Linear Algebra with Applications

Linear Algebra with Applications

An adaptation for MAT223 UTM

XINLI WANG

Contents

Acknowledgement

1. System of Linear Equations

1.1 Solutions and elementary operations

Practical problems in many fields of study—such as biology, business, chemistry, computer science, economics, electronics, engineering, physics and the social sciences—can often be reduced to solving a system of linear equations. Linear algebra arose from attempts to find systematic methods for solving these systems, so it is natural to begin this book by studying linear equations.

If a, b, and c are real numbers, the graph of an equation of the form

$$ax + by = c$$

is a straight line (if a and b are not both zero), so such an equation is called a ***linear*** equation in the variables x and y. However, it is often convenient to write the variables as $x_1, x_2, \ldots, x_n$, particularly when more than two variables are involved. An equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

is called a **linear equation** in the n variables $x_1, x_2, \ldots, x_n$. Here $a_1, a_2, \ldots, a_n$ denote real numbers (called the **coefficients** of $x_1, x_2, \ldots, x_n$, respectively) and b is also a number (called the **constant term** of the equation). A finite collection of linear equations in the variables $x_1, x_2, \ldots, x_n$ is called a **system of linear equations** in these variables. Hence,

$$2x_1 - 3x_2 + 5x_3 = 7$$

is a linear equation; the coefficients of x_1, x_2, and x_3 are 2, -3, and 5, and the constant term is 7. Note that each variable in a linear equation occurs to the first power only.

Given a linear equation $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$, a sequence $s_1, s_2, \ldots, s_n$ of n numbers is called a **solution** to the equation if

$$a_1 s_1 + a_2 s_2 + \cdots + a_n s_n = b$$

that is, if the equation is satisfied when the substitutions $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ are made. A sequence of numbers is called **a solution to a system** of equations if it is a solution to every equation in the system.

A system may have no solution at all, or it may have a unique solution, or it may have an infinite family of solutions. For instance, the system $x + y = 2, x + y = 3$ has no solution because the sum of two numbers cannot be 2 and 3 simultaneously. A system that has no solution is called **inconsistent**; a system with at least one solution is called **consistent**.

Show that, for arbitrary values of s and t,

$$x_1 = t - s + 1$$
$$x_2 = t + s + 2$$
$$x_3 = s$$
$$x_4 = t$$

is a solution to the system

$$x_1 - 2x_2 + 3x_3 + x_4 = -3$$
$$2x_1 - x_2 + 3x_3 - x_4 = 0$$

Simply substitute these values of x_1, x_2, x_3, and x_4 in each equation.

$$x_1 - 2x_2 + 3x_3 + x_4 = (t - s + 1) - 2(t + s + 2) + 3s + t = -3$$
$$2x_1 - x_2 + 3x_3 - x_4 = 2(t - s + 1) - (t + s + 2) + 3s - t = 0$$

Because both equations are satisfied, it is a solution for all choices of s and t.

The quantities s and t in this example are called **parameters**, and the set of solutions, described in this way, is said to be given in **parametric form** and is called the **general solution** to the system. It turns out that the solutions to **every** system of equations (if there **are** solutions) can be given in parametric form (that is, the variables x_1, x_2, ... are given in terms of new independent variables s, t, etc.).

When only two variables are involved, the solutions to systems of linear equations can be described geometrically because the graph of a linear equation $ax + by = c$ is a straight line if a and b are not both zero. Moreover, a point $P(s, t)$ with coordinates s and t lies on the line if and only if $as + bt = c$–that is when $x = s, y = t$ is a solution to the equation. Hence the solutions to a **system** of linear equations correspond to the points $P(s, t)$ that lie on **all** the lines in question.

In particular, if the system consists of just one equation, there must be infinitely many solutions because there are infinitely many points on a line. If the system has two equations, there are three possibilities for the corresponding straight lines:

- The lines intersect at a single point. Then the system has **a unique solution** corresponding to that point.
- The lines are parallel (and distinct) and so do not intersect. Then the system has **no solution**.
- The lines are identical. Then the system has **infinitely many** solutions—one for each point on the (common) line.

With three variables, the graph of an equation $ax + by + cz = d$ can be shown to be a plane and so again provides a "picture" of the set of solutions. However, this graphical method has its limitations: When more than three variables are involved, no physical image of the graphs (called hyperplanes) is possible. It is necessary to turn to a more "algebraic" method of solution.

