the definition of linear transformations and their representing matrices.*

B. *Be able to recognize whether a function $F : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and find its representing matrix.*

C. *Be able to determine the range of a linear transformation from the columns of the representing matrix; understand the concept of rank.*

D. *Gain a geometric understanding of linear transformations and the various ways in which lines and planes are given by linear transformations.*

5.2.1. Real-valued linear transformations

Definition 5.7: Real-valued linear transformations

- *A real-valued linear transformation of one variable is of the form $T(x) = ax$.*
- *A real-valued linear transformation of two variables is of the form $T(x,y) = ax + by$.*
- *A real-valued linear transformation of three variables is of the form $T(x,y,z) = ax + by + cz$.*

Note that if $T : \mathbb{R} \to \mathbb{R}$ is a linear transformation, i.e., $T(x) = ax$, then the graph $y = T(x)$ of T is a straight line through the origin of slope a. Moreover every line through the origin that is not vertical (i.e., not parallel to the y-axis) is the graph of a linear transformation. We now give an analogous description of linear transformations $T : \mathbb{R}^2 \to \mathbb{R}$.

As we have seen, the scalar equation of every plane in $\mathbb{R}^3$ is of the form $ax + by + cz = d$, where $\langle a,b,c \rangle$ is normal to the plane. If the plane passes through the origin, then $d = 0$. We will say that a plane is vertical if it is parallel to the z-axis. The vertical planes are precisely those for which $c = 0$. (Indeed, if the plane is vertical, then the vector $\langle 0,0,1 \rangle$ is parallel to the plane and thus orthogonal to $\langle a,b,c \rangle$. Thus $\langle 0,0,1 \rangle \cdot \langle a,b,c \rangle = 0$, so $c = 0$. The converse is similar.)

If the plane isn't vertical, then $c \neq 0$ and we can solve for z. For example, the plane $3x - 4y + 2z = 0$ can be written as

$$z = -\frac{3}{2}x + 2y.$$

This is the graph of the linear transformation $T(x,y) = -\frac{3}{2}x + 2y$. More generally, we have:

Proposition 5.8

- *Let $T : \mathbb{R}^2 \to \mathbb{R}$ be a linear transformation, i.e., $T(x,y) = ax + by$ where a and b are constants. Then the graph $z = T(x,y)$ is a plane through the origin.*
- *If $\mathscr{P}$ is any plane through the origin that is not vertical, then $\mathscr{P}$ can be expressed as $z = T(x,y)$ where $T : \mathbb{R}^2 \to \mathbb{R}$ is a linear transformation.*

Example 5.9: Contour map

Consider a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}$ such as $T(x,y) = 2x + 3y$. Show that the contour map of T consists of a family of parallel lines. (We leave this to the reader to check.)

For real-valued functions of three variables, we can no longer visualize a graph. However, we can still talk about the contour map, which will consist of level surfaces of the function.

Example 5.10: Contour map

Consider a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}$ such as $T(x,y) = 2x + 3y + 4z$. Show that the contour map of T consists of a family of parallel planes. (We leave this to the reader to check.)

Remark 5.11: Linear transformations and first degree homogeneous polynomials

You are familiar with polynomials in one variable x; for example, $3x^2 + 2x + 5$ is a polynomial of degree 2. Polynomials in two variables x, y are defined similarly; every term is of the form $cx^n y^m$ where c is a real number and n and m are non-negative integers. For example,

$$P(x,y) = 3x^2 + 2xy + 4y^2 + 5x + 7y + 10$$

is a polynomial of degree two. The first three terms (those involving x^2, xy and y^2) are said to be of degree 2 since the sum of the powers of x and y is 2. The next two terms are of degree one and the final constant term is of degree zero. One analogously defines polynomials in three or more variables, e.g., $x^3 + 3xyz + \frac{1}{2}yz^2 + x^2 + 2xz + 7z$ is a polynomial in three variables.

A polynomial is homogeneous if all the terms have the same degree. Thus, for example, the polynomial P above is not homogeneous, while the polynomials $3x^2 + 2xy + 4y^2$ and $5x + 7y$ are homogeneous of degree 2 and degree 1, respectively.

Definition 5.7 is equivalent to:

A non-zero real-valued linear transformation T is a homogeneous polynomial of degree one.

5.2.2. Linear Transformations and their Representing Matrices

Now that we have defined real-valued linear transformations (see Definition 5.7), we can define vector-valued linear transformations as follows:

Definition 5.12: Linear Transformations

A vector-valued function is said to be a linear transformation if each of the component functions is a real-valued linear transformation. (Equivalently, each component function is either zero or a homogeneous polynomial of degree one as in Remark 5.11.)

Here are a few examples of linear transformations:

1. A linear transformation $S : \mathbb{R}^2 \to \mathbb{R}^3$:

$$S(x,y) = \begin{bmatrix} 2x+3y \\ 5x+0y \\ 3x-8y \end{bmatrix}$$

   (It's of course fine to write the second component as $5x$.)

2. A linear transformation $T : \mathbb{R} \to \mathbb{R}^3$:

$$T(x) = \begin{bmatrix} 2x \\ 5x \\ 0 \end{bmatrix}$$

3. A linear transformation $F : \mathbb{R}^3 \to \mathbb{R}^3$:

$$F(x,y,z) = \begin{bmatrix} 2x+3y-5z \\ -x+y-2z \\ x+4y+z \end{bmatrix}$$

In the same way that we associated an augmented matrix to every system of linear equations in Chapter 1, we can associate a matrix to every linear transformation. Moreover, the linear transformation can be expressed by matrix multiplication. We illustrate with the three linear transformations S, T and F above:

1. The linear transformation $S : \mathbb{R}^2 \to \mathbb{R}^3$ was given by $S(x,y) = \left[\begin{smallmatrix} 2x+3y \\ 5x+0y \\ 3x-8y \end{smallmatrix}\right]$. Check that

$$S(x,y) = \begin{bmatrix} 2 & 3 \\ 5 & 0 \\ 3 & -8 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

2. The linear transformation $T : \mathbb{R} \to \mathbb{R}^3$ above can be written as

$$T(x) = \begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix} \begin{bmatrix} x \end{bmatrix}.$$

3. The linear transformation $F : \mathbb{R}^3 \to \mathbb{R}^3$ can be written as

$$F(x,y,z) = \begin{bmatrix} 2 & 3 & -5 \\ -1 & 1 & -2 \\ 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Definition 5.13: Representing Matrix

Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, the matrix associated to T as in the examples above is called the representing matrix of T. We will usually denote the representing matrix by $[T]$ (i.e., by putting brackets around the name of the linear transformation). Observe:

- *The rows of the representing matrix correspond to the component functions of T.*
- *The columns of the representing matrix correspond to the independent variables; the entries in the column are the coefficients of the corresponding variable in the various component functions. Thus the size of the representing matrix $[T]$ is $m \times n$.*

Conversely, starting with any $m \times n$ matrix A, we can define a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ with representing matrix $[T] = A$ by setting $T(\mathbf{x}) = A[\mathbf{x}]$, just as we did in Example 5.3 in Section 5.1.

Notational Conventions 5.14: Viewing elements of $\mathbb{R}^n$ both as points and vectors

- *Following the notational convention 5.1, we will often denote variables in $\mathbb{R}^2$ or $\mathbb{R}^3$ by bold-face letters such as $\mathbf{x}$ and denote column matrices by expressions such as $[\mathbf{x}]$. E.g, we write $[\mathbf{x}] = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ for a column vector in $\mathbb{R}^2$. Especially when working with linear transformations, it is convenient to think of elements of $\mathbb{R}^2$ or $\mathbb{R}^3$ simultaneously as points and as vectors. Thinking of them as points will enable us to describe linear transformations geometrically. Thinking of them as vectors will enable us, for example, to take linear combinations.*
- *While we will normally write the output of a vector-valued linear transformation as a column, we will use both expressions such as (x,y,z) and $\mathbf{x}$ for the input. Thus for example, for a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ with representing matrix $[T] = \left[\begin{smallmatrix} 2 & 4 & 5 \\ 1 & 3 & 4 \end{smallmatrix}\right]$, we will write both*

$$T(x,y,z) = \begin{bmatrix} 2 & 4 & 5 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{and} \quad T(\mathbf{x}) = \begin{bmatrix} 2 & 4 & 5 \\ 1 & 3 & 4 \end{bmatrix} [\mathbf{x}]$$

Many functions that arise naturally turn out to be linear transformations even though it is not immediately obvious. We illustrate with the following example.

Example 5.15: Cross Product

Define $F : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$F(\mathbf{x}) = \langle 1,2,3 \rangle \times \mathbf{x}$$

the cross product of the vectors $\langle 1,2,3 \rangle$ and $\mathbf{x} = \langle x,y,z \rangle$. Show that F is a linear transformation and find its representing matrix.

Solution. By computing the cross product, we find that $F(x,y,z) = \langle -3y+2z, 3x-z, -2x+y \rangle$, or in column form:

$$F(x,y,z) = \begin{bmatrix} -3y+2z \\ 3x-z \\ -2x+y \end{bmatrix} = \begin{bmatrix} 0x-3y+2z \\ 3x+0y-z \\ -2x+y+0z \end{bmatrix}$$

Since each of the component functions is of the form $ax+by+cz$, we see that F is a linear transformation, and we can read off its representing matrix

$$[F] = \begin{bmatrix} 0 & -3 & 2 \\ 3 & 0 & -1 \\ -2 & 1 & 0 \end{bmatrix}$$

♠

Similarly, you will see in the section exercises that vector and scalar projections are linear transformations.

Example 5.16: The Identity Transformation

Let

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

be the 2×2 *identity matrix as in Example 3.12. Observe that the corresponding linear transformation is given by*

$$T(x,y) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

Thus T *maps every element of* $\mathbb{R}^2$ *to itself.* T *is called the* identity transformation *of* $\mathbb{R}^2$.
One defines the identity transformation from $\mathbb{R}^3$ *to* $\mathbb{R}^3$ *similarly using* I_3.

5.2.3. What the Columns of the Representing Matrix Tell Us

Prerequisite 5.17

Before reading this subsection, it is important to review Sections 2.1 and 2.3.

In Subsection 5.2.1, we saw that the real-valued linear transformations $T : \mathbb{R}^2 \to \mathbb{R}$ are precisely those functions of two variables whose graphs are non-vertical planes through the origin in $\mathbb{R}^3$, while real-valued linear transformations of a single variable are the functions whose graphs are non-vertical straight lines through the origin in $\mathbb{R}^2$.

We next address the geometry of vector-valued linear transformations. The key to understanding their geometry is to look at the *columns* of the representing matrix. Recall that when we multiply a matrix A by a column matrix B, the product AB is the linear combination of the columns of A with coefficients specified by the entries of B. This was how we initially defined matrix multiplication in Definition 3.6. We then later observed that a second way to obtain the entries of the product is by taking dot products of the rows of A with the column B. (The two methods of course yielded the same result.) The first viewpoint is especially helpful when working with linear transformations.

The reader should go through the example below carefully before reading the theorem that follows.

Example 5.18

Let $T:\mathbb{R}^2 \to \mathbb{R}^3$ be the linear transformation with representing matrix $[T] = \begin{bmatrix} 2 & 1 \\ 3 & 5 \\ 4 & 7 \end{bmatrix}$. We have

$$T(x,y) = \begin{bmatrix} 2 & 1 \\ 3 & 5 \\ 4 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} + y \begin{bmatrix} 1 \\ 5 \\ 7 \end{bmatrix}. \tag{5.2}$$

Observe:

- *$T(x,y)$ is a linear combination of the columns of $[T]$.*
- *As x and y vary, $T(x,y)$ varies over* all *linear combinations of the columns. Thus the range of T is the subspace of $\mathbb{R}^3$ spanned by the two columns of $[T]$. (This is a plane since the two columns are linearly independent.)*
- $$T(1,0) = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \quad \text{and} \quad T(0,1) = \begin{bmatrix} 1 \\ 5 \\ 7 \end{bmatrix}$$

  which are the two columns of $[T]$. If we think of elements of the domain as vectors (and thus write $\mathbf{i}$ and $\mathbf{j}$ in place of $(1,0)$ and $(0,1)$), then the two columns of $[T]$ are $T(\mathbf{i})$ and $T(\mathbf{j})$.

Remark 5.19

If we change the names of the independent variables in the example above to s and t and use vector notation rather than column notation for $T(s,t)$, then Equation 5.2 becomes

$$T(s,t) = s\langle 2,3,4\rangle + t\langle 1,5,7\rangle$$

which is a familiar type of equation. As in section 2.3, T can be viewed as the vector equation of the plane through the origin spanned by $\langle 2,3,4\rangle$ and $\langle 1,5,7\rangle$. We will return to this viewpoint later.

As in Example 5.18, we have:

Theorem 5.20: Columns Span the Range

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then:

1. *The columns of T are the images of the standard basis vectors of $\mathbb{R}^n$. For example, if $T : \mathbb{R}^3 \to \mathbb{R}^m$ (so the domain is $\mathbb{R}^3$), then the three columns of $[T]$ are given by $T(\mathbf{i})$, $T(\mathbf{j})$ and $T(\mathbf{k})$, respectively, i.e.,*

$$[T] = \begin{bmatrix} | & | & | \\ T(\mathbf{i}) & T(\mathbf{j}) & T(\mathbf{k}) \\ | & | & | \end{bmatrix}$$

2. *T maps every element of $\mathbb{R}^n$ to a linear combination of the columns of $[T]$. For example, if the domain of T is $\mathbb{R}^3$, then*

$$T(x,y,z) = xT(\mathbf{i}) + yT(\mathbf{j}) + zT(\mathbf{k}).$$

3. *Viewing the columns of $[T]$ as vectors, then the range of T is the subspace of $\mathbb{R}^m$ spanned by the columns.*

Note the contrast between linear transformations and many other functions such as $f(x) = x^2$. The range of the function $f(x) = x^2$ is only part of a line, the ray $[0,\infty)$. In contrast, Theorem 5.20 tells us that the range of a linear transformation is always a subspace of $\mathbb{R}^m$. For example, if the target space is $\mathbb{R}^3$, then the range is one of $\{0\}$, a line through the origin, a plane through the origin, or all of $\mathbb{R}^3$.

Definition 5.21: Column Span and Rank

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformations.

- *We will refer to the subspace of $\mathbb{R}^m$ spanned by the columns of $[T]$ as the column span. Thus Theorem 5.20 tells us that the range of T is the column span.*
- *The rank of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is the dimension of the range, equivalently of the column span.*

Example 5.22

Frind the rank and describe the range of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$T(x,y,z) = \begin{bmatrix} 2x+y+5z \\ x+2y+4z \\ 4x+3y+11z \end{bmatrix}$$

Solution. T has representing matrix

$$[T] = \begin{bmatrix} 2 & 1 & 5 \\ 1 & 2 & 4 \\ 4 & 3 & 11 \end{bmatrix}$$

The columns of $[T]$ are not all parallel, so we know that the column span is either a plane or all of $\mathbb{R}^3$. We leave it to the reader to show that the three column vectors are coplanar. Thus the column span, equivalently the range of T, is a plane through the origin in $\mathbb{R}^3$. The rank of T is two since a plane is two-dimensional. ♠

5.2.4. Visualizing linear transformations

5.2.4.1. Linear transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$

How do we visualize a linear transformation $T:\mathbb{R}^2 \to \mathbb{R}^2$? We can't draw the graph, since both the domain and the target space are 2-dimensional. Instead, a good way to try to visualize such linear transformations is to draw two copies of $\mathbb{R}^2$, one representing the domain and the other the target space. If we want to draw the standard coordinate axes in the two copies, we will use different labels in order to tell them apart. E.g, we could label the axes x, y in the first copy and u and v in the second, with the u-axis horizontal and the v-axis vertical. Sometimes, however, the axes just clutter the picture and we may choose not to draw them. We next draw lines, squares, etc. in the first copy of $\mathbb{R}^2$ and draw the images under T of these lines, squares, etc. in the second copy.

We illustrate with an example of a linear transformation $T:\mathbb{R}^2 \to \mathbb{R}^2$ of rank 2. Let T be the linear transformation with standard representing matrix

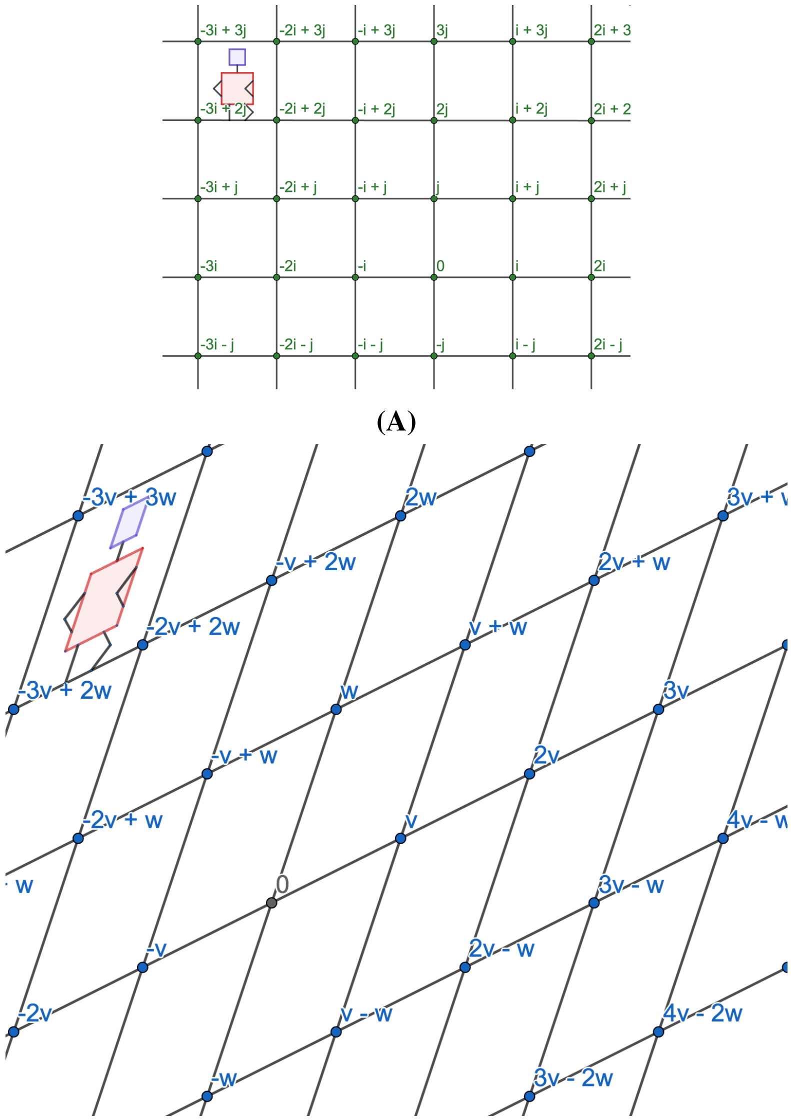
$$[T] = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Letting $\mathbf{v} = \left[\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right]$ and $\mathbf{w} = \left[\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right]$, then $T(\mathbf{i}) = \mathbf{v}$, $T(\mathbf{j}) = \mathbf{w}$, and

$$T(x\mathbf{i} + y\mathbf{j}) = x\mathbf{v} + y\mathbf{w} \tag{5.3}$$

for all x, y.