Before describing the method, we introduce a concept that simplifies the computations involved. Consider the following system

$$
\begin{aligned}
3x_1 + 2x_2 - x_3 + x_4 &= -1 \\
2x_1 \qquad - x_3 + 2x_4 &= 0 \\
3x_1 + x_2 + 2x_3 + 5x_4 &= 2
\end{aligned}
$$

of three equations in four variables. The array of numbers

$$
\left[\begin{array}{cccc|c}
3 & 2 & -1 & 1 & -1 \\
2 & 0 & -1 & 2 & 0 \\
3 & 1 & 2 & 5 & 2
\end{array}\right]
$$

occurring in the system is called the **augmented matrix** of the system. Each row of the matrix consists of the coefficients of the variables (in order) from the corresponding equation, together with the constant term. For clarity, the constants are separated by a vertical line. The augmented matrix is just a different way of describing the system of equations. The array of coefficients of the variables

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

is called the **coefficient matrix** of the system and

$\left[\begin{array}{c} -1 \\ 0 \\ 2 \end{array} \right]$ is called the **constant matrix** of the system.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=5#h5p-130

Elementary Operations

The algebraic method for solving systems of linear equations is described as follows. Two such systems are said to be **equivalent** if they have the same set of solutions. A system is solved by writing a series of systems, one after the other, each equivalent to the previous system. Each of these systems has the same set of solutions as the original one; the aim is to end up with a system that is easy to solve. Each system in the series is obtained from the preceding system by a simple manipulation chosen so that it does not change the set of solutions.

As an illustration, we solve the system $x + 2y = -2$, $2x + y = 7$ in this manner. At each stage, the corresponding augmented matrix is displayed. The original system is

$$\begin{aligned} x + 2y &= -2 \\ 2x + \ y &= \ \ 7 \end{aligned} \qquad \left[\begin{array}{cc|c} 1 & 2 & -2 \\ 2 & 1 & 7 \end{array} \right]$$

First, subtract twice the first equation from the second. The resulting system is

$$\begin{aligned} x + 2y &= -2 \\ -\ 3y &= \ 11 \end{aligned} \qquad \left[\begin{array}{cc|c} 1 & 2 & -2 \\ 0 & -3 & 11 \end{array} \right]$$

which is equivalent to the original. At this stage we obtain $y = -\frac{11}{3}$ by multiplying the second equation by $-\frac{1}{3}$. The result is the equivalent system

$$\begin{aligned} x + 2y &= \ -2 \\ y &= -\tfrac{11}{3} \end{aligned} \qquad \left[\begin{array}{cc|c} 1 & 2 & -2 \\ 0 & 1 & -\tfrac{11}{3} \end{array} \right]$$

Finally, we subtract twice the second equation from the first to get another equivalent system.

$$\begin{aligned} x &= \ \ \tfrac{16}{3} \\ y &= -\tfrac{11}{3} \end{aligned} \qquad \left[\begin{array}{cc|c} 1 & 0 & \tfrac{16}{3} \\ 0 & 1 & -\tfrac{11}{3} \end{array} \right]$$

Now **this** system is easy to solve! And because it is equivalent to the original system, it provides the solution to that system.

Observe that, at each stage, a certain operation is performed on the system (and thus on the augmented matrix) to produce an equivalent system.

Definition 1.1 **Elementary Operations**

The following operations, called **elementary operations**, can routinely be performed on systems of linear equations to produce equivalent systems.

1. Interchange two equations.
2. Multiply one equation by a nonzero number.
3. Add a multiple of one equation to a different equation.

Theorem 1.1.1

Suppose that a sequence of elementary operations is performed on a system of linear equations. Then the resulting system has the same set of solutions as the original, so the two systems are equivalent.

Elementary operations performed on a system of equations produce corresponding manipulations of the *rows* of the augmented matrix. Thus, multiplying a row of a matrix by a number k means multiplying *every entry* of the row by k. Adding one row to another row means adding *each entry* of that row to the corresponding entry of the other row. Subtracting two rows is done similarly. Note that we regard two rows as equal when corresponding entries are the same.

In hand calculations (and in computer programs) we manipulate the rows of the augmented matrix rather than the equations. For this reason we restate these elementary operations for matrices.

Definition 1.2 Elementary Row Operations

The following are called **elementary row operations** on a matrix.

1. Interchange two rows.
2. Multiply one row by a nonzero number.
3. Add a multiple of one row to a different row.

In the illustration above, a series of such operations led to a matrix of the form

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

where the asterisks represent arbitrary numbers. In the case of three equations in three variables, the goal is to produce a matrix of the form

$$\begin{bmatrix} 1 & 0 & 0 & | & * \\ 0 & 1 & 0 & | & * \\ 0 & 0 & 1 & | & * \end{bmatrix}$$

This does not always happen, as we will see in the next section. Here is an example in which it does happen.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=5#h5p-131

Example 1.1.3 Find all solutions to the following system of equations.