The vectors $\mathbf{v}$ and $\mathbf{w}$ are the same vectors that we considered in Subsection 5.2.4, now written in column form. Figure 5.1 is a copy of Figures 2.3 and 2.4 from that example. Figure 5.1(A) shows the tiling we get of $\mathbb{R}^2$ with corners given by all the linear combinations of $a\mathbf{i} + b\mathbf{j}$ with a and b integers, and Figure 5.1(B) shows the analogous tiling with $\mathbf{i}$ and $\mathbf{j}$ replaced by $\mathbf{v}$ and $\mathbf{w}$. Interpreting Equation (5.3) geometrically, we see that the linear transformation T maps each of the squares in Figure 5.1(A) to the corresponding parallelogram in Figure 5.1(B). (E.g., the square with lower left corner $2\mathbf{i} + 3\mathbf{j}$ goes to the parallelogram with lower left corner $2\mathbf{v} + 3\mathbf{w}$.) Horizontal lines are mapped to lines parallel to $\mathbf{v}$ and vertical lines are mapped to lines parallel to $\mathbf{w}$. In other words, the effect of the linear transformation is to replace the grid of perpendicular streets in Figure 5.1(A) to the grid of streets in Figure 5.1(B). You can imagine taking a picture of a city with the street system in Figure 5.1(A) and redrawing it in slanted fashion – all the same houses and buildings, etc. – but everything gets slanted and stretched. In fact, we've drawn a picture of a (perhaps giant!) pedestrian walking along a street in Figure 5.1(A) and the image of that pedestrian in Figure 5.1(B). That's the effect of T.


-3i + 3j
-2i + 3j
-i + 3j
3j
i + 3j
-3i + 2j
-2i + 2j
-i + 2j
2j
i + 2j
-3i + j
-2i + j
-i + j
j
i + j
2i + j
-3i
-2i
-i
0
i
2i
-3i - j
-2i - j
-i - j
-j
i - j
2i - j
-3v + 3w
2w
-v + 2w
2v + w
-2v + 2w
v + w
-3v + 2w
w
3v
-v + w
2v
-2v + w
v
4v - w
0
3v - w
-v
2v - w
-2v
v - w
4v - 2w
-w
3v - 2w

(A)

(B)

Figure 5.1

Remark 5.23

- *The pictures above suggest why we use the language "linear transformation". In the example above, you can think of T as transforming the plane, e.g., by doing things like stretching in certain directions and changing angles.*
- *Linear transformations $T : \mathbb{R}^3 \to \mathbb{R}^3$ of rank 3 can be viewed similarly, although they are more difficult to draw. You can visualize the tiling of the plane $\mathbb{R}^3$ by unit cubes being transformed to a tiling by parallelapipeds.*
- *For linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^2$ of rank one, the two column vectors of $[T]$ are parallel so the range (the column span) is only a line in $\mathbb{R}^2$. One can visualize collapsing the plane $\mathbb{R}^2$ (the domain) to a line (the range). Similarly, for a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ of rank only 1 or 2, the range will be only a line or a plane in $\mathbb{R}^3$. In Exercise 5.2.15, you will see an example given by a vector projection.*

5.2.4.2. Linear transformations from $\mathbb{R}$ to $\mathbb{R}^2$ or $\mathbb{R}^3$

Consider a non-zero linear transformation $T : \mathbb{R} \to \mathbb{R}^2$.

Example 5.24

For a specific example, suppose T has matrix

$$[T] = \begin{bmatrix} -2 \\ -1 \end{bmatrix}$$

Denoting the independent variable by t, we have

$$T(t) = \begin{bmatrix} -2 \\ -1 \end{bmatrix} [t] \quad \text{or in vector notation} \quad T(t) = t\langle -2, -1 \rangle.$$

In Section 2.3, we saw that such an equation can be viewed as the vector equation of a line ℓ through the origin in $\mathbb{R}^2$.

If you prefer a picture analogous to that in Figure 5.1, we can view T as transforming the real line to the line ℓ as in Figure 5.2.

More generally we have:

Theorem 5.25

Vector equations of lines through the origin in $\mathbb{R}^2$ or $\mathbb{R}^3$ are given by linear transformations with domain $\mathbb{R}$ and target space $\mathbb{R}^2$ or $\mathbb{R}^3$, respectively.

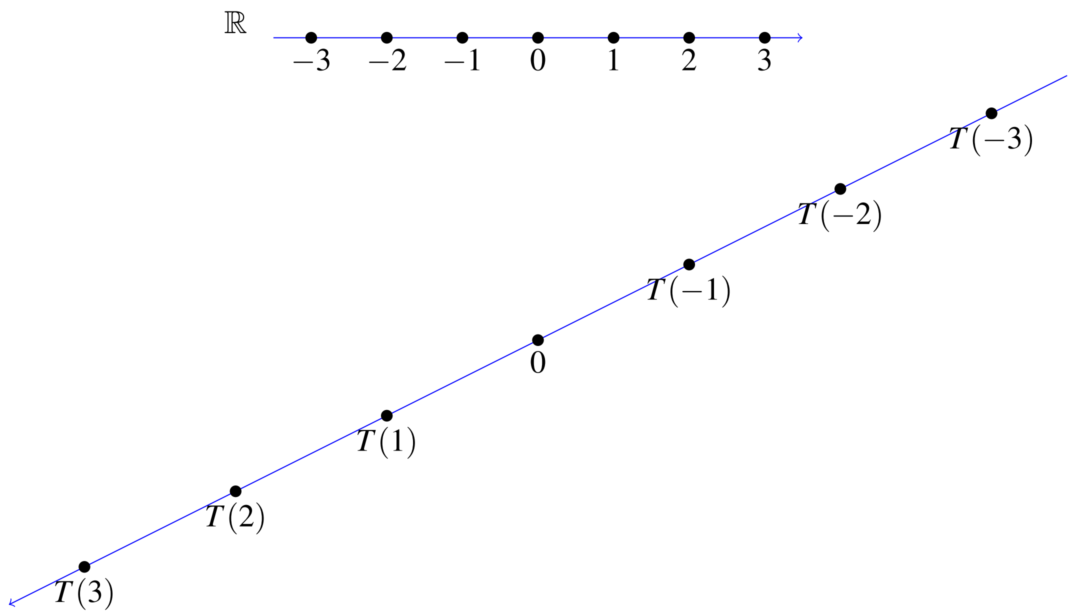


Figure 5.2

5.2.4.3. Linear transformations from $\mathbb{R}^2$ to $\mathbb{R}^3$

Linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^3$ correspond to 3×2 matrices. Since $[T]$ has two columns, the rank (the dimension of the column span) must be one of 0, 1, or 2. We will focus here on the case of rank 2.

Recall Example 5.18 and the subsequent Remark 5.19. There we saw that the linear transformation with representing matrix

$$[T] = \begin{bmatrix} 2 & 1 \\ 3 & 5 \\ 4 & 7 \end{bmatrix}$$

can be viewed as the vector equation of the plane spanned by the column vectors:

$$\mathscr{P} = \mathrm{Span}(\langle 2,3,4\rangle, \langle 1,5,7\rangle).$$

Indeed, expressed in vector form rather than column form, we saw that $T(s,t) = s\langle 2,3,4\rangle + t\langle 1,5,7\rangle$.

Again, if you prefer a picture analogous to Figure 5.1, you can picture the plane $\mathbb{R}^2$ transformed to the plane $\mathscr{P}$ in $\mathbb{R}^3$. The standard tiling of $\mathbb{R}^2$ by squares, is transformed to the tiling of the plane $\mathscr{P}$ by parallelograms whose sides are given by the vectors $\langle 2,3,4\rangle$ and $\langle 1,5,7\rangle$.

More generally,

Theorem 5.26

The vector equation of every plane through the origin in $\mathbb{R}^3$ is given by a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^3$ of rank 2.

5.2.5. Section summary

- Non-zero real-valued linear transformations are homogeneous polynomials of degree one. A vector-valued function is a linear transformation if each of its component functions is either zero or a homogeneous polynomial of degree one.
- The graph of any linear transformation $T : \mathbb{R} \to \mathbb{R}$ is a non-vertical line through the origin in $\mathbb{R}^2$.
- The graph of any linear transformation $T : \mathbb{R}^2 \to \mathbb{R}$ is a non-vertical plane through the origin in $\mathbb{R}^3$.
- Every linear transformation can be expressed as $T(\mathbf{x}) = [T][\mathbf{x}]$ (matrix multiplication), where $[T]$ is the representing matrix of $[T]$.
- The range of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is the subspace of $\mathbb{R}^m$ spanned by the columns of $[T]$, also called the "column span". The "rank" of T is the dimension of the range.
- If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then the columns of $[T]$ are the images of the standard basis vectors of $\mathbb{R}^n$. E.g., if $T : \mathbb{R}^2 \to \mathbb{R}^m$, then the two columns of $[T]$ are $T(\mathbf{i})$ (i.e., $T(1,0)$) and $T(\mathbf{j})$ (i.e., $T(0,1)$).
- The range of a non-zero linear transformation with domain $\mathbb{R}$ and target space $\mathbb{R}^2$ or $\mathbb{R}^3$ is a line through the origin in $\mathbb{R}^2$ or $\mathbb{R}^3$ and the linear transformation may be viewed as giving the vector equation of this line.
- The range of a linear transformation of rank two with domain $\mathbb{R}^2$ and target space $\mathbb{R}^3$ is a plane through the origin in $\mathbb{R}^3$ and the linear transformation may be viewed as giving the vector equation of this plane.
- You can picture a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ of rank 2 by drawing two copies of $\mathbb{R}^2$ and illustrating how the first is "transformed" by T into the second one as in Figure 5.1.

Exercises

Exercise 5.2.1 *Decide whether each of the following functions is a linear transformation.*

(a) $T(x) = -2x + 5$

(b) $T(x,y) = x - y$

(c) $T(x,y) = x^2 + xy + y^2$

(d) $T(x,y,z) = \begin{bmatrix} x+y \\ y+z \end{bmatrix}$

(e) $T(x,y,z) = \begin{bmatrix} x+1 \\ y+1 \\ z+1 \end{bmatrix}$

(f) $T(x,y,z) = \begin{bmatrix} xy \\ yz \\ zx \end{bmatrix}$

Exercise 5.2.2 *For each function T in Exercise 5.2.1 that is a linear transformation, indicate its domain and target space, and then write down the representing matrix.*

Exercise 5.2.3 *Define $F : \mathbb{R}^3 \to \mathbb{R}^3$ by*

$$F(\mathbf{x}) = \langle 1,1,1\rangle \times \mathbf{x}$$

(the cross product of $\langle 1,1,1\rangle$ with $\mathbf{x}$).

(a) *Check that F is a linear transformation and find its representing matrix.*

(b) *Find $\mathbf{x}$ such that $F(\mathbf{x})$ is parallel to $\langle 1,-2,1\rangle$.*

Exercise 5.2.4 *Define $T : \mathbb{R}^3 \to \mathbb{R}$ by*

$$T(\mathbf{x}) = \langle 1,2,4\rangle \cdot \mathbf{x}$$

(the dot product of $\langle 1,2,4\rangle$ with $\mathbf{x}$). Check that T is a linear transformation and write down its representing matrix.

Exercise 5.2.5 *For each of the following linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^2$, draw a parallelogram tiling analogous to Figure 5.1(b) showing the image under T of the square tiling in Figure 5.1(a). Include in your picture the image of the pedestrian, placing him in the correct parallelogram. Be sure to make his orientation clear and stretch or compress him as directed by T.*

(a) $[T] = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

(b) $[T] = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$

(c) $[T] = \begin{bmatrix} -1 & -2 \\ 2 & 1 \end{bmatrix}$

(d) $[T] = \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}$

Exercise 5.2.6 *Write down the representing matrix of a linear transformation $T : \mathbb{R} \to \mathbb{R}^3$ that gives the vector equation (in the sense of Example 5.24) for the line in $\mathbb{R}^3$ through the points $(0,0,0)$ and $(1,2,3)$.*

Exercise 5.2.7 *Write down the representing matrix of a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ that gives the vector equation (in the sense of Theorem 5.26 and the example that precedes it) for the plane in $\mathbb{R}^3$ through the points $(0,0,0)$, $(1,2,3)$ and $(4,1,4)$.*

Exercise 5.2.8 *Write down a linear transformation $S : \mathbb{R}^2 \to \mathbb{R}$ such that the graph $z = S(x,y)$ of S is the plane in Exercise 5.2.7.*

Exercise 5.2.9 *For each of the following planes, write down a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}$ whose graph $z = T(x,y)$ is the indicated plane.*

(a) $z = 5x - 2y$

(b) $x + 3y - 2z = 0$

(c) *the plane through the origin orthogonal to the line given parametrically by* $x = 1 + t,\ y = 2 - 3t,\ z = t$.

(d) *the plane whose vector equation is expressed (in the sense of Theorem 5.26) by the linear transformation* $S : \mathbb{R}^2 \to \mathbb{R}^3$ *whose representing matrix is*

$$[S] = \begin{bmatrix} 2 & 1 \\ 4 & 5 \\ 1 & 0 \end{bmatrix}$$

Exercise 5.2.10 *Draw the contour map of the linear transformation* $T(x,y) = 3x + 4y$.

Exercise 5.2.11 *Describe the contour map of the linear transformation* $T(x,y,z) = 2x + y + 3z$.

Exercise 5.2.12 *We have talked about level sets of real-valued functions. It also makes sense to talk about level sets of functions whose target space has higher dimension. For example, consider the linear transformation* T *with matrix*

$$[T] = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \end{bmatrix}$$

(a) *Show that the range of* T *is all of* $\mathbb{R}^2$.

(b) *Find the set of all* (x,y,z) *in* $\mathbb{R}^3$ *such that* $T(x,y,z) = \begin{bmatrix} 6 \\ 11 \end{bmatrix}$. *Then describe your answer in words.*

This is an example of a level set of T. *(More generally, the level set of* T *corresponding to* $\begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$ *in* $\mathbb{R}^2$ *is the solution set of* $T(x,y,z) = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$. *)*

(c) *The contour map of* T *consists of all the level sets. Describe the contour map explicitly. (E.g., if the level sets are lines, indicate their directions.)*

Exercise 5.2.13 *For each of the following, you are given the representing matrix* $[T]$ *of a linear transformation with target space* $\mathbb{R}^2$. *In each case, determine whether the range of* T *is a line, all of* $\mathbb{R}^2$ *or just the origin. Also indicate the rank.*

(a) $[T] = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

(b) $[T] = \begin{bmatrix} 1 & 2 \\ 4 & 8 \end{bmatrix}$

(c) $[T] = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 1 \end{bmatrix}$

(d) $[T] = \begin{bmatrix} 1 & 2 & 3 \\ -4 & -8 & -12 \end{bmatrix}$

Exercise 5.2.14 *For each of the following, you are given the representing matrix $[T]$ of a linear transformation with target space $\mathbb{R}^3$. In each case, determine whether the range of T is a line through the origin, a plane through the origin, all of $\mathbb{R}^3$ or just the origin. Also indicate the rank.*

(a) $[T] = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \end{bmatrix}$

(b) $[T] = \begin{bmatrix} 1 & 2 & -5 \\ -2 & -4 & 10 \\ 3 & 6 & -15 \end{bmatrix}$

(c) $[T] = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 1 \\ 1 & 0 & 1 \end{bmatrix}$

(d) $[T] = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 3 & 4 & 7 \end{bmatrix}$

(e) $[T] = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 3 & 4 & 1 \end{bmatrix}$

Exercise 5.2.15 *Let*

$$\mathbf{a} = \langle 1, 2, 1 \rangle.$$

(a) *Define $S : \mathbb{R}^3 \to \mathbb{R}$ by $S(\mathbf{x}) = \mathrm{comp}_{\mathbf{a}}\, \mathbf{x}$ (the scalar projection of $\mathbf{x}$ on $\mathbf{a}$). By writing $\mathbf{x} = \langle x, y, z \rangle$ and evaluating $S(\mathbf{x})$, check that F is a linear transformation and write down its representing matrix.*

(b) *Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(\mathbf{x}) = \mathrm{proj}_{\mathbf{a}}(\mathbf{x})$. Find an explicit expression for T, verify that T is a linear transformation and write down its representing matrix. (More generally, one can show that $\mathrm{proj}_{\mathbf{a}}$ is a linear transformation for any choice of $\mathbf{a}$.)*

(c) *Write down a basis for the subspace of $\mathbb{R}^3$ spanned by the columns of $[T]$, where T is the linear transformation in part (b). Why is this answer what you would have expected even before you computed the matrix $[T]$? (Think about what T means geometrically.)*

5.3 Linearity Properties

Outcomes

- *Understand how the properties of matrix multiplication, such as the distributive law, result in two special properties of linear transformations, called the linearity properties, that other functions don't have.*
- *Be able to state the two linearity properties as equations.*
- *Most important outcome: Understand the two linearity properties geometrically.*

Since linear transformations are given by matrix multiplication, the properties of matrix multiplication give us some special properties of linear transformations that distinguish them from more general functions.

The properties we will state are valid in all dimensions but you will probably find it helpful to think of both the domain $\mathbb{R}^n$ and the target space $\mathbb{R}^m$ as $\mathbb{R}^2$. We will illustrate the properties with drawings in $\mathbb{R}^2$.

In order to state the linearity properties, we need to think of elements in both the domain and range as vectors rather than as points.

Theorem 5.27

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be any linear transformation. For each vector $\mathbf{v}$ in the domain $\mathbb{R}^n$ and each scalar c, we have

$$\text{First linearity property: } T(c\mathbf{v}) = cT(\mathbf{v}). \tag{5.4}$$

Geometric interpretation: Letting ℓ be the line through the origin in $\mathbb{R}^n$ with direction vector $\mathbf{v}$, we have:

- *If $T(\mathbf{v}) \neq 0$, then T maps the line ℓ through the origin in $\mathbb{R}^n$ with direction vector $\mathbf{v}$ to the line through the origin in $\mathbb{R}^m$ with direction vector $T(\mathbf{v})$. Moreover, it does so in a proportionate way as illustrated in Figure 5.3; i.e, points equally spaced along ℓ are mapped to points equally spaced along the image of ℓ.*
- *If $T(\mathbf{v}) = 0$, then T maps the entire line ℓ to the origin in $\mathbb{R}^m$.*

Proof. We have

$$T(c\mathbf{v}) = [T][c\mathbf{v}] = c[T][\mathbf{v}] = cT(\mathbf{v})$$

where the second equality comes from the fact that scalars pull out of matrix multiplication. ♠

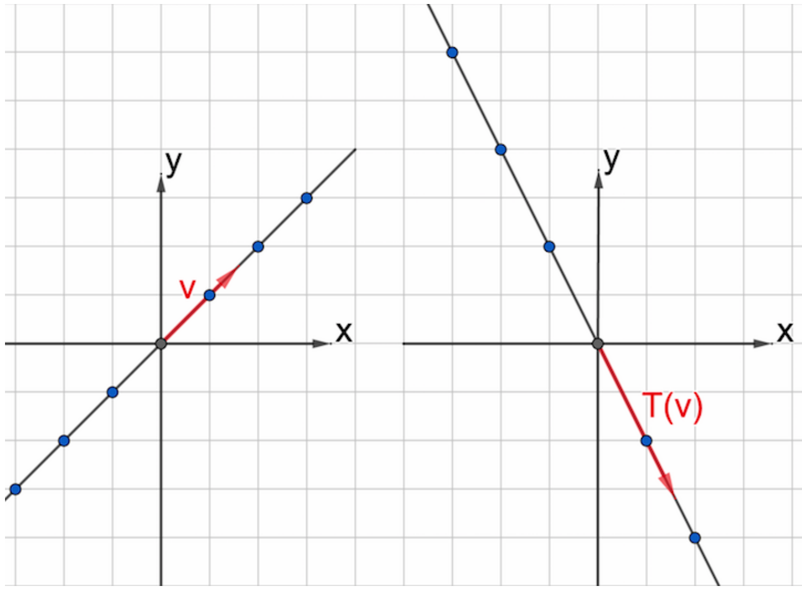


Figure 5.3

Before stating the second linearity property, we introduce notation:

Notation 5.28

If $\mathbf{v}$ *and* $\mathbf{w}$ *are not parallel, then* $\text{Par}(\mathbf{v},\mathbf{w})$ *will denote the parallelogram with sides* $\mathbf{v}$ *and* $\mathbf{w}$.