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

Solution:
The augmented matrix of the original system is

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

To create a 1 in the upper left corner we could multiply row 1 through by $\frac{1}{3}$. However, the 1 can be obtained without introducing fractions by subtracting row 2 from row 1. The result is

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

The upper left 1 is now used to "clean up" the first column, that is create zeros in the other positions in that column. First subtract 2 times row 1 from row 2 to obtain

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

Next subtract 4 times row 1 from row 3. The result is

$$\begin{bmatrix} 1 & 1 & 1 & \bigm| & 1 \\ 0 & 1 & -2 & \bigm| & -2 \\ 0 & -1 & -5 & \bigm| & -6 \end{bmatrix}$$

This completes the work on column 1. We now use the 1 in the second position of the second row to clean up the second column by subtracting row 2 from row 1 and then adding row 2 to row 3. For convenience, both row operations are done in one step. The result is

$$\begin{bmatrix} 1 & 0 & 3 & \big| & 3 \\ 0 & 1 & -2 & \big| & -2 \\ 0 & 0 & -7 & \big| & -8 \end{bmatrix}$$

Note that the last two manipulations **did not affect** the first column (the second row has a zero there), so our previous effort there has not been undermined. Finally we clean up the third column. Begin by multiplying row 3 by $-\frac{1}{7}$ to obtain

$$\begin{bmatrix} 1 & 0 & 3 & \big| & 3 \\ 0 & 1 & -2 & \big| & -2 \\ 0 & 0 & 1 & \big| & \frac{8}{7} \end{bmatrix}$$

Now subtract 3 times row 3 from row 1, and then add 2 times row 3 to row 2 to get

$$\begin{bmatrix} 1 & 0 & 0 & \big| & -\frac{3}{7} \\ 0 & 1 & 0 & \big| & \frac{2}{7} \\ 0 & 0 & 1 & \big| & \frac{8}{7} \end{bmatrix}$$

The corresponding equations are $x = -\frac{3}{7}$, $y = \frac{2}{7}$, and $z = \frac{8}{7}$, which give the (unique) solution.

1.2 Gaussian elimination

The algebraic method introduced in the preceding section can be summarized as follows: Given a system of linear equations, use a sequence of elementary row operations to carry the augmented matrix to a "nice" matrix (meaning that the corresponding equations are easy to solve). In Example 1.1.3, this nice matrix took the form

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

The following definitions identify the nice matrices that arise in this process.

---

*Definition 1.3 **row–echelon form** (reduced)*

A matrix is said to be in **row-echelon form** (and will be called a **row-echelon matrix** if it satisfies the following three conditions:

1. All **zero rows** (consisting entirely of zeros) are at the bottom.
2. The first nonzero entry from the left in each nonzero row is a 1, called the leading 1 for that row.
3. Each leading 1 is to the right of all **leading** 1s in the rows above it.

A row-echelon matrix is said to be in **reduced row-echelon form** (and will be called a **reduced row-echelon matrix** if, in addition, it satisfies the following condition:

4.   Each leading 1 is the only nonzero entry in its column.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=5#h5p-3

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=5#h5p-6

The row-echelon matrices have a "staircase" form, as indicated by the following example (the asterisks indicate arbitrary numbers).

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

The leading 1s proceed "down and to the right" through the matrix. Entries above and to the right of the leading 1s are arbitrary, but all entries below and to the left of them are zero. Hence, a matrix in row-echelon form is in reduced form if, in addition, the entries directly above each leading 1 are all zero. Note that a matrix in row-echelon form can, with a few more row operations, be carried to reduced form (use row operations to create zeros above each leading one in succession, beginning from the right).

The importance of row-echelon matrices comes from the following theorem.

Theorem 1.2.1

Every matrix can be brought to (reduced) row-echelon form by a sequence of elementary row operations.

In fact we can give a step-by-step procedure for actually finding a row-echelon matrix. Observe that while there are many sequences of row operations that will bring a matrix to row-echelon form, the

one we use is systematic and is easy to program on a computer. Note that the algorithm deals with matrices in general, possibly with columns of zeros.

Gaussian Algorithm

Step 1. If the matrix consists entirely of zeros, stop—it is already in row-echelon form.

Step 2. Otherwise, find the first column from the left containing a nonzero entry (call it a), and move the row containing that entry to the top position.

Step 3. Now multiply the new top row by $1/a$ to create a leading 1.

Step 4. By subtracting multiples of that row from rows below it, make each entry below the leading 1 zero. This completes the first row, and all further row operations are carried out on the remaining rows.

Step 5. Repeat steps 1–4 on the matrix consisting of the remaining rows.

The process stops when either no rows remain at step 5 or the remaining rows consist entirely of zeros.

Observe that the gaussian algorithm is recursive: When the first leading 1 has been obtained, the procedure is repeated on the remaining rows of the matrix. This makes the algorithm easy to use on a computer. Note that the solution to Example 1.1.3 did not use the gaussian algorithm as written because the first leading 1 was not created by dividing row 1 by 3. The reason for this is that it avoids

fractions. However, the general pattern is clear: Create the leading 1s from left to right, using each of them in turn to create zeros below it. Here is one example.

Solution:

The corresponding augmented matrix is

$$
\left[\begin{array}{ccc|c}
3 & 1 & -4 & -1 \\
1 & 0 & 10 & 5 \\
4 & 1 & 6 & 1
\end{array} \right]
$$

Create the first leading one by interchanging rows 1 and 2

$$
\left[\begin{array}{ccc|c}
1 & 0 & 10 & 5 \\
3 & 1 & -4 & -1 \\
4 & 1 & 6 & 1
\end{array} \right]
$$

Now subtract 3 times row 1 from row 2, and subtract 4 times row 1 from row 3. The result is

$$\begin{bmatrix} 1 & 0 & 10 & 5 \\ 0 & 1 & -34 & -16 \\ 0 & 1 & -34 & -19 \end{bmatrix}$$

Now subtract row 2 from row 3 to obtain

$$\begin{bmatrix} 1 & 0 & 10 & 5 \\ 0 & 1 & -34 & -16 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

This means that the following reduced system of equations

$$\begin{aligned} x \quad + 10z &= \quad 5 \\ y - 34z &= -16 \\ 0 &= \ -3 \end{aligned}$$

is equivalent to the original system. In other words, the two have the same solutions. But this last system clearly has no solution (the last equation requires that x, y and z satisfy $0x + 0y + 0z = -3$, and no such numbers exist). Hence the original system has no solution.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=5#h5p-9

To solve a linear system, the augmented matrix is carried to reduced row-echelon form, and the variables corresponding to the leading ones are called **leading variables**. Because the matrix is in reduced form, each leading variable occurs in exactly one equation, so that equation can be solved to give a formula for the leading variable

in terms of the nonleading variables. It is customary to call the nonleading variables "free" variables, and to label them by new variables $s, t, \ldots$, called **parameters**. Every choice of these parameters leads to a solution to the system, and every solution arises in this way. This procedure works in general, and has come to be called

---

Gaussian Elimination

To solve a system of linear equations proceed as follows:

1. Carry the augmented matrix\index{augmented matrix}\index{matrix!augmented matrix} to a reduced row-echelon matrix using elementary row operations.
2. If a row $\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$ occurs, the system is inconsistent.
3. Otherwise, assign the nonleading variables (if any) as parameters, and use the equations corresponding to the reduced row-echelon matrix to solve for the leading variables in terms of the parameters.

There is a variant of this procedure, wherein the augmented matrix is carried only to row-echelon form. The nonleading variables are assigned as parameters as before. Then the last equation (corresponding to the row-echelon form) is used to solve for the last leading variable in terms of the parameters. This last leading variable is then substituted into all the preceding equations. Then, the second last equation yields the second last leading variable,

which is also substituted back. The process continues to give the general solution. This procedure is called **back-substitution**. This procedure can be shown to be numerically more efficient and so is important when solving very large systems.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=5#h5p-10

Rank

It can be proven that the *reduced* row-echelon form of a matrix A is uniquely determined by A. That is, no matter which series of row operations is used to carry A to a reduced row-echelon matrix, the result will always be the same matrix. By contrast, this is not true for row-echelon matrices: Different series of row operations can carry the same matrix A to different row-echelon matrices. Indeed, the matrix $A = \begin{bmatrix} 1 & -1 & 4 \\ 2 & -1 & 2 \end{bmatrix}$ can be carried (by one row operation) to the row-echelon matrix $\begin{bmatrix} 1 & -1 & 4 \\ 0 & 1 & -6 \end{bmatrix}$, and then by another row operation to the (reduced) row-echelon matrix $\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -6 \end{bmatrix}$. However, it **is** true that the number r of leading 1s must be the same in each of these row-echelon matrices (this will be proved later). Hence, the number r depends only on A and not on the way in which A is carried to row-echelon form.

Example 1.2.5

Compute the rank of $A = \begin{bmatrix} 1 & 1 & -1 & 4 \\ 2 & 1 & 3 & 0 \\ 0 & 1 & -5 & 8 \end{bmatrix}$.

Solution:

The reduction of A to row-echelon form is

$$A = \begin{bmatrix} 1 & 1 & -1 & 4 \\ 2 & 1 & 3 & 0 \\ 0 & 1 & -5 & 8 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & -1 & 4 \\ 0 & -1 & 5 & -8 \\ 0 & 1 & -5 & 8 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & -1 & 4 \\ 0 & 1 & -5 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Because this row-echelon matrix has two leading 1s, rank $A = 2$
.