Theorem 5.29

Let $T:\mathbb{R}^n \to \mathbb{R}^m$ *be any linear transformation. For every pair of vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *in the domain* $\mathbb{R}^n$, *we have*

$$\textit{Second linearity property}: \quad T(\mathbf{v}+\mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w}). \tag{5.5}$$

Geometric interpretation: Let $\mathbf{v}$ *and* $\mathbf{w}$ *be any pair of vectors in the domain* $\mathbb{R}^n$ *that aren't parallel. Except in the special case that* $T(\mathbf{v})$ *and* $T(\mathbf{w})$ *happen to be parallel,* T *maps the diagonal of* $\text{Par}(\mathbf{v},\mathbf{w})$ *to the diagonal of* $\text{Par}(T(\mathbf{v}),T(\mathbf{w}))$. *See Figure 5.4.*

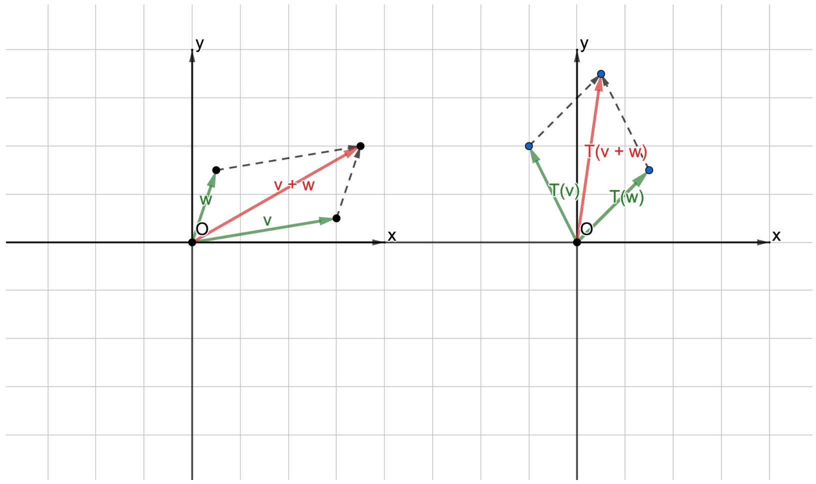


Figure 5.4

One can combine the two linearity properties to obtain:

Corollary 5.30

A linear transformation takes linear combinations of vectors to the corresponding linear combinations of their images, i.e.,

$$T(a\mathbf{v}+b\mathbf{w}) = aT(\mathbf{v})+bT(\mathbf{w})$$

for all vectors $\mathbf{v}$ *and* $\mathbf{w}$ *in the domain and all scalars* a *and* b.

The same property holds for linear combinations of any number of vectors. E.g.,

$$T(a\mathbf{u}+b\mathbf{v}+c\mathbf{w}) = aT(\mathbf{u})+bT(\mathbf{v})+cT(\mathbf{w}).$$

We have already seen the special case in which $\mathbf{u} = \mathbf{i}$, $\mathbf{v} = \mathbf{j}$ and $\mathbf{w} = \mathbf{k}$ in Theorem 5.20.

Example 5.31

Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^3$ *is a linear transformation and that*

$$T(1,2) = \begin{bmatrix} 4 \\ 7 \\ 8 \end{bmatrix} \quad \text{and} \quad T(3,1) = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

Find $T(5,5)$ *and* $T(1,1)$.

Solution. We think of the input as vectors rather than points. We first express the vector $\langle 5,5\rangle$ as a linear combination of $\langle 1,2\rangle$ and $\langle 3,1\rangle$. We leave it to the reader to solve for the coefficients, obtaining

$$\langle 5,5\rangle = 2\langle 1,2\rangle + \langle 3,1\rangle.$$

By Corollary 5.30, we thus have

$$T(5,5) = 2T(1,2)+T(3,1) = 2\begin{bmatrix} 4 \\ 7 \\ 8 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 9 \\ 14 \\ 18 \end{bmatrix}$$

The easiest way to find $T(1,1)$ now is to use the fact that $\langle 1,1\rangle = \frac{1}{5}\langle 5,5\rangle$. Thus by the first linearity property,

$$T(1,1) = \frac{1}{5}T(5,5) = \begin{bmatrix} 9/5 \\ 14/5 \\ 18/5 \end{bmatrix}$$

♠

Example 5.32

Suppose you are told only that $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation and that

$$T(2,1) = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} \quad \textit{and} \quad T(1,1) = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}.$$

Find $T(1,0)$ and $T(0,1)$. Then write down the representing matrix of T.
Note: Once you've done this, you will know the linear transformation T completely!

Solution. First write $\mathbf{i}$ as a linear combination of $\langle 2,1\rangle$ and $\langle 1,1\rangle$:

$$\mathbf{i} = \langle 1,0\rangle = \langle 2,1\rangle - \langle 1,1\rangle.$$

By Corollary 5.30, we thus have (expressing the inputs and outputs of T as vectors):

$$T(\mathbf{i}) = T(\langle 2,1\rangle) - T(\langle 1,1\rangle) = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \\ -1 \end{bmatrix}$$

Similarly, $\mathbf{j} = \langle 0,1\rangle = 2\langle 1,1\rangle - \langle 2,1\rangle$, so

$$T(\mathbf{j}) = 2T(\langle 1,1\rangle) - T(\langle 2,1\rangle) = 2\begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ 3 \end{bmatrix}$$

Thus by Theorem 5.20, we have $[T] = \begin{bmatrix} | & | \\ T(\mathbf{i}) & T(\mathbf{j}) \\ | & | \end{bmatrix} = \begin{bmatrix} -2 & 5 \\ -1 & 6 \\ -1 & 3 \end{bmatrix}$ ♠

Remark 5.33: Every function that satisfies the linearity properties is a linear transformation

We have seen that linear transformations satisfy the two linearity properties $T(c\mathbf{v}) = cT(\mathbf{v})$ and $T(\mathbf{v}+\mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$. In fact, linear transformations are the only functions that satisfy both linearity properties. In other words, to check whether a function is a linear transformation, it's enough to show that it satisfies the linearity properties. In more advanced courses, linear transformations are defined to be functions that satisfy the linearity properties.

Aside: To see why linear transformations are the only functions that satisfy these properties, suppose for example that $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a function that satisfies the linearity properties. We will show that F is a linear transformation. Write $F(\mathbf{i}) = \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ and $F(\mathbf{j}) = \left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$. (Here a, b, c, d are specific real numbers determined by F.) Since we are assuming that F satisfies the two linearity properties, we have

$$F(x\mathbf{i} + y\mathbf{j}) = F(x\mathbf{i}) + F(y\mathbf{j}) = xF(\mathbf{i}) + yF(\mathbf{j}) = x\begin{bmatrix} a \\ b \end{bmatrix} + y\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

(The first equality comes from the second linearity property and the second inequality comes from the first linearity property.)
Thus F is the linear transformation with representing matrix $\left[\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right]$.

5.3.1. Section summary

- The first linearity property says that $T(c\mathbf{v}) = cT(\mathbf{v})$ for all vectors $\mathbf{v}$ in the domain and all scalars c. Geometrically, this says that the image of a line through the origin is another line through the origin, scaled proportionately. (Special case: if $T(\mathbf{v}) = 0$, then the image of the line with direction vector $\mathbf{v}$ is just the origin rather than another line.)

- The second linearity property says that $T(\mathbf{v}+\mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$. This says that the image of the diagonal of the parallelogram $\mathrm{Par}(\mathbf{v}, \mathbf{w})$ is the diagonal of $\mathrm{Par}(T(\mathbf{v}), T(\mathbf{w}))$.

  (The formula $T(\mathbf{v}+\mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$ is valid always, but our geometric interpretation only makes sense when $T(\mathbf{v})$ and $T(\mathbf{w})$ are not parallel.)

- The two linearity properties together tell us that T takes linear combinations of vectors to linear combinations of their images: $T(a\mathbf{v} + b\mathbf{w}) = aT(\mathbf{v}) + bT(\mathbf{w})$.

- Linear transformations are the only functions that satisfy the linearity properties; this is one of the reasons that linear transformations are pretty special!

Exercises

Exercise 5.3.1 *Draw a sketch analogous to Figure 5.4 illustrating that $T(\mathbf{v}-\mathbf{w}) = T(\mathbf{v}) - T(\mathbf{w})$, when T is a linear transformation.*

Exercise 5.3.2 *Define $F : \mathbb{R}^3 \to \mathbb{R}^3$ by*

$$F(\mathbf{x}) = \langle 1,2,3 \rangle \times \mathbf{x}$$

. Use properties of the cross product to verify that T satisfies the two linearity properties.

Exercise 5.3.3 *Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation, let $\mathbf{v} = \langle 4,1 \rangle$ let $\mathbf{w} = \langle 2,1 \rangle$, and let $\mathbf{u} =$ $2\mathbf{v}+3\mathbf{w} = \langle 14,5 \rangle$. If*

$$T(4,1) = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \quad \text{and } T(2,1) = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

find $T(14,5)$.

Exercise 5.3.4 *Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation and that $T(1,1) = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}$ and $T(2,3) =$ $\begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$. Find the standard representing matrix $[T]$.*

Exercise 5.3.5 *Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation and that $T(2,3) = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ and $T(1,2) =$ $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$. Find the standard representing matrix $[T]$.*

5.4 Example: Rotations and Reflections of $\mathbb{R}^2$

Outcomes

- *Understand what we mean by rotations about the origin and reflections across lines.*
- *Gain a geometric understanding of why rotations and reflections are linear transformations.*
- *Be able to write down the representing matrices of rotations and reflections.*

Many geometrically natural functions can be seen to be linear transformations. In Exercise 5.2.15, we saw that projections are linear transformations. In that exercise, you first wrote down the explicit formula for the projection and then recognized it as a linear transformation. In this subsection, we are going to take a different approach. We begin with a geometric transformation, verify geometrically that it satisfies the linearity properties and thus is a linear transformation, and use this information to find an explicit formula.

5.4.1. Rotations

Definition 5.34

Fix an angle θ. Define $\mathrm{Rot}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ to be the function that rotates each vector through the angle θ about the origin.

In Figure 5.5, we have illustrated the effect of $\mathrm{Rot}_{\pi/3}$ (counterclockwise rotation through angle $\frac{\pi}{3}$) on a vector $\mathbf{v}$.

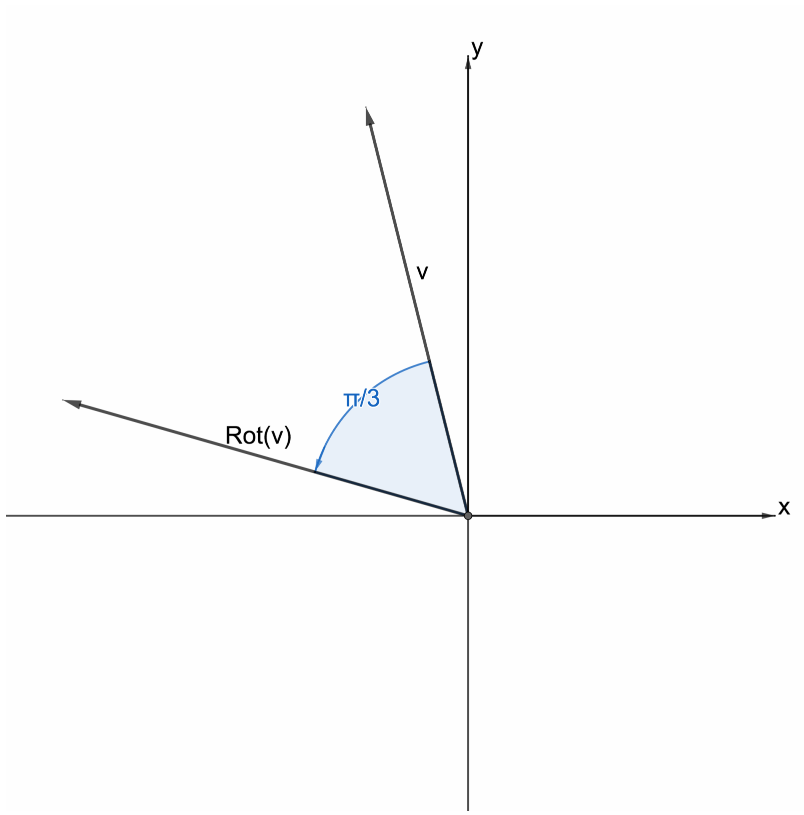


Figure 5.5

Rotations arise extensively in many different settings. An engineer may need to figure out how to rotate a robot's arm. In computer graphics, all or part of a picture may need to be rotated. The reader can think of many more applications.

Here is our plan to find an exact expression for Rot_θ.

Step 1: First show geometrically that Rot_θ is a linear transformation.

Step 2: Compute $\mathrm{Rot}_\theta(\mathbf{i})$ and $\mathrm{Rot}_\theta(\mathbf{j})$. These will be the columns of the representing matrix $[\mathrm{Rot}_\theta]$ and we're done!

Note that the second step would not have been enough without the first. For general functions $F : \mathbb{R}^2 \to \mathbb{R}^2$, knowing $F(\mathbf{i})$ and $F(\mathbf{j})$ does not tell you anything about $F(\mathbf{x})$ for other values of $\mathbf{x}$.

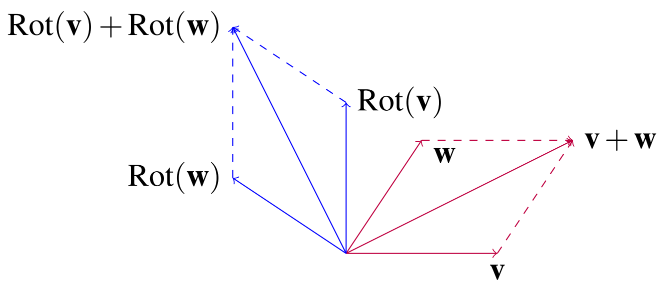


Figure 5.6

Step 1. To see that Rot_θ is a linear transformation, Remark 5.33 tells us that we just need to check that Rot_θ satisfies the two linearity properties in Theorems 5.27 and 5.29. We do this geometrically. We leave the first property to the reader; i.e., check that rotations take lines through the origin to lines through the origin in a proportionate way. For the second linearity property, see Figure 5.6. There we have illustrated the effect of rotation through an angle slightly larger than $\frac{\pi}{2}$. Even though we normally draw the image vectors on a separate set of axes, here we have drawn them on the same set of axes to make it more obvious how all the vectors are rotated. As the drawing shows, parallelograms rotate to parallelograms and the diagonal of the first parallelogram goes to the diagonal of the rotated one.
Step 2. Compute $\mathrm{Rot}_\theta(\mathbf{i})$ and $\mathrm{Rot}_\theta(\mathbf{j})$. See Figure 5.7. (We have rotated through an acute angle θ in Figure 5.7. The reader is encouraged to draw pictures using other angles as well.) We obtain:

$$\mathrm{Rot}_\theta(\mathbf{i}) = \begin{bmatrix}\cos(\theta)\\ \sin(\theta)\end{bmatrix} \quad \mathrm{Rot}_\theta(\mathbf{j}) = \begin{bmatrix}-\sin(\theta)\\ \cos(\theta)\end{bmatrix} \tag{5.6}$$

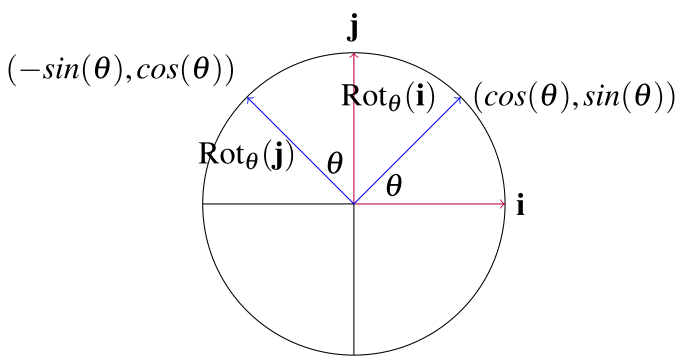


Figure 5.7

From Equation (5.6), we can write down the representing matrix for Rot_θ:

$$[\mathrm{Rot}_\theta] = \begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix} \tag{5.7}$$

Example 5.35

Find the result of rotating the point $(6,2)$ *through an angle of* $-\frac{\pi}{3}$ *about the origin.*
Solution. *We first find a general formula for the rotation* $\mathrm{Rot}_{-\pi/3}$ *using Equation (5.7).*

$$[\mathrm{Rot}_{-\pi/3}] = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$

Thus

$$\mathrm{Rot}_{-\pi/3}(6,2) = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 6 \\ 2 \end{bmatrix} = \begin{bmatrix} 3+\sqrt{3} \\ -3\sqrt{3}+1 \end{bmatrix}$$

Thus the point $(6,2)$ *is rotated to the point* $(3+\sqrt{3}, -3\sqrt{3}+1)$. ♠

5.4.2. Reflections

You are probably familiar with the notion of reflecting a point across the x-axis or the y-axis. For example, if we reflect the point (2,5) across the y-axis, we get the point (-2,5). More generally, reflection across the y-axis is given by the function $T(x,y) = \left[\begin{smallmatrix} -x \\ y \end{smallmatrix}\right]$, which you will recognize to be a linear transformation.

Similarly, if ℓ is any line through the origin in $\mathbb{R}^2$ and $P = (x,y)$ is any point in $\mathbb{R}^2$, we obtain the reflection of P across ℓ as follows:

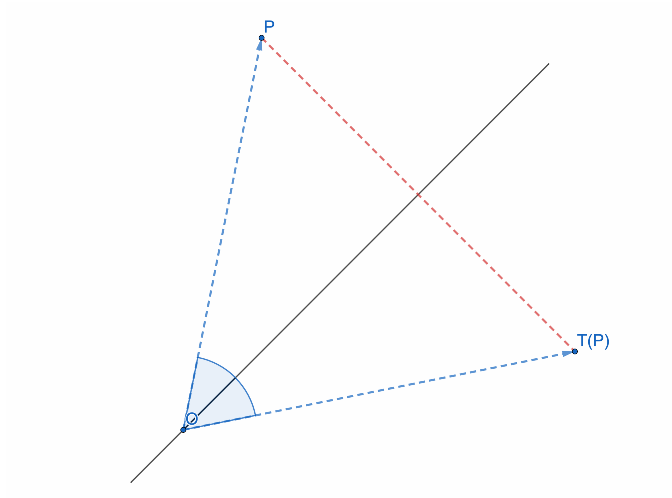


Figure 5.8

- If P doesn't lie on ℓ, draw a perpendicular from P to ℓ and continue along it until you reach a point Q equidistant from ℓ. The point Q is the reflection of P across ℓ. (In other words, Q is the unique point such that ℓ is the perpendicular bisector of the segment PQ. The perpendicular segment appears in red in Figure 5.8.)

- If P happens to lie on ℓ, then reflection doesn't change P.