Suppose that rank $A = r$, where A is a matrix with m rows and n columns. Then $r \leq m$ because the leading 1s lie in different rows, and $r \leq n$ because the leading 1s lie in different columns. Moreover, the rank has a useful application to equations. Recall that a system of linear equations is called consistent if it has at least one solution.

Theorem 1.2.2

Suppose a system of m equations in n variables is **consistent**, and that the rank of the augmented matrix is r.

1. The set of solutions involves exactly $n - r$ parameters.
2. If $r < n$, the system has infinitely many solutions.
3. If $r = n$, the system has a unique solution.

Proof:

The fact that the rank of the augmented matrix is r means there are exactly r leading variables, and hence exactly $n - r$ nonleading variables. These nonleading variables are all assigned

as parameters in the gaussian algorithm, so the set of solutions involves exactly $n - r$ parameters. Hence if $r < n$, there is at least one parameter, and so infinitely many solutions. If $r = n$, there are no parameters and so a unique solution.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=5#h5p-11

Theorem 1.2.2 shows that, for any system of linear equations, exactly three possibilities exist:

1. **No solution**. This occurs when a row $\begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$ occurs in the row-echelon form. This is the case where the system is inconsistent.
2. **Unique solution**. This occurs when every variable is a leading variable.
3. **Infinitely many solutions**. This occurs when the system is consistent and there is at least one nonleading variable, so at least one parameter is involved.

GeoGebra Exercise: Linear Systems:

https://www.geogebra.org/m/cwQ9uYCZ
Please answer these questions after you open the webpage:

1. For the given linear system, what does each one of them represent?

2. Based on the graph, what can we say about the solutions? Does the system have one solution, no solution or infinitely many solutions? Why

3. Change the constant term in every equation to 0, what changed in the graph?

4. For the following linear system:

$$*x + y = 0, y + z = 0, x + z = 0$$

Can you solve it using Gaussian elimination? When you look at the graph, what do you observe?

Many important problems involve **linear inequalities** rather than **linear equations** For example, a condition on the variables x and y might take the form of an inequality $2x - 5y \leq 4$ rather than an equality $2x - 5y = 4$. There is a technique (called the **simplex algorithm**) for finding solutions to a system of such inequalities that maximizes a function of the form $p = ax + by$ where a and b are fixed constants.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=5#h5p-137

1.3 Homogeneous equations

A system of equations in the variables $x_1, x_2, \ldots, x_n$ is called **homogeneous** if all the constant terms are zero—that is, if each equation of the system has the form

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0$$

Clearly $x_1 = 0, x_2 = 0, \ldots, x_n = 0$ is a solution to such a system; it is called the **trivial solution**. Any solution in which at least one variable has a nonzero value is called a **nontrivial solution.**

Our chief goal in this section is to give a useful condition for a homogeneous system to have nontrivial solutions. The following example is instructive.

Example 1.3.1

Show that the following homogeneous system has nontrivial solutions.

$$x_1 - x_2 + 2x_3 - x_4 = 0$$
$$2x_1 + 2x_2 + x_4 = 0$$
$$3x_1 + x_2 + 2x_3 - x_4 = 0$$

Solution:

The reduction of the augmented matrix to reduced row-echelon form is outlined below.

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

The leading variables are x_1, x_2, and x_4, so x_3 is assigned as a parameter—say $x_3 = t$. Then the general solution is $x_1 = -t$, $x_2 = t$, $x_3 = t$, $x_4 = 0$. Hence, taking $t = 1$ (say), we get a nontrivial solution: $x_1 = -1, x_2 = 1, x_3 = 1, x_4 = 0$.

The existence of a nontrivial solution in Example 1.3.1 is ensured by the presence of a parameter in the solution. This is due to the fact that there is a **nonleading** variable (x_3 in this case). But there *must* be a nonleading variable here because there are four variables and only three equations (and hence *at most* three leading variables). This discussion generalizes to a proof of the following fundamental theorem.

Theorem 1.3.1

If a homogeneous system of linear equations has more variables than equations, then it has a nontrivial solution (in fact, infinitely many).

Proof:

Suppose there are m equations in n variables where $n > m$, and let R denote the reduced row-echelon form of the augmented matrix. If there are r leading variables, there are $n - r$ nonleading variables, and so $n - r$ parameters. Hence, it suffices to show that $r < n$. But $r \leq m$ because R has r leading 1s and m rows, and $m < n$ by hypothesis. So $r \leq m < n$, which gives $r < n$.

Note that the converse of Theorem 1.3.1 is not true: if a homogeneous system has nontrivial solutions, it need not have more variables than equations (the system $x_1 + x_2 = 0$, $2x_1 + 2x_2 = 0$ has nontrivial solutions but $m = 2 = n$.)

Theorem 1.3.1 is very useful in applications. The next example provides an illustration from geometry.

Example 1.3.2

We call the graph of an equation
$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$ a **conic** if the numbers a, b, and c are not all zero. Show that there is at least one conic through any five points in the plane that are not all on a line.

Solution:

Let the coordinates of the five points be (p_1, q_1), (p_2, q_2), (p_3, q_3), (p_4, q_4), and (p_5, q_5). The graph of $ax^2 + bxy + cy^2 + dx + ey + f = 0$ passes through (p_i, q_i) if

$$ap_i^2 + bp_iq_i + cq_i^2 + dp_i + eq_i + f = 0$$

This gives five equations, one for each i, linear in the six variables $a, b, c, d, e,$ and f. Hence, there is a nontrivial solution by Theorem 1.1.3. If $a = b = c = 0$, the five points all lie on the line with equation $dx + ey + f = 0$, contrary to assumption. Hence, one of a, b, c is nonzero.

Linear Combinations and Basic Solutions

As for rows, two columns are regarded as **equal** if they have the same number of entries and corresponding entries are the same. Let x and y be columns with the same number of entries. As for elementary row operations, their **sum** $x + y$ is obtained by adding corresponding entries and, if k is a number, the **scalar product** kx is defined by multiplying each entry of x by k. More precisely:

$$\text{If } x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ and } y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \text{ then } x+y = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \text{ and } kx = \begin{bmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_n \end{bmatrix}.$$

A sum of scalar multiples of several columns is called a **linear combination** of these columns. For example, $sx + ty$ is a linear combination of x and y for any choice of numbers s and t.

Example 1.3.3

If $x = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

then $2x + 5y = \begin{bmatrix} 6 \\ -4 \end{bmatrix} + \begin{bmatrix} -5 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$

Example 1.3.4

Let $x = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, y = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$

and $z = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$. If $v = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}$

and $w = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$,

determine whether v and w are linear combinations of x, y and z.

Solution:

For v, we must determine whether numbers r, s, and t exist such that $v = rx + sy + tz$, that is, whether

$$\begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} = r \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + s \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} r + 2s + 3t \\ s + t \\ r + t \end{bmatrix}$$

Equating corresponding entries gives a system of linear equations $r + 2s + 3t = 0$, $s + t = -1$, and $r + t = 2$ for r, s, and t. By gaussian elimination, the solution is $r = 2 - k$, $s = -1 - k$, and $t = k$ where k is a parameter. Taking $k = 0$, we see that $v = 2x - y$ is a linear combination of x, y, and z.

Turning to w, we again look for r, s, and t such that $w = rx + sy + tz$; that is,

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = r \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + s \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} r + 2s + 3t \\ s + t \\ r + t \end{bmatrix}$$

leading to equations $r + 2s + 3t = 1$, $s + t = 1$, and $r + t = 1$ for real numbers r, s, and t. But this time there is **no** solution as the reader can verify, so w is **not** a linear combination of x, y, and z.

Our interest in linear combinations comes from the fact that they provide one of the best ways to describe the general solution of a homogeneous system of linear equations. When solving such a system with n variables $x_1, x_2, \ldots, x_n$, write the

variables as a column matrix: $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$. The trivial solution is

denoted $0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. As an illustration, the general solution in

Example 1.3.1 is $x_1 = -t, x_2 = t, x_3 = t$, and $x_4 = 0$, where t is a parameter, and we would now express this by

saying that the general solution is $x = \begin{bmatrix} -t \\ t \\ t \\ 0 \end{bmatrix}$, where t is

arbitrary.

Now let x and y be two solutions to a homogeneous system with n variables. Then any linear combination $sx + ty$ of these solutions turns out to be again a solution to the system. More generally:

Any linear combination of solutions to a homogeneous system is again a solution.

In fact, suppose that a typical equation in the system is $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0$, and suppose that

$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ are solutions. Then

$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0$ and

$$a_1 y_1 + a_2 y_2 + \cdots + a_n y_n = 0.$$

Hence $sx + ty = \begin{bmatrix} sx_1 + ty_1 \\ sx_2 + ty_2 \\ \vdots \\ sx_n + ty_n \end{bmatrix}$ is also a solution because

$$a_1(sx_1 + ty_1) + a_2(sx_2 + ty_2) + \cdots + a_n(sx_n + ty_n)$$
$$= [a_1(sx_1) + a_2(sx_2) + \cdots + a_n(sx_n)] + [a_1(ty_1) + a_2(ty_2) + \cdots + a_n(ty_n)]$$
$$= s(a_1 x_1 + a_2 x_2 + \cdots + a_n x_n) + t(a_1 y_1 + a_2 y_2 + \cdots + a_n y_n)$$
$$= s(0) + t(0)$$
$$= 0$$

A similar argument shows that Statement 1.1 is true for linear combinations of more than two solutions.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=5#h5p-129

The remarkable thing is that **every** solution to a homogeneous system is a linear combination of certain particular solutions and, in fact, these solutions are easily computed using the gaussian algorithm. Here is an example.