You can also visualize reflection by drawing the position vector of P. The position vector of the reflection of P makes the same angle with ℓ and is the same length as that of P. See Figure 5.8.

Note the symmetry here: P and Q are reflections of each other acros ℓ.

Notation 5.36

Let ℓ_θ denote the line through the origin in $\mathbb{R}^2$ that makes the angle θ with the positive x-axis. (It doesn't matter in any of the computations, but you can always choose θ to lie in the interval $[0,\pi)$ if you wish.) Denote by $\mathrm{Ref}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ the function that sends every point to its reflection across ℓ_θ.

In Figure 5.9, we have drawn a line ℓ_θ and various regions such as a pentagon and a figure of a duck. The images under Ref_θ of the various regions are drawn in the same color. (Note: we haven't specified which is the original and which is the image since each is the image of the other under reflection!)

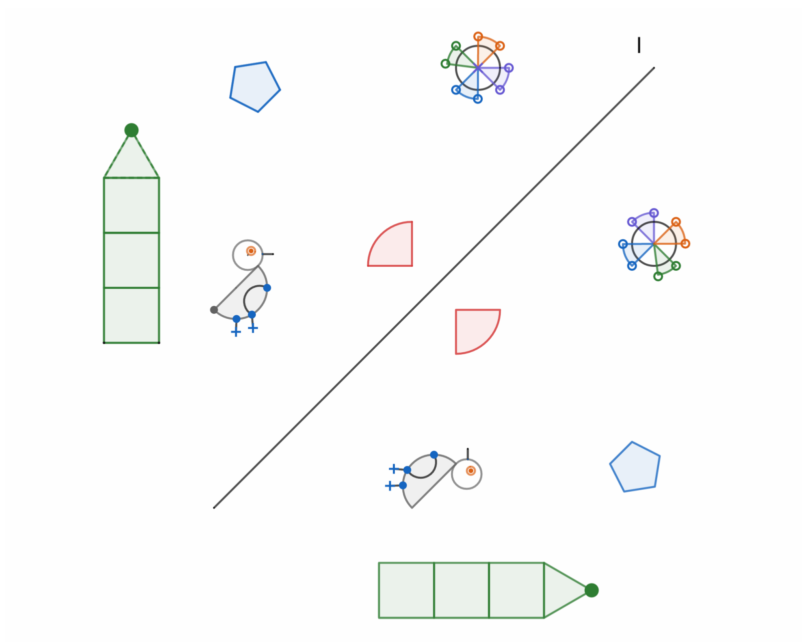
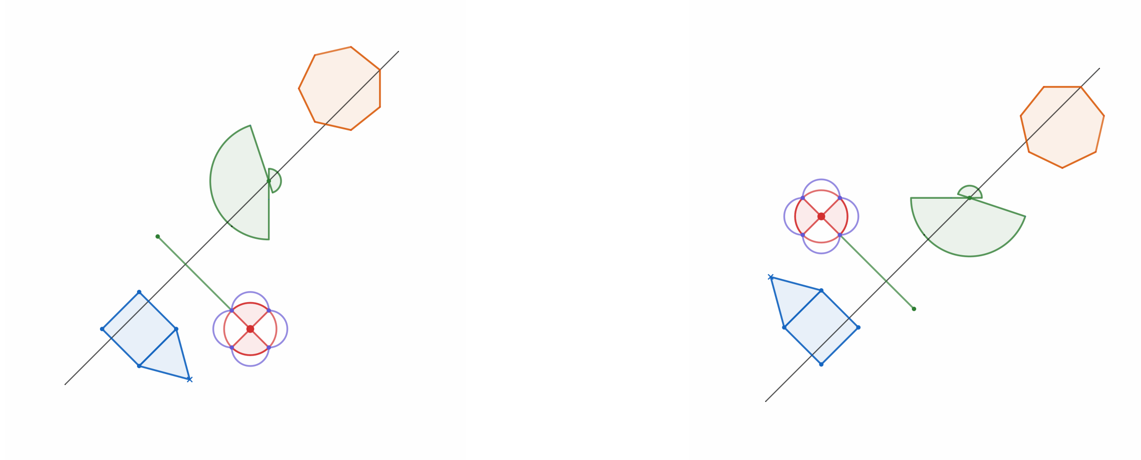


Figure 5.9

In Figure 5.10(A), we have drawn sets that passes through ℓ_θ; Figure 5.10(B) shows their images under Ref_θ.



(A) **(B)**

Figure 5.10

We can now follow the same two step process as we did for rotations to see that Ref_θ is a linear transformation and to find its matrix.

Step 1. Ref_θ is a linear transformation.

We leave it to the reader to draw pictures illustrating that reflections carry lines through the origin proportionately to lines through the origin and similarly for parallelograms.

Step 2. We compute $\mathrm{Ref}_\theta(\mathbf{i})$ and $\mathrm{Ref}_\theta(\mathbf{j})$.

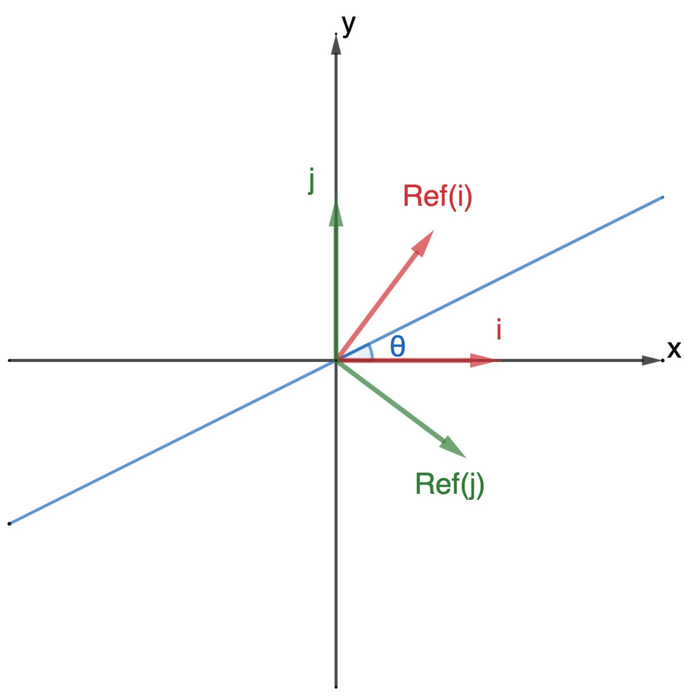


Figure 5.11

From Figure 5.11, we see that

$$\mathrm{Ref}_\theta(\mathbf{i}) = \begin{bmatrix} \cos(2\theta) \\ \sin(2\theta) \end{bmatrix} \tag{5.8}$$

The computation of $\mathrm{Ref}_\theta(\mathbf{j})$ is a bit trickier. One can figure out the angles directly from Figure 5.11. A way to avoid that computation, however, is to observe that $\mathrm{Ref}_\theta(\mathbf{j}) \perp \mathrm{Ref}_\theta(\mathbf{i})$, as one can see from the picture. (In fact, reflections always preserves shapes, so the angle between the reflected vectors is the same as the angle between the original vectors.) Since there are only two unit vectors orthogonal to $\langle \cos(2\theta), \sin(2\theta) \rangle$, namely $\pm\langle -\sin(2\theta), \cos(2\theta) \rangle$, one only has to determine the sign. We get

$$\mathrm{Ref}_\theta(\mathbf{j}) = \begin{bmatrix} \sin(2\theta) \\ -\cos(2\theta) \end{bmatrix}$$

We conclude:

$$[\mathrm{Ref}_\theta] = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix} \tag{5.9}$$

(Aside: You might notice that if we were to change the signs of the entries in the second column, we would get the matrix for $\mathrm{Rot}_{2\theta}$! Rotations and reflections are clearly different, so this confirms that we made the correct choice of sign when we computed $\mathrm{Ref}_\theta(\mathbf{j})$.)

Remark 5.37

The analogous concept in 3 dimensions is reflection across a plane in $\mathbb{R}^3$. For example, when you look at yourself in a mirror, you are seeing your reflection across the plane of the mirror as in Figure 5.12. Here the plane of reflection (the mirror) is perpendicular to the plane of the paper. If we think of the x-axis as coming towards us out of the paper, the y-axis as horizontal and the z-axis as vertical, then we are reflecting across the xz-plane.

Figure 5.12

5.4.3. Section summary

- We saw that rotations and reflections are linear transformations by showing geometrically that they satisfy the two linearity properties.
- We then found their representing matrices by finding the columns $T(\mathbf{i})$ and $T(\mathbf{j})$.
- Now that we have their representing matrices, we can rotate any vector (or point) about the origin or reflect any vector (or point) across any line through the origin and write down the result explicitly.

Exercises

Exercise 5.4.1 *Draw the capital letter R with its lower left corner at the origin. Then:*

(a) *Draw it's image under* Rot_θ*, where* $\theta = \frac{2\pi}{3}$.

(b) *Draw it's image under* Ref_θ*, where* $\theta = \frac{2\pi}{3}$.

Exercise 5.4.2

(a) *Write down the representing matrix of the rotation* $\mathrm{Rot}_{\frac{\pi}{4}}$.

(b) *Find the result of rotating each of the points* $(1,3)$ *and* $(-5,4)$ *through an angle of* $\frac{\pi}{4}$ *about the origin.*

(c) *Write down the representing matrix of the reflection* $\mathrm{Ref}_{\frac{5\pi}{6}}$.

(d) *Find the result of reflecting each of the points* $(1,3)$ *and* $(-5,4)$ *across the line* $\ell_{\frac{5\pi}{6}}$.

Exercise 5.4.3 *Find the result of rotating the point* $(2,5)$ *through an angle of* $\frac{4\pi}{3}$ *about the origin.*

Exercise 5.4.4 *Find the result of reflecting the points* $(1,3)$ *and* $(-5,4)$ *across the line* $y = \sqrt{3}x$.

Exercise 5.4.5 *For each of the following, identify the linear transformation whose representing matrix is given. If it is a rotation, specify the angle of rotation. If it is a reflection across a line, indicate the angle that the line makes with the positive x-axis.*

(a) $\begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$

(b) $\begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

Exercise 5.4.6 *Draw pictures similar to that in Figure 5.6 to illustrate geometrically that reflection across a line ℓ through the origin is a linear transformation. For the line ℓ, choose any line other than the x or y-axes. Be sure to illustrate* both *the linearity properties, using separate drawings for each of the two properties.*

5.5 What the Determinant of the Representing Matrix Tells Us

Prerequisite 5.38

Before reading this section, it is important to recall how to compute the area of a parallelogram $\text{Par}(\mathbf{v},\mathbf{w})$ *in* $\mathbb{R}^2$ *with sides given by vectors* $\mathbf{v} = \langle a,b\rangle$ *and* $\mathbf{w} = \langle c,d\rangle$*. You have learned that the area is the absolute value of the determinant of the matrix with rows* $\mathbf{v}$ *and* $\mathbf{w}$*, i.e, the matrix* $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$*. Thus the area is* $|ad-bc|$*. While you learned to make* $\mathbf{v}$ *and* $\mathbf{w}$ *the* rows *of the matrix, you get the same result if you instead make them the* columns*, since* $\det(\left[\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right]) = ad-bc = \det(\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right])$.

Outcomes

- *Understand what the determinant of the representing matrix of a linear transformation T tells you about how the area of a region in the domain and the area of its image are related.*
- *Understand what the determinant tells you about orientation.*

If $T:\mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation, then $[T]$ is a 2×2 matrix, and we can compute its determinant. We will see that the determinant contains a lot of information about the linear transformation.

5.5.1. Area

Let's return to the example of the linear transformation we looked at in Subsection 5.2.4. We have

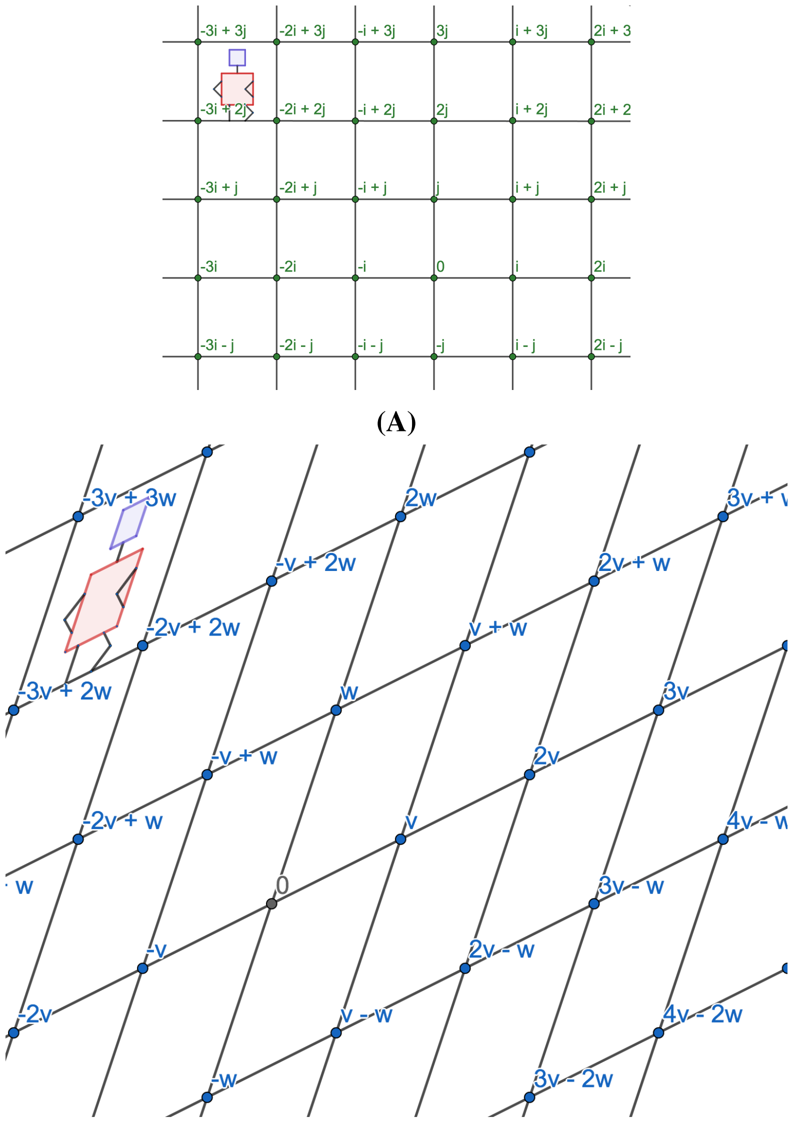
$$[T] = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Letting $\mathbf{v} = \langle 2,1\rangle$ and $\mathbf{w} = \langle 1,3\rangle$ (the two column vectors of $[T]$, we saw that T maps the grid of unit squares to the grid of parallelograms $\text{Par}(\mathbf{v},\mathbf{w})$, i.e., parallelograms with sides $\mathbf{v}$ and $\mathbf{w}$, as in Figure 5.13. (This is the same illustration that appeared in Subsection 5.2.4, repeated here for convenience.)

As noted in the Prerequisite 5.38, we have

$$\text{Area of } \operatorname{Par}(\mathbf{v},\mathbf{w}) = |\det(\left[\begin{smallmatrix} 2 & 1 \\ 1 & 3 \end{smallmatrix}\right])| = |\det([T])| = 5.$$

Thus each of the parallelograms in Figure 5.1(B) has area 5. In other words, T maps the unit squares in the first grid to parallelograms of area 5.



(A)

(B)

Figure 5.13

One can in fact say much more:

- The person shown in the second grid is five times as big as the person in the first grid.
- More generally, if you take any region R in the domain (e.g., a disk, a triangle,,,,), the area of the image of R will be 5 times larger than the area of R.

Theorem 5.39: Areas of Images

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be any linear transformation, and let R be any region in $\mathbb{R}^2$. Then the area of the image of R under T is given by $|\mathrm{Det}[T]|Area(R)$. In words, to get the area of the image of R, multiply the area of R by the absolute value of the determinant of $[T]$.

The analogous theorem with "area" replaced by "volume" holds for linear transformations $T : \mathbb{R}^3 \to \mathbb{R}^3$.

Example 5.40

- *If you visualize a rotation, you can see that the area of a region doesn't change when you rotate it. Let's see that this observation is consistent with Theorem 5.39. In Subsection 5.4.1, we saw that the representing matrix of* Rot_θ *is* $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$. *We have*

$$\det \begin{bmatrix} cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} = \cos^2(\theta) + \sin^2(\theta) = 1.$$

  Thus Theorem 5.39 tells us that when we rotate a region R, the area doesn't change, just as our geometric intuition told us.

- *Similarly, we saw in Subsection 5.4.2, that reflection across a line through the origin has representing matrix of the form* $\det(\mathrm{Ref}_\theta) = \det \begin{bmatrix} cos(2\theta) & \sin(2\theta) \\ -\sin(2\theta) & \cos(2\theta) \end{bmatrix}$. *We then get*

$$\det(\mathrm{Ref}_\theta) = -1 \;\; so \;\; |\det(\mathrm{Ref}_\theta)| = 1.$$

  This again agrees with our geometric understanding: reflecting a region across a line doesn't change its area.

Example 5.41

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation. We saw in Theorem 5.20 that the range of T is the subspace of $\mathbb{R}^2$ spanned by the columns of $[T]$. If the two columns of $[T]$ are parallel (and not both zero), then the range is a line. The image of any region in the domain must be contained in this line and thus have area zero. Let's check whether this agrees with Theorem 5.39. Since the two columns of $[T]$ are parallel, $[T]$ is of the form

$$[T] = \begin{bmatrix} a & ka \\ c & kc \end{bmatrix}$$

so

$$\det([T]) = a(kc) = c(ka) = 0.$$

Thus Theorem 5.39 confirms that the image of any region has area zero.

5.5.2. Orientation

The sign of the determinant also gives us valuable information. Note that the sign only makes sense if the determinant is non-zero, i.e., if T has rank 2. (See the previous example.) This is the case precisely when $T(\mathbf{i})$ and $T(\mathbf{j})$ are non-parallel.

Definition 5.42

We say that an ordered pair $(\mathbf{v}, \mathbf{w})$ of non-parallel vectors is positively oriented if the smaller of the two angles from $\mathbf{v}$ to $\mathbf{w}$ is the one that goes counterclockwise as in Figure 5.14). We say it is negatively oriented if the smaller angle is the one that goes clockwise.

For example, the ordered pair $(\mathbf{i}, \mathbf{j})$ is positively oriented and the ordered pair $(\mathbf{j}, \mathbf{i})$ is negatively oriented.

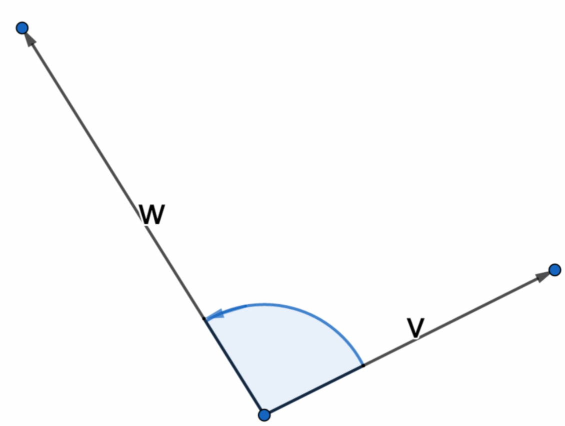


Figure 5.14

Definition 5.43: Orientation-preserving/reversing

A linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ of rank two is said to be orientation-preserving *(respectively,* orientation-reversing*) if the ordered pair $(T(\mathbf{i}), T(\mathbf{j}))$ is positively oriented (respectively, negatively oriented).*

For linear transformations of rank less than two, we can't talk about orientation since $T(\mathbf{i})$ and $T(\mathbf{j})$ are parallel in that case. (Recall these are the columns of $[T]$.)