Example 1.3.5

Solve the homogeneous system with coefficient matrix

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

Solution:

The reduction of the augmented matrix to reduced form is

$$\begin{bmatrix} 1 & -2 & 3 & -2 & | & 0 \\ -3 & 6 & 1 & 0 & | & 0 \\ -2 & 4 & 4 & -2 & | & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -2 & 0 & -\frac{1}{5} & | & 0 \\ 0 & 0 & 1 & -\frac{3}{5} & | & 0 \\ 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$$

so the solutions are $x_1 = 2s + \frac{1}{5}t$, $x_2 = s$, $x_3 = \frac{3}{5}$, and $x_4 = t$ by gaussian elimination. Hence we can write the general solution x in the matrix form

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2s + \frac{1}{5}t \\ s \\ \frac{3}{5}t \\ t \end{bmatrix} = s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} \frac{1}{5} \\ 0 \\ \frac{3}{5} \\ 1 \end{bmatrix} = sx_1 + tx_2.$$

Here $x_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $x_2 = \begin{bmatrix} \frac{1}{5} \\ 0 \\ \frac{3}{5} \\ 1 \end{bmatrix}$ are particular solutions
determined by the gaussian algorithm.

The solutions x_1 and x_2 in Example 1.3.5 are denoted as follows:

Definition 1.5 Basic Solutions

The gaussian algorithm systematically produces solutions to any homogeneous linear system, called **basic solutions**, one for every parameter.

Moreover, the algorithm gives a routine way to express ***every*** solution as a linear combination of basic solutions as in Example 1.3.5, where the general solution x becomes

$$x = s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} \frac{1}{5} \\ 0 \\ \frac{3}{5} \\ 1 \end{bmatrix} = s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \frac{1}{5}t \begin{bmatrix} 1 \\ 0 \\ 3 \\ 5 \end{bmatrix}$$

Hence by introducing a new parameter $r = t/5$ we can multiply the original basic solution x_2 by 5 and so eliminate fractions.

For this reason:

Convention:

Any nonzero scalar multiple of a basic solution will still be called a basic solution.

In the same way, the gaussian algorithm produces basic solutions to *every* homogeneous system, one for each parameter (there are *no* basic solutions if the system has only the trivial solution). Moreover every solution is given by the algorithm as a linear combination of

these basic solutions (as in Example 1.3.5). If A has rank r, Theorem 1.2.2 shows that there are exactly $n - r$ parameters, and so $n - r$ basic solutions. This proves:

Theorem 1.3.2

Let A be an $m \times n$ matrix of rank r, and consider the homogeneous system in n variables with A as coefficient matrix. Then:

1. The system has exactly $n - r$ basic solutions, one for each parameter.
2. Every solution is a linear combination of these basic solutions.

Example 1.3.6

Find basic solutions of the homogeneous system with coefficient matrix A, and express every solution as a linear combination of the basic solutions, where

$$A = \begin{bmatrix} 1 & -3 & 0 & 2 & 2 \\ -2 & 6 & 1 & 2 & -5 \\ 3 & -9 & -1 & 0 & 7 \\ -3 & 9 & 2 & 6 & -8 \end{bmatrix}$$

Solution:

The reduction of the augmented matrix to reduced row-echelon form is

$$\begin{bmatrix} 1 & -3 & 0 & 2 & 2 & | & 0 \\ -2 & 6 & 1 & 2 & -5 & | & 0 \\ 3 & -9 & -1 & 0 & 7 & | & 0 \\ -3 & 9 & 2 & 6 & -8 & | & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -3 & 0 & 2 & 2 & | & 0 \\ 0 & 0 & 1 & 6 & -1 & | & 0 \\ 0 & 0 & 0 & 0 & 0 & | & 0 \\ 0 & 0 & 0 & 0 & 0 & | & 0 \end{bmatrix}$$

so the general solution is $x_1 = 3r - 2s - 2t$, $x_2 = r$, $x_3 = -6s + t$, $x_4 = s$, and $x_5 = t$ where r, s, and t are parameters. In matrix form this is

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 3r - 2s - 2t \\ r \\ -6s + t \\ s \\ t \end{bmatrix} = r \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -2 \\ 0 \\ -6 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