Example 5.44

By looking at drawings, we see:

- *Let T be the linear transformation with representing matrix* $[T] = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$ *and illustrated in Figure 5.13. Recall that $T(\mathbf{i})$ and $T(\mathbf{j})$ are the vectors labelled $\mathbf{v}$ and $\mathbf{w}$, respectively, in Figure 5.13. Observe that $\mathbf{w}$ is counterclockwise of $\mathbf{v}$, so T is orientation-preserving.*
- *Rotations Rot_θ are orientation-preserving.*
- *Reflections Ref_θ are orientation-reversing. (This is analogous to the fact that your reflection in a mirror is oriented oppositely to yourself.)*

Theorem 5.45: Sign of Determinant and Orientation

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation.

1. *If $\mathrm{Det}[T] > 0$, then T is orientation-preserving.*
2. *If $\mathrm{Det}[T] < 0$, then T is orientation-reversing.*

The reader is encouraged to compute the determinants of the representing matrices of the linear transformations in Example 5.44 and verify that they are consistent with Theorem 5.45.

While we only have to check two vectors $T(\mathbf{i})$ and $T(\mathbf{j})$ to determine whether T is orientation-preserving or reversing, it turns out that orientation-preserving (respectively, reversing) linear transformations preserve, respectively reverse, orientation in many ways as the next two propositions indicate.

Proposition 5.46: Circles to Ellipses

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation of rank 2. Then the image under T of every circle in $\mathbb{R}^2$ is an ellipse. If T is orientation-preserving (respectively, reversing), then as a point moves counterclockwise around a circle, the image under T moves counterclockwise (respectively, clockwise) around the image ellipse.

Circles are themselves ellipses. For rotations and reflections, the image of any circle is actually a circle. For the linear transformation in Subsection 5.2.4, on the other hand, the images of circles are not circles but, as guaranteed by the proposition, they are ellipses.

Proposition 5.47

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation of rank 2. Suppose that T is orientation-preserving (respectively, reversing). If $(\mathbf{v}, \mathbf{w})$ is any *ordered pair of non-parallel vectors in $\mathbb{R}^2$, then $(T(\mathbf{v}), T(\mathbf{w}))$ has the same (respectively, opposite) orientation as $(\mathbf{v}, \mathbf{w})$.*

Thus an orientation-preserving linear transformation preserves the direction of *all* angles.

5.5.3. Section summary

Let $T : \mathbf{T}^2 \to \mathbb{R}^2$ be any linear transformation.

- Given any region R in the domain, let's write $T(R)$ for the image of R. Then to find the area of $T(R)$, you just need to multiply the area of R by $|\det(T)|$. In particular, the area of every region (disk, parallelogram, triangle,....) in $\mathbb{R}^2$ expands or shrinks by the same ratio $|\det(T)|$.
- If $\det(T) > 0$, then T preserves the orientation of all angles. In other words, if $\mathbf{w}$ is counterclockwise from $\mathbf{v}$, then $T(\mathbf{w})$ is counterclockwise from $T(\mathbf{v})$. If $\det(T) < 0$, then T reverses the orientation of all angles.

Exercises

Exercise 5.5.1 *Let T be a linear transformation with representating matrix*

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

.

(a) Find the area of the image under T of the unit square.

(b) Find the area of the image under T of a disk whose radius is 3.

(c) Decide whether T is orientation-preserving or reversing by computing the determinant.

(d) Verify your answer to part (c) by drawing the vectors $T(\mathbf{i})$ and $T(\mathbf{j})$. Explain.

5.6 Composition of Linear Transformations

Definition 5.48: Composition of Functions

Given functions $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^m \to \mathbb{R}^k$ (so the range of F is contained in the domain of G), we define the composition of G with F, denoted $G \circ F$ by

$$G \circ F(\mathbf{x}) = G(F(\mathbf{x})).$$

You are already familiar with the composition of functions from $\mathbb{R}$ to $\mathbb{R}$. E.g., if $f(x) = x^2$ and $g(x) = \sin(x)$, then

$$g \circ f(x) = g(f(x)) = g(x^2) = \sin(x^2).$$

When you first learned about compositions of real-valued functions on $\mathbb{R}$, you might have found it easier to change the name of the variable in the second function, say writing $g(u) = \sin(u)$ and then writing $g(f(x)) = g(u) = \sin(u)$ where $u = x^2$.

Example 5.49

Let

$$F(x,y) = \begin{bmatrix} x^2y \\ x+2y \\ y^3 \end{bmatrix} \text{ and } G(x,y,z) = \sin(x+z)y^4.$$

The range of F is contained in $\mathbb{R}^3$, which is the domain of G; thus we can form the composition $G \circ F$. Find this composition explicitly.

Solution. We need to compute $G \circ F(x,y) = G(F(x,y))$. You might find it easier to rename the variables in the domain of G, say u, v, w, and write $G(u,v,w) = \sin(u+w)v^4$. Then $G(F(x,y)) = G(u,v,w)$ where $u = x^2y$, $v = x+2y$, and $w = y^3$. This yields,

$$G(F(x,y)) = G(x^2y, x+2y, y^3) = \sin(x^2y+y^3)(x+2y)^4.$$

♠

For linear transformations, the representing matrices simplify the process of composition.

Theorem 5.50: Composing Linear Transformation

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^k$ are linear transformations. Then the composition $S \circ T$ is also a linear transformation, and its representing matrix is the product of the representing matrices of S and T, i.e.,

$$[S \circ T] = [S][T].$$

Proof. We have

$$S \circ T(\mathbf{x}) = S(T(\mathbf{x})) = [S]\left([T][\mathbf{x}]\right).$$

Since matrix multiplication is associative, we have $[S]\left([T][\mathbf{x}]\right) = \left([S][T]\right)[\mathbf{x}]$. Substituting this into the previous equation, we obtain

$$S \circ T(\mathbf{x}) = \left([S][T]\right)[\mathbf{x}].$$

Thus $S \circ T$ can be expressed as multiplication by the matrix $[S][T]$. This tells us that $S \circ T$ is a linear transformation and its representing matrix is $[S][T]$.

♠

Example 5.51

Let

$$T(x,y) = \begin{bmatrix} 5x+3y \\ x+y \end{bmatrix} \text{ and } S(x,y,z) = \begin{bmatrix} 2x+y+4z \\ x+2z \end{bmatrix}$$

Does either of $S \circ T$ or $T \circ S$ make sense? If so, compute it.

Solution. First a note about the language being used here: When we ask whether a composition $F \circ G$ "makes sense", we are asking whether the range of G is contained in the domain of F. If the answer is yes, then the composition makes sense and we can compute it. If the answer is no, then we can't form the composition $F \circ G$.

In this example, we have $T : \mathbb{R}^2 \to \mathbb{R}^2$ and $S : \mathbb{R}^3 \to \mathbb{R}^2$. The composition $S \circ T$ does not make sense, since the range of T lies in $\mathbb{R}^2$, whereas the domain of S is $\mathbb{R}^3$. However the composition $T \circ S$ does make sense. Writing down the representing matrices of T and S and multiplying, we get

$$[T \circ S] = \begin{bmatrix} 5 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 4 \\ 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 13 & 5 & 26 \\ 3 & 1 & 6 \end{bmatrix}$$

Thus

$$T \circ S(x,y,z) = \begin{bmatrix} 13x+5y+26z \\ 3x+y+6z \end{bmatrix}$$

♠

Now for a more interesting example:

Example 5.52: Composing Rotations

Consider two rotations of $\mathbb{R}^2$ about the origin, say Rot_α and Rot_β as in Definition 5.34. Since the domain and range of both rotations are $\mathbb{R}^2$, both compositions Rot_α and Rot_β make sense. Find them.

Solution. Before we actually compute, let's THINK GEOMETRICALLY! You actually know the answer without doing any computation! (Think about it on your own before reading the next sentence.)

If you first rotate $\mathbf{x}$ through angle β and then rotate the result through angle α, it's the same as rotating $\mathbf{x}$ through angle $\alpha+\beta$ as in Figure 5.15. Thus

$$[\mathrm{Rot}_\alpha \circ \mathrm{Rot}_\beta] = \begin{bmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{bmatrix} \tag{5.10}$$

by Equation (5.7) in Subsection 5.4.1

Now let's compute the composition using matrix multiplication and compare with our answer above:

$$[\mathrm{Rot}_\alpha \circ \mathrm{Rot}_\beta] = [\mathrm{Rot}_\alpha][\mathrm{Rot}_\beta] = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} \cos(\beta) & -\sin(\beta) \\ \sin(\beta) & \cos(\beta) \end{bmatrix}$$

$$= \begin{bmatrix} \cos(\alpha)\cos(\beta)-\sin(\alpha)\sin(\beta) & -\cos(\alpha)\sin(\beta)-\sin(\alpha)\cos(\beta) \\ \sin(\alpha)\cos(\beta)+\cos(\alpha)\sin(\beta) & -\sin(\alpha)\sin(\beta)+\cos(\alpha)\cos(\beta) \end{bmatrix} \tag{5.11}$$

Stare at this for a second and compare with Equation (5.10). What do you notice? If you ever forget the trig formulas for the sine and cosine of the sum of two angles, rotation matrices are a great way to reconstruct them!

♠

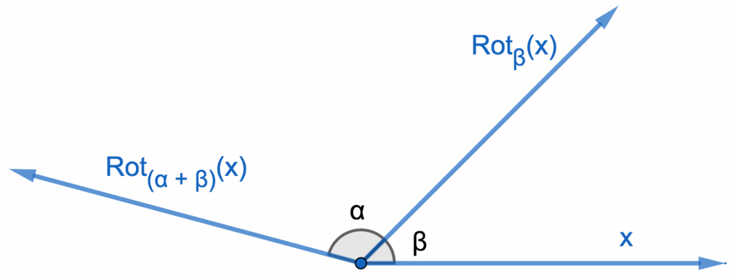


Figure 5.15

We note – thinking geometrically again – that it doesn't matter which way you compose the two rotations. If you instead rotate first through angle α and then through angle β, you'll get the same result as in the example. In terms of the matrices, this says that the representing matrices of different rotations about the origin commute with each other, as you can easily check. In general though, when composing two linear transformations, you need to be careful to multiply the matrices in the correct order. If only one of the compositions $S \circ T$ or $T \circ S$ makes sense, then the size of the matrices will only allow them to be multiplied in the correct order. But extra caution is needed when both make sense.

Exercises

Exercise 5.6.1 *Let*

$$R(x,y) = \begin{bmatrix} 2x+y \\ x-y \\ 5x \end{bmatrix} \quad S(x,y) = \begin{bmatrix} -3x-y \\ x+y \end{bmatrix} \quad \text{and} \quad T(x,y,z) = \begin{bmatrix} x+y+z \\ y-z \end{bmatrix}$$

Decide whether each of the following compositions makes sense.

(a) $R \circ S$

(b) $S \circ R$

(c) $R \circ T$

(d) $T \circ S$

(e) $R \circ S \circ T$

(f) $T \circ S \circ R$

Exercise 5.6.2 *Find the resulting linear transformation and write down its standard representating matrix for each of the compositions in Exercise 5.6.1 that makes sense.*

Exercise 5.6.3 *Let*

$$T(x,y) = 3x+4y \text{ and } S(x,y,z) = \begin{bmatrix} x+2y \\ 3x+y+2z \end{bmatrix}.$$

(a) *Compute $T \circ S(x,y,z)$ directly without using matrices.*

(b) *Write down the representing matrices $[T]$ and $[S]$ and use matrix mutliplication to compute $[T \circ S]$.*

(c) *Verify that your answers to (a) and (b) agree.*

Exercise 5.6.4 *Let T be the reflection of $\mathbb{R}^2$ across the line ℓ_θ where $\theta = \frac{\pi}{3}$. Find $T \circ T$.*

Exercise 5.6.5 *Show that for any reflection T of $\mathbb{R}^2$, $T \circ T$ is the identity transformation. (See Example 5.16 for the definition of the identity transformation.)*

Exercise 5.6.6 *Consider two rotations of $\mathbb{R}^2$ about the origin, Rot_α and Rot_β, where $\alpha = \frac{\pi}{4}$ and $\beta = -\frac{\pi}{4}$. Use matrix multiplication to show that $\mathrm{Rot}_\alpha \circ \mathrm{Rot}_\beta$ is the identity transformation. Then give a geometric explanation, i.e., explain how you could have known that this composition would give you the identity without doing the computation.*

Exercise 5.6.7 *Let $\mathbf{v} = \langle 1,2,1 \rangle$ and let $T(x,y,z) = \mathrm{proj}_{\mathbf{v}} \langle x,y,z \rangle$.*

(a) *Write down the standard representing matrices of T and of $T \circ T$.*

(b) *Compare the two matrices in (a). What do you notice? (Answer: they are the same! If you don't see this, go back and check your computation in part (a).)*

(c) *Give a geometric explanation of part (b). I.e., use the geometry of projections to explain how you could know that the relationship in part (b) would hold before even computing.*

Exercise 5.6.8 *Let T be the linear transformation in problem 7, let $\mathbf{w} = \langle 2,-1,0 \rangle$, and let $S(x,y,z) = \mathrm{proj}_{\mathbf{w}} \langle x,y,z \rangle$.*

(a) *Write down the standard representing matrices of S and of $S \circ T$.*

(b) *What do you notice about $S \circ T$. (Answer: it's zero! If you didn't get this, go back and check your computations.)*

(c) *Give a geometric explanation of the observation in part (b). I.e., using the geometry of projections, how could you know that the relationship in part (b) would hold before even computing? (Hint: what is the angle between $\mathbf{v}$ and $\mathbf{w}$?)*

Exercise 5.6.9

(a) *Find the representing matrix of the linear transformation* $T = \mathrm{Rot}_{\frac{\pi}{6}} \circ \mathrm{Ref}_0 \circ \mathrm{Rot}_{(-\frac{\pi}{6})}$. *(Here* Ref_0 *denotes reflection across the x-axis.)*

(b) *Identify the linear transformation* T *in part (a). (It should be a reflection. Across what line?)*

(c) *Give a geometric explanation of parts (a) and (b). I.e., explain geometrically why the composition in part (a) should result in the reflection identified in part (b).*

5.7 Affine Transformations

We hope that the previous sections have conveyed the simplicity and geometric appeal of linear transformations. Even ignoring the especially nice ones like rotations, try to imagine trying to visualize a function such as $F(x,y) = (\sin(xy)e^y, \sqrt{x^2y - y^4})$ and contrast with the relative ease with which we visualized in Figure 5.1 the rather generic linear transformation in Subsection 5.2.4.

You might have wondered initially why a function such as $f(x) = 2x+5$ is not called a linear function or a linear transformation, given that its graph is a straight line. Why do we not allow constant terms? The answer is quite simple: if we add in constants then we loose all the advantages afforded by the fact that linear transformations are given by matrix multiplication. We also loose the linearity properties. (One still gets reasonably nice analogues of the linearity properties when constants are added in, but they're not as elegant and nice to work with.)

However, functions that are given by adding constants to linear transformations do have a special name. They are called affine transformations.

Definition 5.53: Affine transformations

An affine transformation $F : \mathbb{R}^n \to \mathbb{R}^m$ *is any function of the form* $F(\mathbf{x}) = T(\mathbf{x}) + \mathbf{b}$, *where* $T : \mathbb{R}^n \to \mathbb{R}^m$ *is a linear transformation and* $\mathbf{b}$ *is a fixed element of* $\mathbb{R}^m$.

Example 5.54

The function $F : \mathbb{R}^3 \to \mathbb{R}^2$ *given by*

$$F(x,y,z) = \begin{bmatrix} 2x+y-z+7 \\ x+z-5 \end{bmatrix}$$

is an affine transformation since

$$F(x,y,z) = \begin{bmatrix} 2x+y-z \\ x+z \end{bmatrix} + \begin{bmatrix} 7 \\ -5 \end{bmatrix} = T(x,y,z) + \mathbf{b}$$

where $T(x,y,z) = \left[\begin{smallmatrix} 2x+y-z \\ x+z \end{smallmatrix}\right]$ *and* $\mathbf{b} = \left[\begin{smallmatrix} 7 \\ -5 \end{smallmatrix}\right]$.

Example 5.55

The plane through the point $(1,2,4)$ *parallel to the plane* $z = 2x+3y$ *(or* $2x+3y-z=0$*) has normal vector* $\langle 2,3,-1\rangle$ *and thus is given by* $2(x-1)+3(y-2)-(z-4)=0$ *or*

$$z-4 = 2(x-1)+3(y-2).$$

We can rewrite this as

$$z = T(x,y)+b$$

where T *is the linear transformation* $T(x,y) = 2x+3y$ *and* $b = -4$.

Preview of Things to Come 5.56

In single variable calculus, you approximated a function $f:\mathbb{R}\to\mathbb{R}$ *near a point* x_0 *by its tangent line* $y-y_0 = m(x-x_0)$. *If you combine all the constant terms, you can write it in the form* $y = mx+b$, *so you're approximating* f *by an affine function* $h(x) = mx+b$. *(It's usually more convenient though to leave it in the form* $y-y_0 = m(x-x_0)$ *since you're working near the point* (x_0,y_0).*)*

The real work in finding the tangent line was in finding its slope $m = f'(x_0)$. *Once you have the slope, you can quickly write down the tangent line. We can rephrase this as saying that the real work was finding the* linear part $T(x) = mx$ *of the tangent approximation. Adding in* b *or writing in the form* $y-y_0 = m(x-x_0)$ *is easy from there. The effect of adding the constant just translates the line so it passes through the point* (x_0,y_0).

We will soon be defining derivatives of functions $F:\mathbb{R}^n\to\mathbb{R}^m$ *at points* $\mathbf{x}_0$ *of* $\mathbb{R}^n$. *They will be expressed as matrices, denoted* $[F'(\mathbf{x}_0)]$ *and can also be viewed as linear transformations. (If the dimension* n *and* m *of the domain and target space are both one, the derivative will be the familiar notion, just viewed as a* 1×1 *matrix.) Once we have the derivative, we can easily add the appropriate constant to the corresponding linear transformation in order to obtain an approximation of* F *near* $\mathbf{x}_0$.

Exercises

Exercise 5.7.1 *For each of the following functions, decide whether it is an affine transformation. If so, write it explicitly as the sum of a linear transformation and a constant vector* **b**.

(a) $F(x) = \begin{bmatrix} x+1 \\ 2x \\ 100x-50 \end{bmatrix}$

(b) $F(x) = \begin{bmatrix} x^2+x \\ -x \end{bmatrix}$

(c) $F(x,y) = x+y+1$

(d) $F(x,y) = \begin{bmatrix} xy+1 \\ -xy+1 \end{bmatrix}$

(e) $F(x,y,z) = \begin{bmatrix} x-y-1 \\ y-z-1 \\ z-x-1 \end{bmatrix}$

(f) $F(x,y,z) = \begin{bmatrix} xyz \\ z^2+1 \end{bmatrix}$

Exercise 5.7.2 *Express each of the following planes explicitly as the graph of an affine transformation $F : \mathbb{R}^2 \to \mathbb{R}$, i.e., write it in the form $z = T(x,y) + b$ where $T : \mathbb{R}^2 \to \mathbb{R}$ is a linear transformation and b is a real number.*

(a) $z - 5 = 2(x+1) + 4(y-2)$

(b) $3x + 4y + 5z = 10$

(c) $2(x-1) + 3y + 2(z-3) = 0$

Exercise 5.7.3 *Show that if $F(x,y) = T(x,y) + \mathbf{b}$ is an affine transformation with $\mathbf{b} \neq 0$, then F does not satisfy the linearity properties.*

6. Derivatives

6.1 Derivatives of real-valued functions of two variables

Motivation

In this section we will consider real-valued functions $f(x,y)$ of two variables. In Section 4.1, we defined directional derivatives, and in Section 4.8, we defined (at least informally) what it means for a surface to have a tangent plane. We then defined f to be differentiable at a point (x_0,y_0) if its graph has a tangent plane at $(x_0,y_0,f(x_0,y_0))$. However, we still haven't said what we mean by the derivative of f. We address that now.

Outcomes

- *Understand and be able to compute derivatives of real-valued functions of two variables.*
- *Learn an easy way to compute directional derivatives of differentiable functions.*
- *Be able to do word problems involving related rates.*

6.1.1. Introducing the derivative

Key Idea 6.1: What information should the derivative give us?

- *For a real-valued differentiable function f of one variable, you learned in single-variable calculus that the derivative of $f'(x_0)$ is the slope of the tangent line to the graph of f at $(x_0, f(x_0)$. Once you know the slope (and of course the point $(x_0, f(x_0))$), you have all the information you need to write down the tangent line*

$$y - y_0 = f'(x_0)(x - x_0). \tag{6.1}$$

- *For a differentiable function of two variables, we analogously want the derivative at (x_0, y_0) – along with the point $(x_0, y_0, f(x_0, y_0))$ – to tell us precisely the information needed to write down the tangent plane. As we have seen, any plane in $\mathbb{R}^3$ that is not vertical (i.e., not parallel to the z axis) can be written in the form*

$$z - z_0 = a(x - x_0) + b(y - y_0).$$

In contrast to lines where we needed only a single real number (the slope), we now need two real numbers a, b in order to obtain the tangent plane. Thus, in contrast to the functions studied in single variable calculus, the derivative of f will not be a single real number but instead must provide us two real numbers a, b.

Assume that f is differentiable at (x_0, y_0). By Theorem 4.12, its tangent plane at (x_0, y_0, z_0), where $z_0 = f(x_0, y_0)$, is given by

$$z - z_0 = f_x(x_0, y_0)\,(x - x_0) + f_y(x_0, y_0)\,(y - y_0).$$

We can write the tangent plane in matrix form:

$$[z - z_0] = \begin{bmatrix} f_x(x_0, y_0) & f_y(x_0, y_0) \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \tag{6.2}$$

Notice that the matrix $\begin{bmatrix} f_x(x_0, y_0) & f_y(x_0, y_0) \end{bmatrix}$ gives us the two items of information we need to determine the tangent plane and thus plays the same role as the slope of the tangent line in Equation (6.1).

Definition 6.2: Derivative

- *Let $f(x,y)$ be a real-valued function and assume that f is differentiable at (x_0,y_0). Then we define the derivative of f at (x_0,y_0), denoted both by $[Df_{(x_0,y_0)}]$ and $[f'(x_0,y_0)]$ to be the matrix*

$$[Df_{(x_0,y_0)}] = [f'(x_0,y_0)] = \begin{bmatrix} f_x(x_0,y_0) & f_y(x_0,y_0) \end{bmatrix}.$$

- *We will denote by $Df_{(x_0,y_0)}$ the linear transformation whose matrix is $[Df_{(x_0,y_0)}] = \begin{bmatrix} f_x(x_0,y_0) & f_y(x_0,y_0) \end{bmatrix}$.*

Aside: In more advanced texts, the derivative is defined to be the linear transformation $Df_{(x_0,y_0)}$, and the matrix $[Df_{(x_0,y_0)}]$ is then called the "derivative matrix" or the "Jacobian matrix" (after the mathematician Jacobi).

If we write

$$[\mathbf{x}] = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad [\mathbf{x}_0] = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

then the equation (6.2) of the tangent plane becomes

$$\boxed{[z - z_0] = [f'(\mathbf{x}_0)][\mathbf{x} - \mathbf{x}_0]}$$

which makes clearer the analogy with the tangent line in Equation (6.1).

Example 6.3

For the function $f(x,y) = x^2y^3$, we have $\begin{bmatrix} f_x & f_y \end{bmatrix} = \begin{bmatrix} 2xy^3 & 3x^2y^2 \end{bmatrix}$. Thus, for example,

$$[f'(2,1) = \begin{bmatrix} 4 & 12 \end{bmatrix}$$

and the tangent plane to the graph $z = f(x,y)$ at $(2,1,4)$ can be written

$$[z-4] = \begin{bmatrix} 4 & 12 \end{bmatrix} \begin{bmatrix} x-2 \\ y-1 \end{bmatrix}$$

6.1.2. Directional Derivatives Revisited

Our definitions of differentiability and of tangent plane to the graph $z = f(x,y)$ required that all the tangent lines to smooth curves in the surface lie in a single plane. We have not yet taken full advantage of that information.

Let's first consider directional derivatives. Let $\mathbf{u} = \langle u_1, u_2 \rangle$ be a unit vector. By taking the tangent vector to an appropriate curve in the surface $z = f(x,y)$, we saw in Proposition 4.11 that the vector $\langle u_1, u_2, D_{\mathbf{u}} f(x_0,y_0) \rangle$ is parallel to the tangent plane. Thus this vector is orthogonal to the normal vector $\mathbf{n} = \langle -f_x(x_0,y_0), -f_y(x_0,y_0), 1 \rangle$ of the plane, so the dot product is zero:

$$\langle -f_x(x_0,y_0), -f_y(x_0,y_0), 1 \rangle \cdot \langle u_1, u_2, D_{\mathbf{u}} f(x_0,y_0) \rangle = 0$$

yielding

$$-f_x(x_0,y_0)u_1 - f_y(x_0,y_0)u_2 + D_{\mathbf{u}}f(x_0,y_0) = 0. \tag{6.3}$$

Thus

Theorem 6.4

If f is differentiable at (x_0,y_0), then

$$D_{\mathbf{u}}f(x_0,y_0) = f_x(x_0,y_0)u_1 + f_y(x_0,y_0)u_2 = [f'(x_0,y_0)]\begin{bmatrix}u_1\\u_2\end{bmatrix} \tag{6.4}$$

Proof. The first equality is immediate from Equation (6.3), and the second equality follows from Definition 6.2 of the derivative $[f'(x_0,y_0)]$. ♠

Whoa!! Strange!! Equation (6.4) enables us to compute all the directional derivatives of a differentiable function once we know the two partial derivatives. Viewing the positive x and y axis as pointing east and north, respectively, and z as elevation, this says that to compute the slope of the "mountain" (the graph of f) at $(x_0,y_0,f(x_0,y_0))$ in any direction $\langle u_1, u_2\rangle$, we only need to know the slope in two specific directions (east and north)! That probably doesn't jive with your experience when hiking on a mountain. If you are climbing up Mount Washington, the slope heading east and the slope heading north from a point on the mountain don't determine the slope heading northeast for example. So the mountains you climb on probably aren't differentiable! We commented in Section 4.8 that the requirement to have a tangent plane and thus to be differentiable was extremely demanding. This theorem is an illustration.

Example 6.5

Let $f(x,y) = x^2y^3$, $(x_0,y_0) = (2,1)$, and $\mathbf{u} = \langle \frac{4}{5}, \frac{3}{5}\rangle$. Then

$$D_{\mathbf{u}}f(x_0,y_0) = \begin{bmatrix}4 & 12\end{bmatrix}\begin{bmatrix}\frac{4}{5}\\ \\ \frac{3}{5}\end{bmatrix} = \frac{52}{5}$$

The reader is encouraged to compute the directional derivative directly and check that it agrees with this result.

6.6: Interpretation of the Derivative

From Theorem 6.4, we see that the derivative $[Df_{(x_0,y_0)}] = [f'(x_0,y_0)]$ captures not only the two partial derivatives but in fact tells us all the directional derivatives at (x_0,y_0). One way to interpret the theorem is that the linear transformation $Df_{(x_0,y_0)}$ associated with the derivative as in Definition 6.2 satisfies

$$Df_{(x_0,y_0)}(\mathbf{u}) = D_{\mathbf{u}}f(x_0,y_0).$$

In other words, if we apply the linear transformation to a unit vector $\mathbf{u}$, we get the directional derivative in the direction $\mathbf{u}$.

Even more is true. Let $\langle x(t), y(t)\rangle$ be any smooth path in the domain of f passing through (x_0,y_0) at time t_0 and write

$$\mathbf{r}(t) = \langle x(t),\, y(t),\, f(x(t),y(t))\rangle$$

for the corresponding path C in the graph $z = f(x,y)$. We have

$$\mathbf{r}'(t_0) = \langle x'(t_0),\, y'(t_0),\, \frac{d}{dt}_{|t=t_0} f(x(t),y(t))\rangle.$$

The tangent line to the curve C at the point $(x_0,y_0,f(x_0,y_0))$ must lie in the tangent plane to the graph of f. Thus $\mathbf{r}'(t_0)$ is parallel to the tangent plane, so it is orthogonal to the normal vector $\langle -f_x(x_0,y_0),\, -f_y(x_0,y_0),\, 1\rangle$ for the tangent plane. A similar computation as we did for directional derivatives in Theorem 6.4 thus gives:

Theorem 6.7

Suppose that f is differentiable at (x_0,y_0). Let $\langle x(t),y(t)\rangle$ be a smooth path in the domain of f passing through (x_0,y_0) at time t_0. Then

$$\frac{d}{dt}_{|t=t_0} f(x(t),y(t)) = f_x(x_0,y_0)x'(t_0) + f_y(x_0,y_0)y'(t_0) = [f'(x_0,y_0)]\begin{bmatrix} x'(t_0) \\ y'(t_0) \end{bmatrix} \tag{6.5}$$

The next example illustrates how we can apply Theorem 6.7 to problems involving "related rates" in which various related quantities are changing over time.

Example 6.8

Suppose that you and a friend are planning to meet at the Hanover Inn on the intersection of Wheelock and Main Streets. You are bicycling on Wheelock and your friend is bicycling on Main St., both heading towards the Inn. We assume (though this isn't really the case!) that the two streets are perfectly straight and that they meet at a right angle so that at each given instant, your position, your friend's position and the Hanover Inn form the vertices of a right triangle. Suppose, at a certain moment t_0, that you are 4 miles from the intersection and bicycling 10 mph, while your friend is 3 miles from the intersection and bicycling at 8 mph. How rapidly is the distance between you and your friend changing at that instant?

Solution. *Let x denote your distance from the meeting point and y your friend's distance. Then the distance $f(x,y)$ between you is the length of the hypotenuse of the right triangle with sides x and y, namely*

$$f(x,y) = \sqrt{x^2+y^2}.$$

Note that x and y both depend on t. Your positions at time t_0 are $x_0 = 4$ miles and $y_0 = 3$ miles, and your rates of change are $x'(t_0) = -10$ mph, and $y'(t_0) = -8$ mph. (The derivatives x' and y' are negative since you are getting closer to the meeting point.) We want to know the rate of change of $f(x(t),y(t))$ at time t_0. Applying Theorem 6.7, we have

$$\frac{d}{dt}_{|t=t_0} f(x(t),y(t)) = f_x(4,3)x'(t_0) + f_y(4,3)y'(t_0) = \frac{4}{5}(-10) + \frac{3}{5}(-8) = -\frac{64}{5} \text{ mph.}$$

♠

6.1.3. Section Summary

- For $f:\mathbb{R}^2\to\mathbb{R}$ a differentiable function, the derivative of f at (x_0,y_0) is the matrix $[Df_{(x_0,y_0)}]=[f'(x_0,y_0)]=\begin{bmatrix} f_x(x_0,y_0) & f_y(x_0,y_0)\end{bmatrix}$.
- Writing $[\mathbf{x}]=[\begin{smallmatrix}x\\y\end{smallmatrix}]$ and $[\mathbf{x}_0]=[\begin{smallmatrix}x_0\\y_0\end{smallmatrix}]$, then the tangent plane to the graph $z=f(x,y)$ at $(x_0,y_0,f(x_0,y_0))$ can be written as $z-z_0=[f'(\mathbf{x}_0)][\mathbf{x}-\mathbf{x}_0]$, in analogy with the equation of the tangent line to the graph $y=f(x)$ of a function of one variable.
- For differentiable functions f, we can compute the directional derivative in the direction $\mathbf{u}=\langle u_1,u_2\rangle$ by
$$D_{\mathbf{u}}f(x_0,y_0)=[f'(x_0,y_0)]\begin{bmatrix}u_1\\u_2\end{bmatrix}$$
- For differentiable functions, we have $\frac{d}{dt}_{|t=t_0}f(x(t),y(t))=f_x(x_0,y_0)x'(t_0)+f_y(x_0,y_0)y'(t_0)$. We can use this to solve related rates problems.

6.1.4. Appendix to Section 6.1: Derivatives as limits

For functions of one variable, you learned that f is differentiable if

$$f'(x_0):=\lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}$$

exists. One can rewrite this as

$$\lim_{x\to x_0}\frac{f(x)-f(x_0)-f'(x_0)(x-x_0)}{x-x_0}=0. \tag{6.6}$$

Writing the tangent line as

$$y=L(x)=f(x_0)+f'(x_0)(x-x_0),$$

equation (6.6) can be rewritten as

$$\lim_{x\to x_0}\frac{f(x)-L(x)}{x-x_0}=0. \tag{6.7}$$

We could replace the denominator by $|x-x_0|$; this wouldn't affect the fact that the limit is zero. Thus this equation is saying that the difference $f(x)-L(x)$ is going to zero VERY quickly as $x\to x_0$; the difference $f(x)-L(x)$ is much smaller in magnitude than the distance $|x-x_0|$ from x_0 to x when x is small. That's why the tangent line is a good approximation of the function.

We took a reverse approach for functions of two variables, defining a real-valued function $f(x,y)$ of two variables to be differentiable at (x_0,y_0) if the graph $z=f(x,y)$ has a tangent plane at $(x_0,y_0,f(x_0,y_0))$. We now express this condition in terms of a limit.

Let f be a real-valued function of two variables, and assume that you've already determined that the two partial derivatives $f_x(x_0,y_0)$ and $f_y(x_0,y_0)$ exist. It then makes sense to write down the equation of the plane through $(x_0,y_0,f(x_0,y_0))$ containing the tangent lines to the curves in the graph in the "east-west and north-south directions". As before, this plane can be expressed as $z=L(x,y)$ where

$$L(x,y)=f(x_0,y_0)+f_x(x_0,y_0)(x-x_0)+f_y(x_0,y_0)(y-y_0). \tag{6.8}$$

This is the only candidate for the tangent plane. However, this plane might not actually be tangent to the graph of f; it may not contain the tangent lines to other curves in the graph of f through $(x_0, y_0, f(x_0, y_0))$. In fact, other curves might not even have tangent lines! Put another way, the plane might not really be approximating the surface $z = f(x,y)$ very well.

How close of an approximation is good enough for $z = L(x,y)$ to actually be a tangent plane to the graph $z = f(x,y)$? As in the case of functions of one variable, we need that the difference $f(x,y) - L(x,y)$ go to zero very fast as $(x,y) \to (x_0, y_0)$, so fast that when we divide by the distance from (x,y) to (x_0, y_0), the quotient still goes to zero, i.e.,

$$\lim_{(x,y)\to(x_0,y_0)} \frac{f(x,y) - L(x,y)}{\sqrt{(x-x_0)^2 + (y-y_0)^2}} = 0.$$

Substituting in the expression for L given in Equation (6.8), we thus have:

Theorem 6.9

Suppose that the partial derivatives $f_x(x_0,y_0)$ and $f_y(x_0,y_0)$ exist. (We are not assuming that the partials are continuous.) Then f is differentiable at (x_0,y_0) if and only if

$$\lim_{(x,y)\to(x_0,y_0)} \frac{f(x,y) - f(x_0,y_0) - f_x(x_0,y_0)(x-x_0) - f_y(x_0,y_0)(y-y_0)}{\sqrt{(x-x_0)^2 + (y-y_0)^2}} = 0. \qquad (6.9)$$

When we learned about limits of functions of two variables, we emphasized that for the limit to exist, you must get the same result (in this case 0) when you approach (x_0, y_0) along any path whatsoever. It is precisely this fact that leads from Equation (6.9) to the condition that the plane $z = L(x,y)$ contains the tangent line to *every* path in the graph $z = f(x,y)$ through $(x_0, y_0, f(x_0, y_0))$, i.e., that this plane really is tangent to the graph.

Example 6.10

Let $f(x,y) = x^{1/3}y^{2/3}$ and let $(x_0,y_0) = (0,0)$. We have $f(x_0,y_0) = 0$. We saw in Example 4.6 that $f_x(0,0) = 0$ and one shows similarly that $f_y(0,0) = 0$. Thus in Equation (6.9), many of the terms in the numerator vanish and Equation (6.9) becomes

$$\lim_{(x,y)\to(0,0)} \frac{x^{1/3}y^{2/3}}{\sqrt{x^2+y^2}}.$$

The theorem tells us that f is differentiable if and only if this limit exists and equals zero. We leave it as an exercise to check whether this is the case.

Exercises

Exercise 6.1.1 *Find $[f'(x_0,y_0)]$ for each of the following:*

(a) $f(x,y) = x^3\ln(y),\;\; (x_0,y_0) = (2,1)$

(b) $f(x,y) = e^{2x-y},\;\; (x_0,y_0) = (1,2)$

(c) $f(x,y) = \frac{xy}{x^2+y^2},\;\; (x_0,y_0) = (1,1).$

Exercise 6.1.2 *For each of the following, find the directional derivative in two ways: first using Theorem 6.4 and then by using the original definition of directional derivative. Check that your answers agree.*

(a) $f(x,y) = x^3\ln(y)$, $(x_0,y_0) = (2,1)$, $\mathbf{u} = \langle -\frac{4}{5}, \frac{3}{5}\rangle$.

(b) $f(x,y) = e^{2x-y}$, $(x_0,y_0) = (1,2)$, $\mathbf{u} = \langle \frac{4}{5}, -\frac{3}{5}\rangle$.

Exercise 6.1.3 *Let $f(x,y) = \frac{xy}{x^2+y^2}$ and $(x_0,y_0) = (1,1)$. Use Theorem 6.4 to compute the directional derivatives of f at the point $(1,1)$ in the following directions:*

(a) in the direction from $(1,1)$ towards $(3,4)$.

(b) in the direction of the vector $\langle 2,3\rangle$.

Exercise 6.1.4 *Suppose that f is differentiable at (x_0,y_0). Let θ be the angle between the unit vector $\mathbf{u}$ and the vector $\mathbf{v} = \langle f_x(x_0,y_0), f_y(x_0,y_0)\rangle$. (Note that $\mathbf{v}$ is the derivative of f viewed as a vector rather than as a row matrix.) Show that $D_{\mathbf{u}}f(x_0,y_0) = |\mathbf{v}|\cos(\theta)$.*

Exercise 6.1.5 *Suppose that the radius r of a cylinder is increasing at 0.5 in/sec and the height is decreasing at 1 in/sec. At the instant when the radius is 5 inches and the height is 10 inches, at what rate is the volume changing?*

Exercise 6.1.6 *Assume that the temperature at any point (x,y) on a metal plate is given by $f(x,y) = 2x^2+5xy-2y^2$ in degrees Fahrenheit. Suppose that you run a temperature probe over the plate. If the location of the probe at time t is given by $(x(t),y(t)) = (2t^2,t)$, how fast is the temperature reading on the probe changing at time $t_0 = 1$? (Assume that the probe registers temperature instantly.)*

Exercise 6.1.7 *In Example 4.6, we saw that the function $f(x,y) = x^{1/3}y^{2/3}$ satisfies $f_x(0,0) = 0$. A similar computation shows that $f_y(0,0) = 0$. In Exercise 4.1.8, you computed $D_{\mathbf{u}}f(0,0)$ for $\mathbf{u} = \langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\rangle$ and got $\frac{1}{\sqrt{2}}$. Why doesn't this contradict Theorem 6.4? What can you conclude about f? Is f differentiable at $(0,0)$?*

Exercise 6.1.8 *For the function $f(x,y) = x^{1/3}y^{2/3}$, show that the limit in Equation 6.9 doesn't exist. (See Example 6.10.)*

Note: if you use the path $y = x$ for one of your test paths, you may want to compare your limit with the directional derivative that you obtained in Exercise 4.1.8(c).

6.2 Derivatives in Higher Dimensions

Outcomes

- *Understand and be able to compute derivatives of functions whose domains and target spaces are of any dimension 1, 2, or 3.*

6.2.1. Real-valued functions of three variables

For functions $f : D \to \mathbb{R}$ where D is contained in $\mathbb{R}^3$, we can no longer visualize the graph $w = f(x,y,z)$ since it lies in four-dimensional space. However, the concepts of derivative, tangent approximations, and directional derivatives are all completely analogous to that of real-valued functions of two variables.

6.11: Derivative and Tangent Approximation

- *A sufficient condition for f to be differentiable at a point (x_0,y_0,z_0) is that the three partials f_x, f_y and f_z all exist and are continuous at (x_0,y_0,z_0).*
- *If f is differentiable at (x_0,y_0,z_0), we define its derivative by*

$$[f'(x_0,y_0,z_0)] = \begin{bmatrix} f_x(x_0,y_0,z_0) & f_y(x_0,y_0,z_0) & f_z(x_0,y_0,z_0) \end{bmatrix}$$

  The derivative is also denoted $[Df_{(x_0,y_0,z_0)}]$.

- *The tangent approximation of f is given by*

$$w - w_0 = [f'(x_0,y_0,z_0)] \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix}$$

  where $w_0 = f(x_0,y_0,z_0)$, equivalently,

$$\boxed{w - w_0 = f_x(x_0,y_0,z_0)(x - x_0) + f_y(x_0,y_0,z_0)(y - y_0) + f_z(x_0,y_0,z_0)(z - z_0).}$$

  We can also write $w = L(x,y,z)$ where

$$\boxed{L(x,y,z) = f(x_0,y_0,z_0) + f_x(x_0,y_0,z_0)(x - x_0) + f_y(x_0,y_0,z_0)(y - y_0) + f_z(x_0,y_0,z_0)(z - z_0).}$$

For $\mathbf{u} = \langle u_1, u_2, u_3 \rangle$ any unit vector in $\mathbb{R}^3$, the directional derivative $D_{\mathbf{u}} f(x_0,y_0,z_0)$ is defined by

$$D_{\mathbf{u}} f(x_0,y_0,z_0) = \frac{d}{dt}_{|t=0} f(x_0 + tu_2,\, y_0 + tu_2,\, z_0 + tu_3).$$

As in the two-dimensional case (see Theorems 6.4 and 6.7), we have:

Theorem 6.12

Assume that f is differentiable at (x_0, y_0, z_0). Then:

1. *For any unit vector $\mathbf{u} = \langle u_1, u_2, u_3 \rangle$, we have*

$$D_{\mathbf{u}} f(x_0, y_0, z_0) = [f'(x_0, y_0, z_0)] \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

2. *More generally, suppose $x(t), y(t), z(t)$ are differentiable at t_0 and $(x(t_0), y(t_0), z(t_0)) = (x_0, y_0, z_0)$. Letting $w(t) = f(x(t), y(t), z(t))$, we have*

$$w'(t_0) = [f'(x_0, y_0, z_0)] \begin{bmatrix} x'(t_0) \\ y'(t_0) \\ z'(t_0) \end{bmatrix}$$

Example 6.13

A rectangular metal box is expanding due to a rise in temperature. At a certain instant t_0, the length, width and height of the box are 5 feet, 2 feet and 3 feet and they are increasing at the rates of 0.2, 0.1 *and* 0.1 *feet per second, respectively. At what rate is the volume increasing at that instant?*

Solution. Let x, y and z denote the length, width and height, respectively. The volume is then given by $f(x, y, z) = xyz$. Since x, y and z are changing over time, they are functions $x = x(t)$, $y = y(t)$, and $z = z(t)$ of time t. Write

$$V(t) = f(x(t), y(t), z(t)),$$

which is the volume of the box at time t. By Theorem 6.12, part (2), we have

$$V'(t_0) = [f'(5, 2, 3)] \begin{bmatrix} x'(t_0) \\ y'(t_0) \\ z'(t_0) \end{bmatrix}.$$

Now $f_x(x, y, z) = yz$, so $f_x(5, 2, 3) = 6$. Similarly $f_y(5, 2, 3) = 15$ and $f_z(5, 2, 3) = 10$. At that instant, we are given that $x'(t_0) = 0.2$ and $y'(t_0) = 0.1 = z'(t_0)$. Thus

$$V'(t_0) = \begin{bmatrix} 6 & 15 & 10 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.1 \\ 0.1 \end{bmatrix} = 3.7 \text{ cubic feet/sec}$$

♠

6.2.2. Higher-dimension target space

We now consider vector-valued functions of one or more variables. We will begin with an example in order to motivate the definition of the derivative.

Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the function

$$F(x,y) = \begin{bmatrix} x^2 - y^2 \\ 2xy \end{bmatrix}$$

and let $(x_0, y_0) = (1,1)$. Writing $u = F_1(x,y) = x^2 - y^2$ and $v = F_2(x,y) = 2xy$, we have

$$\begin{bmatrix} u \\ v \end{bmatrix} = F(x,y) = \begin{bmatrix} F_1(x,y) \\ F_2(x,y) \end{bmatrix} = \begin{bmatrix} x^2 - y^2 \\ 2xy \end{bmatrix}$$

In particular, $u = x^2 - y^2$ and $v = 2xy$.

We leave it to the reader to compute the partial derivatives of F_1 and F_2 at (x_0, y_0) to verify that the tangent planes to the graphs $u = x^2 - y^2$ and $v = 2xy$ are given by:

$$u - 0 = 2(x-1) - 2(y-1) = \begin{bmatrix} 2 & -2 \end{bmatrix} \begin{bmatrix} x-1 \\ y-1 \end{bmatrix}$$

and

$$v - 2 = 2(x-1) + 2(y-1) = \begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} x-1 \\ y-1 \end{bmatrix}$$

Putting these two tangent approximations together, we obtain:

$$\begin{bmatrix} u-0 \\ v-2 \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x-1 \\ y-1 \end{bmatrix}$$

This is our tangent approximation to F at $(1,1)$. The derivative $[F'(1,1)]$ (as we will define in Definition 6.14) is the matrix

$$[F'(1,1)] = \begin{bmatrix} 2 & -2 \\ 2 & 2 \end{bmatrix} \tag{6.10}$$

Geometric interpretation. We are working in too many dimensions to talk about tangent lines or tangent planes here. So what is the tangent approximation above telling us? To try to get an idea of the behavior of the function F, we first consider the linear transformation whose matrix is $[F'(1,1)]$. This is actually a somewhat familiar linear transformation. To see this pull out a factor of $2\sqrt{2}$ from the matrix in Equation (6.10):

$$[F'(1,1)] = 2\sqrt{2} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = 2\sqrt{2}\,[\mathrm{Rot}_{\frac{\pi}{4}}].$$

Thus the derivative is the matrix of the linear transformation T given by rotation about the origin through the angle $\frac{\pi}{4}$ followed by expansion by a factor $2\sqrt{2}$. We illustrate this linear transformation in Figure 6.1. The small disk on the left is in the domain (the xy-plane). The larger disk on the right is the image of this disk in the target space (the uv-plane). The purple segment illustrates that T not only enlarges the disk but also rotates it.

Figure 6.2 illustrates the tangent approximation of F at $(1,1)$. The picture is the same as before except it takes disks centered at the point $(1,1)$ in the xy-plane to the larger rotated disk now centered at $F(1,1) = (0,2)$.

This tangent approximation gives an idea of the behavior of the function F near the point $(1,1)$. A very small disk centered at (1,1) is carried by F *approximately* to a disk centered about the point $(0,2)$ whose radius is $2\sqrt{2}$ times as large and which is rotated through an angle of $\frac{\pi}{4}$. The area expands approximately by the square of $2\sqrt{2}$ or 8, which is the determinant of the derivative matrix.

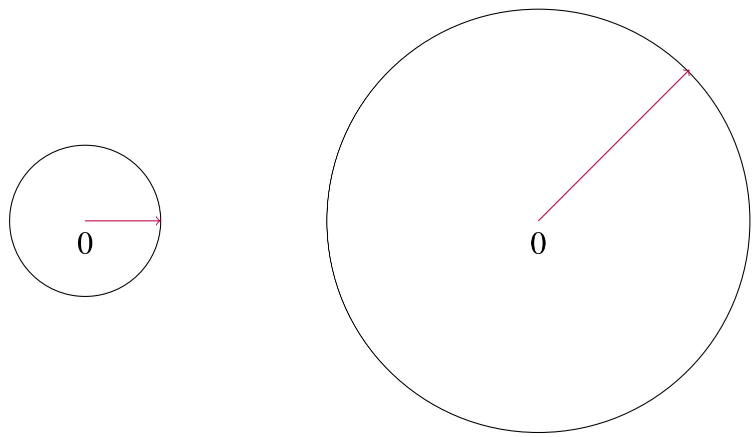


Figure 6.1

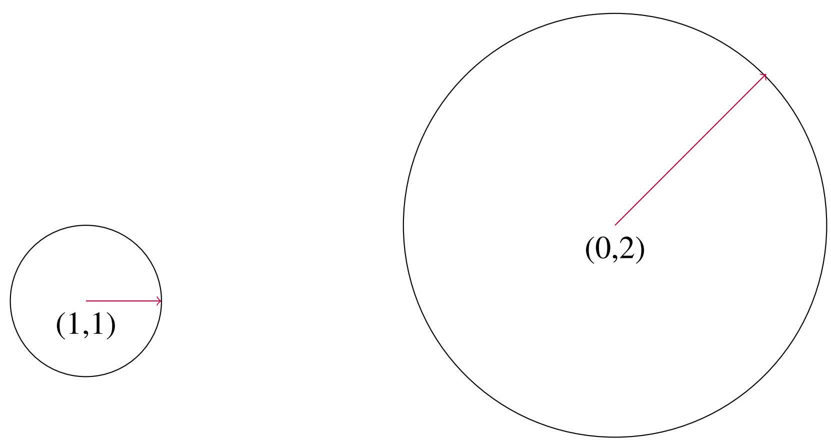


Figure 6.2

As the example illustrates, the derivative of a differentiable function whose target space may have higher dimension is given as follows:

Definition 6.14

Let F be a vector-valued function of one or more variables. We say that F is differentiable if each of its component functions is differentiable. The derivative of F at a point p, denoted $[F'(p)]$ or $[DF_p]$ is then the matrix whose rows are the derivatives of the component functions of F at p.

Example 6.15: Functions of one variable

Define $F : \mathbb{R} \to \mathbb{R}^3$ *by*

$$F(t) = \begin{bmatrix} e^t \\ 1+2t \\ \sin(t) \end{bmatrix}.$$

Compute $[F'(0)]$.

Solution.

The component functions of F are $F_1(t) = e^t$, $F_2(t) = 1+2t$ and $F_3(t) = \sin(t)$. Computing their derivatives, we get $F_1'(0) = 1$, $F_2'(0) = 2$ and $F_3'(0) = 1$. Thus

$$[F'(0)] = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

♠

6.16: Compare

Compare this computation with the familiar way of computing tangent vectors to vector-valued functions of one variable. Let $\mathbf{r}(t) = \langle e^t, 1+2t, \sin(t)\rangle$; *this is the same as the function* F *but is written in vector form. Recall that* $\mathbf{r}(t)$ *defines a curve in* $\mathbb{R}^3$, *and the tangent vector to the curve at a point is* $\mathbf{r}'(t) = \langle e^t, 2, \cos(t)\rangle$. *In particular, the tangent vector at* $(1,1,0)$ *is* $\mathbf{r}'(0) = \langle 1,2,1\rangle$. *This tangent vector is precisely* $[F'(0)]$ *written in vector rather than column form.*

Example 6.17

Define a function $F : \mathbb{R}^2 \to \mathbb{R}^3$ *by*

$$F(x,y) = \begin{bmatrix} x^2 y \\ 3xy \\ 5x+4y \end{bmatrix}.$$

Let $(x_0, y_0) = (1,1)$. *Compute* $[F'(1,1)]$

Solution. The component functions are $F_1(x,y) = x^2 y$, $F_2(x,y) = 3xy$, and $F_3(x,y) = 5x+4y$. We have $[F_1'(x,y)] = \begin{bmatrix} 2xy & x^2 \end{bmatrix}$ so

$$[F_1'(1,1)] = \begin{bmatrix} 2 & 1 \end{bmatrix}.$$

Similarly

$$[F_2'(1,1)] = \begin{bmatrix} 3 & 3 \end{bmatrix}$$

and

$$[F_3'(1,1)] = \begin{bmatrix} 5 & 4 \end{bmatrix}$$

Thus

$$[F'(1,1)] = \begin{bmatrix} 2 & 1 \\ 3 & 3 \\ 5 & 4 \end{bmatrix}.$$

♠

6.2.3. Section summary

- For real-valued functions of 3 variables, derivatives, directional derivatives and related rates problems are all analogous to those for real-valued functions of 2 variables.
- A vector-valued function is differentiable if and only if each of its component functions is differentiable. The derivative is the matrix whose rows are the derivatives of the component functions.
- Vector-valued functions of one-variable can be viewed as curves. The derivative, as defined in this section, is the tangent vector to the curve expressed in column form.

Exercises

Exercise 6.2.1 *Let* $f(x,y,z) = \frac{xy}{z}$.

(a) Find $[f'(2,1,1)]$.

(b) Write down the tangent approximation $L(x,y,z)$ *of* f *at* $(2,1,1)$ *and use it to approximate* $f(1.9,0.8,1.1)$.

(c) Evaluate $D_{\mathbf{u}} f(2,1,1)$ *in the direction from* $(2,1,1)$ *towards the origin.*

Exercise 6.2.2 *Let* $f(x,y,z) = e^{2x-yz}$.

(a) Find $[f'(1,2,1)]$.

(b) Write down the tangent approximation $L(x,y,z)$ *of* f *at* $(1,2,1)$ *and use it to approximate* $f(0.9,1.8,1.1)$.

(c) Evaluate the directional derivative of f *at* $(1,2,1)$ *in the direction of the vector* $\langle 1,2,3\rangle$.

Exercise 6.2.3 *Let* $f(x,y,z) = y\tan(x) + y^2 z$.

(a) Find $[f'(0,1,5)]$.

(b) Write down the tangent approximation $L(x,y,z)$ *of* f *at* $(0,1,5)$ *and use it to approximate* $f(0.1,1.1,4.8)$.

(c) Evaluate the directional derivative of f *at* $(0,1,5)$ *in the direction from* $(0,1,5)$ *towards* $(2,3,6)$.

Exercise 6.2.4 *Find the derivatives of each of the following functions at the indicated point:*

(a) $F(x,y,z) = \begin{bmatrix} e^{3x-y-z} \\ x^2y^2 \end{bmatrix}$ *at* $(x_0,y_0,z_0) = (1,1,2)$

(b) $F(x,y) = \begin{bmatrix} (x+y)^3 \\ x \\ \sin(y) \end{bmatrix}$ *at* $(x_0,y_0) = (2,0)$.

(c) $F(t) = \begin{bmatrix} t^2 \\ t^3 \\ t^4 \end{bmatrix}$ *at* $t_0 = 1$.

(d) $F(x,t,z) = \begin{bmatrix} xy\cos(\pi z) \\ e^{2x-yz} \\ xyz \end{bmatrix}$ *at* $(x_0,y_0,z_0) = (1,2,1)$.

Exercise 6.2.5 *Suppose* $F : \mathbb{R}^3 \to \mathbb{R}^3$ *is differentiable at* (x_0,y_0,z_0) *and that its derivative at* (x_0,y_0,z_0) *is*

$$[F'(x_0,y_0,z_0)] = \begin{bmatrix} 2 & 5 & 1 \\ 3 & 6 & 4 \\ 7 & 8 & 9 \end{bmatrix}$$

Read off the following partials from the derivative matrix:

(a) $\frac{\partial F_2}{\partial x}(x_0,y_0,z_0)$

(b) $\frac{\partial F_3}{\partial y}(x_0,y_0,z_0)$

(c) $\frac{\partial F_1}{\partial z}(x_0,y_0,z_0)$

Exercise 6.2.6 *Let* T *be the linear transformation given by*

$$T = \begin{bmatrix} 2x+y+4z \\ x+7y+8z \end{bmatrix}$$

Show that for every point (x_0,y_0,z_0) *in* $\mathbb{R}^3$ *we have*

$$[T'(x_0,y_0,z_0)] = [T].$$

Once you do the computation, it should be clear that the same relationship between T *and its derivative holds for every linear transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$.

6.3 Tangents to level sets

Motivation

This short section is an accompaniment to Section 14.6 in Stewart. It gives an alternative way of understanding equation 19 on page 994 of Stewart.

Figure 6.3 shows part of the graph of the function $f(x,y) = 10 - x^2 - y^2$. The grey plane below it is the xy-plane. The brown plane is the tangent plane to the graph $z = f(x,y)$ at a point $(x_0, y_0, f(x_0,y_0))$, where $z_0 = f(x_0,y_0)$. We can express the tangent plane as $z = L(x,y)$, where $L(x,y)$ is the tangent approximation of f near (x_0,y_0). The light blue plane is the plane $z = z_0$. The plane $z = z_0$ intersects the graph in a circle, whose projection to the xy-plane is the level curve $f(x,y) = z_0$. The plane $z = z_0$ and the tangent plane intersect in a line, whose projection to the xy-plane is the level curve $L(x,y) = z_0$. (Theses level curves to f and to L are the circle and the line shown in the gray plane.) As the picture illustrates, the level set $L(x,y) = z_0$ is tangent to the level set $f(x,y) = z_0$ at (x_0,y_0).

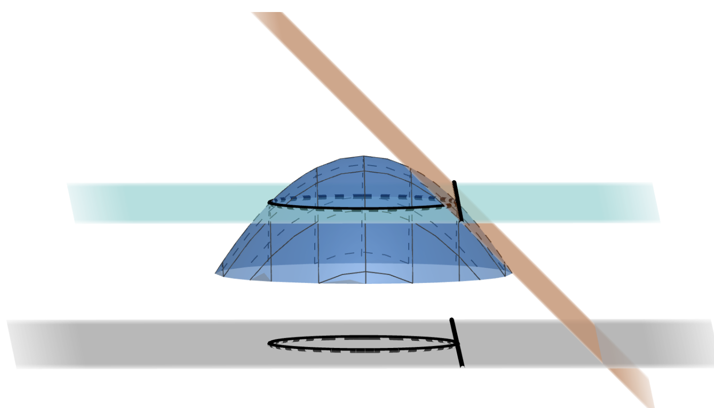


Figure 6.3

Recall that

$$L(x,y) = z_0 + f_x(x_0,y_0)(x - x_0) + f_y(x_0,y_0)(y - y_0).$$

Thus its level set $L(x,y) = z_0$ is given by $f_x(x_0,y_0)(x - x_0) + f_y(x_0,y_0)(y - y_0) = 0$. Generalizing this observation:

Theorem 6.18

1. *Let $f(x,y)$ be a real-valued function. Assume that f is differentiable at the point (x_0,y_0). Let $L(x,y)$ be the tangent approximation to f at (x_0,y_0). Let $z_0 = f(x_0,y_0)$. Then the tangent line to the level set $f(x,y) = z_0$ at (x_0,y_0) is the level set $L(x,y) = z_0$ and is given by*

$$f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0) = 0. \tag{6.11}$$

2. *Similarly, let g be a real-valued function of three variables. Assume that g is differentiable at the point (x_0,y_0,z_0). Let $L(x,y,z)$ be the tangent approximation to g at (x_0,y_0,z_0) and let $w_0 = f(x_0,y_0,z_0)$. Then the tangent plane to the level surface $f(x,y,z) = w_0$ at (x_0,y_0,z_0) is the level surface $L(x,y,z) = w_0$, given by*

$$f_x(x_0,y_0,z_0)(x-x_0) + f_y(x_0,y_0,z_0)(y-y_0) + f_z(x_0,y_0,z_0)(z-z_0) = 0. \tag{6.12}$$

As the following example illustrates, often one can use the theorem to find the tangent plane to a surface S by first finding a function whose level surface is S.

Example 6.19

Let S be the surface $xyz = x^2 + y + 8$. Find the tangent plane to this surface at the point $(2,4,2)$.

Solution. Writing the equation for the surface as $xyz - x^2 - y = 8$, we see that it is a level set of the function $f(x,y,z) = xyz - x^2 - y$. Using Equation (6.12) with $(x_0,y_0,z_0) = (2,4,2)$ we obtain (after computing that $f_x(2,4,2) = 4$, etc.) that

$$4(x-2) + 3(y-4) + 8(z-2) = 0.$$

♠

6.4 The Chain Rule

Motivation

The chain rule is one of the most powerful formulas you learned in single-variable calculus. We introduce the analogous chain rule for multi-variable functions.

Outcomes

- *Be able to compute the derivative of a composition of functions $F \circ G$.*
- *From the derivative, be able to read off the various partial derivatives of the components of $F \circ G$. Understand the relationship with the chain rule stated in Stewart.*

Recall the chain rule for real-valued functions of one variable:

6.20: Familiar chain rule

If $g : \mathbb{R} \to \mathbb{R}$ is differentiable at x_0 and $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $y_0 = g(x_0)$, then $f \circ g$ is differentiable at x_0 and

$$(f \circ g)'(x_0) = f'(y_0)g'(x_0).$$

Using matrices, the chain rule in higher dimensions looks identical to the familiar chain rule. In the theorem below, $\mathbb{R}^n$, $\mathbb{R}^m$ and $\mathbb{R}^p$ can each be any of $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$. (The statement is also valid in higher dimensions but as usual, we will just work in dimensions 1, 2, and 3.)

Theorem 6.21: Chain Rule

If $G : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0$ and $F : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $\mathbf{y}_0 = G(\mathbf{x}_0)$, then $F \circ G$ is differentiable at $\mathbf{x}_0$ and

$$[(F \circ G)'(\mathbf{x}_0)] = [F'(\mathbf{y}_0)][G'(\mathbf{x}_0)].$$

(Aside: If we view the derivative of a function F as a linear transformation, the chain rule says the derivative of the composition $F \circ G$ is the composition of their derivatives.)

Intuition. Writing $[\mathbf{z}_0] = F(\mathbf{y}_0)$, the tangent approximations say that

$$[\mathbf{y} - \mathbf{y}_0] \sim [G'(\mathbf{x}_0)][\mathbf{x} - \mathbf{x}_0]$$

and

$$[\mathbf{z} - \mathbf{z}_0] \sim [F'(\mathbf{y}_0)][\mathbf{y} - \mathbf{y}_0].$$

Substituting the approximation for $\mathbf{y} - \mathbf{y}_0$ in the first equation into the second one above, we get

$$[\mathbf{z} - \mathbf{z}_0] \sim [F'(\mathbf{y}_0)][G'(\mathbf{x}_0)][\mathbf{x} - \mathbf{x}_0]. \tag{6.13}$$

On the other hand, the tangent approximation for $F \circ G$ near $\mathbf{x}_0$ is given by

$$[\mathbf{z} - \mathbf{z}_0] \sim [(F \circ G)'(\mathbf{x}_0)][\mathbf{x} - \mathbf{x}_0]. \tag{6.14}$$

A comparison of the approximations (6.13) and (6.14) suggests the chain rule

$$[(F \circ G)'(\mathbf{x}_0)] = [F'(\mathbf{y}_0)][G'(\mathbf{x}_0)]$$

(The actual proof of the chain rule, which we omit, involves limits.)

Example 6.22

Let $f(x,y) = x^2y^3$ and $\mathbf{r}(t) = \langle e^t, 1+t \rangle$, or in column form (which is recommended before taking the derivative), $\mathbf{r}(t) = \begin{bmatrix} e^t \\ 1+t \end{bmatrix}$. Let's compute $[(f \circ \mathbf{r})'(0)]$. We have $\mathbf{r}(0) = (1,2)$, so the chain rule tells us that

$$[(f \circ \mathbf{r})'(0)] = [f'(1,2)][\mathbf{r}'(0)] = \begin{bmatrix} 16 & 12 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = [28].$$

Since the composition $f \circ \mathbf{r}$ is a real-valued function of just one variable, you might want to leave out the brackets around the 28.

The composition in Example 6.22 can be written as $f \circ \mathbf{r}(t) = f(x(t)), y(t))$, where $x(t) = e^t$ and $y(t) = 1 + t$. You might recall that we've computed $\frac{d}{dt} f(x(t)), y(t))$ before without using the chain rule:

6.23: Compare

Compare the case of the chain rule illustrated in Example 6.22 with Theorem 6.7. Both results say the same thing: if $x(t)$ and $y(t)$ are differentiable at t_0 and $(x(t_0), y(t_0)) = (x_0, y_0)$, then

$$\frac{d}{dt}_{|t=t_0} f(x(t), y(t)) = [f'(x_0, y_0)] \begin{bmatrix} x'(t_0) \\ y'(t_0) \end{bmatrix}$$

In Theorem 6.7, we obtained this result only using the fact that the tangent plane to the graph of f contains the tangent line to all smooth curves in the graph $z = f(x, y)$.

Example 6.24

Define $G : \mathbb{R}^2 \to \mathbb{R}^3$ and $f : \mathbb{R}^3 \to \mathbb{R}$ by

$$G(s,t) = \begin{bmatrix} s^2 t \\ s + 2t^2 \\ st \end{bmatrix} \quad \text{and} \quad f(x,y,z) = e^{2x-y+z}.$$

Compute the derivative of $f \circ G$ at $(s_0, t_0) = (1, 1)$.

Solution. Note that $g(1,1) = (1,3,1)$. The chain rule says that

$$[(f \circ G)'(1,1)] = [f'(1,3,1)][G'(1,1)].$$

We compute:

$$[G'(s,t)] = \begin{bmatrix} 2st & s^2 \\ 1 & 4t \\ t & s \end{bmatrix} \quad \text{so} \quad [G'(1,1)] = \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 1 & 1 \end{bmatrix}$$

Similarly check that

$$[f'(1,3,1)] = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix}$$

Thus

$$[(f \circ G)'(1,1)] = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \end{bmatrix}$$

♠

6.25: Compare

We use Example 6.24 to compare Theorem 6.21 with the version of the chain rule in Stewart. Writing $w = f \circ G(s,t)$, *we can read off from the derivative matrix computed in the example that at* $(s_0,t_0) = (1,1)$,

$$\frac{\partial w}{\partial s} = 4 \text{ and } \frac{\partial w}{\partial t} = -1.$$

Writing $w = f(x,y,z)$ *and* $(x,y,z) = G(s,t)$, *the chain rule in Stewart says that*

$$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial s}.$$

At $(s,t) = (1,1)$ *and* $(x,y,z) = (1,3,1)$, *we get*

$$\frac{\partial w}{\partial s} = 2(2) + (-1)(1) + (1)(1) = 4. \tag{6.15}$$

This agrees with what we obtained using the matrices. Notice that the computation in Equation 6.15 is equivalent to multiplying the first (and only) row of the matrix $[f'(1,3,1)]$ *by the first column of the matrix* $[G'(1,1)]$. *Similarly, the formula in Stewart for* $\frac{\partial w}{\partial t}$ *corresponds to multiplying the row of* $f'(1,3,1)$ *by the second column of* $[G'(1,1)]$. *The matrix multiplication gives us both partials at once. (If you are only interested in one of the partials, then the formula in Stewart is a bit faster. If you want all the partials, the matrix method is more convenient.)*

Example 6.26

Let G be the function in Example 6.24 and define $F : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$F(x,y,z) = \begin{bmatrix} e^{2x-y+z} \\ xyz \end{bmatrix}$$

We can compute (check!)

$$[F'(1,3,1)] = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 1 & 3 \end{bmatrix}$$

so

$$[(F \circ G)'(1,1)] = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 1 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 10 & 10 \end{bmatrix} \tag{6.16}$$

To compare with the chain rule in Stewart, let $u = e^{2x-y+z}$ and $v = xyz$ (so u and v are the two components functions of F, and write $x = s^2t,\ y = s + 2t^2,\ z = st$ (these are the three component functions of G.) We then have

$$\begin{bmatrix} u \\ v \end{bmatrix} = F(x,y,z) \ \text{ and } \begin{bmatrix} x \\ y \\ z \end{bmatrix} = G(s,t)$$

. The composition allows us to view u and v as functions of s and t. From Equation (6.16), we can read off the partials of u and v with respect to s and t at $(s_0,t_0) = (1,1)$. The rows of the matrix correspond to u and v, the two component functions and the columns correspond to the two variables s and t. We thus read off

$$\frac{\partial u}{\partial s} = 4,\ \frac{\partial u}{\partial t} = -1,\ \frac{\partial v}{\partial s} = 10,\ \frac{\partial v}{\partial t} = 10.$$

6.4.1. Section summary

- $[(F \circ G)'(\mathbf{x}_0)] = [F'(\mathbf{y}_0)][G'(\mathbf{x}_0)]$ where $\mathbf{y}_0 = G(\mathbf{x}_0)$. Except for the fact that these are matrices rather than real numbers, this is identical to the single-variable chain rule.
- If you just want to compute a single partial derivative, it is easier to use the version of the chain rule in Stewart. If you want all the partials, it's faster to use the matrix version and then read off the various partials from the matrix entries.

Exercises

Exercise 6.4.1 *For each of the following, compute the derivative $[(F\circ G)']$ at the indicated point.*

(a) $F(x,y) = \begin{bmatrix} e^x y^2 \\ (x+1)y^3 \end{bmatrix}$, $G(t) = \begin{bmatrix} \sin(t) \\ e^t \end{bmatrix}$, $t_0 = 0$.

(b) $F(x,y,z) = \begin{bmatrix} xy/z \\ y\tan(z-1) \\ x^2yz \end{bmatrix}$, $G(s,t) = \begin{bmatrix} s^2t \\ 2se^{t-2} \\ t-2s+1 \end{bmatrix}$, $(s_0,t_0) = (1,2)$.

(c) $F(x,y) = e^x y^2$, $G(x,y,z) = \begin{bmatrix} 3x+2y+5z \\ 2\sqrt{x+y+z} \end{bmatrix}$, $(x_0,y_0,z_0) = (1,1,-1)$.

(Note: it doesn't matter that we used the same names for some of the variables. You can always rename the variables for F if you want to.)

Exercise 6.4.2 *This exercise refers to problem 22 in Stewart Section 14.5.*

(a) *Express the given functions as $T = F(u,v)$ and $\left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right] = G(p,q,r)$.*

(b) *Write down the derivative matrices for F and G at the indicated point. (Here the indicated point is $(p,q,r) = (2,1,4)$ and $(u,v) = G(2,1,4)$. You first need to compute $G(2,1,4)$.)*

(c) *Use the matrix version of the Chain Rule to compute the derivative $[(F\circ G)'(2,1,4)]$.*

(d) *Read off from your matrix in part (c) the partial derivatives that are requested in problem 22.*

(e) *Find $\frac{\partial T}{\partial p}$ using a tree diagram as in Stewart and check that your answer agrees with the value obtained in part (d).*

Exercise 6.4.3 *This exercise refers to problem 23 in Stewart Section 14.5.*

(a) *Follow the procedure outlined in steps (a)-(d) of Exercise 6.4.2 to express w as the composition of two functions F and G, to find the derivative of this composition at the indicated point, and to read off the partial derivatives $\frac{\partial w}{\partial r}$ and $\frac{\partial w}{\partial \theta}$.*

(b) *Use a tree diagram as in Stewart to compute $\frac{\partial w}{\partial r}$ and compare your answer with that in part (a).*

Exercise 6.4.4 *This exercise refers to problem 24 in Stewart Section 14.5.*

(a) *Follow the procedure outlined in steps (a)-(d) of Exercise 6.4.2 to express P as the composition of two functions F and G, to find the derivative of this composition at the indicated point, and to read off the partial derivatives $\frac{\partial P}{\partial x}$ and $\frac{\partial P}{\partial y}$.*

(b) *Use a tree diagram as in Stewart to compute $\frac{\partial P}{\partial y}$ and compare your answer with that in part (a).*

Exercise 6.4.5 *This and the next exercise illustrate that the product rule that you learned for differentiating functions such as $f(x) = x^2 \sin(x)$ is actually a special case of the chain rule. Express $z = x^2 \sin(x)$ as*

$$z = uv, \text{ where } u = x^2 \text{ and } v = \sin(x). \tag{6.17}$$

. Now apply the chain rule (either version you wish) to Equation (6.17) to compute $\frac{dz}{dx}$.

Exercise 6.4.6 *This exercise is a continuation of Exercise 6.4.5. Now let $u(x)$ and $v(x)$ be any differentiable real-valued functions. (Exercise 6.4.5 was the special case $u(x) = x^2$ and $v(x) = \sin(x)$.) Write $z = uv$ and use the chain rule to show that*

$$\frac{dz}{dx} = u'(x)v(x) + u(x)v'(x).$$

Thus the product rule follows from the chain rule.

7. Answers to Selected Exercises

Chapter 1. Systems of Equations

Systems Of Equations, Algebraic Procedures

1.2.1(a) row-echelon form , not reduced row-echelon form 1.2.2(a) inconsistent, (b) $x=5-2t,\ y=t,\ z=4$ where t is any real number 1.2.13 (f) $x=-1,\ y=2,\ z=-1$; (g) $x=1-2t,\ y=t,\ z=1$ where t is any real number

Chapter 2. Vectors: a Linear Viewpoint

Linear Combinations and Spans

2.1.3 (a) $\mathbf{u}=2\mathbf{v}-\mathbf{w}$. (d) $\mathbf{u}=3\mathbf{v}+\mathbf{w}$. (e) not a linear combination
2.1.4 (a) no. (b) yes.
2.1.6 (a) $y=\frac{6}{5}x$.

Using a basis to provide a map of a subspace

2.2.1 (a) linearly dependent. (d) linearly dependent.
2.2.2 (a) $\{\langle 3,1\rangle\}$. (d) $\{\langle 3,1,1\rangle,\langle 1,2,1\rangle\}$.
2.2.3 (a) $(\frac{1}{2},-\frac{1}{2})$.

Vector Equations of Lines and Planes

2.3.1 (a) $\langle 1,5,6\rangle+t\langle 4,7,8\rangle$. (c) $\langle 1,4,5\rangle+t\langle 1,-3,1\rangle$.
2.3.2 (a) $x=1+4t, y=5+7t, z=6+8t$. (c) $x=1+t, y=4-3t, z=5+t$.
2.3.3 (a) $y=\frac{8}{3}x$.
2.3.4 (a) $\langle 4,-1\rangle+t\langle 4,-3\rangle$.

Chapter 3. Introduction to Matrices and Matrix Operations

Matrix Operations

3.2.7.

a. $\begin{bmatrix} -3 & -6 & -9 \\ -6 & -3 & -21 \end{bmatrix}$

b. $\begin{bmatrix} 8 & -5 & 3 \\ -11 & 5 & -4 \end{bmatrix}$

c. Not possible

d. $\begin{bmatrix} -3 & 3 & 4 \\ 6 & -1 & 7 \end{bmatrix}$

e. Not possible

f. Not possible

Chapter 4. Directional Derivatives and Differentiability

Directional Derivatives of Real-valued Functions

4.1.1 (a) $\frac{12}{5}$. (d) $\frac{5}{2\sqrt{2}}$.
4.1.2 (b) $15\sqrt{13}$.
4.1.3 (a) $\langle -\frac{4}{5}, \frac{3}{5}, \frac{12}{5} \rangle$

Chapter 5. Linear Transformations

Linear Transformations and Their Representing Matrices

5.2.1 (a) no. (b) yes
5.2.2 (b) $\begin{bmatrix} 1 & -1 \end{bmatrix}$
5.2.7 $\begin{bmatrix} 1 & 4 \\ 2 & 1 \\ 3 & 4 \end{bmatrix}$
5.2.13: (a) $\mathbb{R}^2$, rank 2. (b) line, rank 1. (c) $\mathbb{R}^2$, rank 2.

Example: Rotations and Reflections of $\mathbb{R}^2$

5.4.2 (b). $(-\sqrt{2}, 2\sqrt{2})$, $(-\frac{9\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$. (d). $(\frac{1-3\sqrt{3}}{2}, \frac{-3-\sqrt{3}}{2})$, $\frac{-5-4\sqrt{3}}{2}, \frac{-4+5\sqrt{3}}{2})$

Composition of LInear Transformations

5.6.1 (a) Yes. (b) No.
5.6.2 (a) Representing matrix $\begin{bmatrix} -5 & -1 \\ -4 & -2 \\ -15 & -5 \end{bmatrix}$

Chapter 6. Derivatives

Directional Derivatives of Real-valued Functions

6.1.1(a) $[0,8]$.

Derivatives in Higher Dimensions

6.2.1 (a) $[1,2,-2]$. (b) $L(x,y,z) = 2+1(x-2)+2(y-1)-2(z-1)$; 1.3. (c) $-\frac{\sqrt{6}}{3}$
6.2.4(a) $\begin{bmatrix} 3 & -1 & -1 \\ 2 & 2 & 0 \end{bmatrix}$

Chain Rule

6.4.1(a) $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$
6.4.2(a) $w = F(x,y,z) = xy+yz+xz$; $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = G(r,\theta) = \begin{bmatrix} r\cos(\theta) \\ r\sin(\theta) \\ r\theta \end{bmatrix}$;
$[F'(0,2,\pi)] = \begin{bmatrix} 2+\pi & \pi & 2 \end{bmatrix}$, $[G'(2,\frac{\pi}{2})] = \begin{bmatrix} 0 & -2 \\ 1 & 0 \\ \pi/2 & 2 \end{bmatrix}$
$[(F\circ G)'(0,2,\pi)] = \begin{bmatrix} 2\pi & -2\pi \end{bmatrix}$

$\frac{\partial w}{\partial r} = 2\pi$, $\frac{\partial w}{\partial \theta} = -2\pi$.