Hence basic solutions are

$$x_1 = \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ x_2 = \begin{bmatrix} -2 \\ 0 \\ -6 \\ 1 \\ 0 \end{bmatrix}, \ x_3 = \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=5#h5p-140

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=5#h5p-128

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=5#h5p-13

2. Matrix Algebra

Introduction

In the study of systems of linear equations in Chapter 1, we found it convenient to manipulate the augmented matrix of the system. Our aim was to reduce it to row-echelon form (using elementary row operations) and hence to write down all solutions to the system. In the present chapter we consider matrices for their own sake. While some of the motivation comes from linear equations, it turns out that matrices can be multiplied and added and so form an algebraic system somewhat analogous to the real numbers. This "matrix algebra" is useful in ways that are quite different from the study of linear equations. For example, the geometrical transformations obtained by rotating the euclidean plane about the origin can be viewed as multiplications by certain 2×2 matrices. These "matrix transformations" are an important tool in geometry and, in turn, the geometry provides a "picture" of the matrices. Furthermore, matrix algebra has many other applications, some of which will be explored in this chapter. This subject is quite old and was first studied systematically in 1858 by Arthur Cayley.

Arthur Cayley (1821-1895) showed his mathematical talent early and graduated from Cambridge in 1842 as senior wrangler. With no employment in mathematics in view, he took legal training and worked as a lawyer while continuing to do mathematics, publishing nearly 300 papers in fourteen years. Finally, in 1863, he accepted the Sadlerian professorship in Cambridge and remained

there for the rest of his life, valued for his administrative and teaching skills as well as for his scholarship. His mathematical achievements were of the first rank. In addition to originating matrix theory and the theory of determinants, he did fundamental work in group theory, in higher-dimensional geometry, and in the theory of invariants. He was one of the most prolific mathematicians of all time and produced 966 papers.

2.1 Matrix Addition, Scalar Multiplication, and Transposition

A rectangular array of numbers is called a **matrix** (the plural is **matrices**), and the numbers are called the **entries** of the matrix. Matrices are usually denoted by uppercase letters: A, B, C, and so on. Hence,

$$A = \left[\begin{array}{ccc} 1 & 2 & -1 \\ 0 & 5 & 6 \end{array} \right] \quad B = \left[\begin{array}{cc} 1 & -1 \\ 0 & 2 \end{array} \right] \quad C = \left[\begin{array}{c} 1 \\ 3 \\ 2 \end{array} \right]$$

are matrices. Clearly matrices come in various shapes depending on the number of **rows** and **columns**. For example, the matrix A shown has 2 rows and 3 columns. In general, a matrix with m rows and n columns is referred to as an $m \times n$ matrix or as having **size** $m \times n$. Thus matrices A, B, and C above have sizes 2×3

$, 2 \times 2$, and 3×1, respectively. A matrix of size $1 \times n$ is called a **row matrix**, whereas one of size $m \times 1$ is called a **column matrix**. Matrices of size $n \times n$ for some n are called **square** matrices.

Each entry of a matrix is identified by the row and column in which it lies. The rows are numbered from the top down, and the columns are numbered from left to right. Then the (i, j)**-entry** of a matrix is the number lying simultaneously in row i and column j. For example,

$$\text{The } (1, 2)\text{-entry of } \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \text{ is } -1.$$

$$\text{The } (2, 3)\text{-entry of } \begin{bmatrix} 1 & 2 & -1 \\ 0 & 5 & 6 \end{bmatrix} \text{ is } 6.$$

A special notation is commonly used for the entries of a matrix. If A is an $m \times n$ matrix, and if the (i, j)-entry of A is denoted as a_{ij}, then A is displayed as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

This is usually denoted simply as $A = [a_{ij}]$. Thus a_{ij} is the entry in row i and column j of A. For example, a 3×4 matrix in this notation is written

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$$

It is worth pointing out a convention regarding rows and columns: *Rows are mentioned before columns*. For example:

- If a matrix has size $m \times n$, it has m rows and n columns.
- If we speak of the (i, j)-entry of a matrix, it lies in row i and column j.
- If an entry is denoted a_{ij}, the first subscript i refers to the row and the second subscript j to the column in which a_{ij} lies.

Two points (x_1, y_1) and (x_2, y_2) in the plane are equal if and only if they have the same coordinates, that is $x_1 = x_2$ and $y_1 = y_2$. Similarly, two matrices A and B are called **equal** (written $A = B$) if and only if:

1. They have the same size.
2. Corresponding entries are equal.

If the entries of A and B are written in the form $A = [a_{ij}]$, $B = [b_{ij}]$, described earlier, then the second condition takes the following form:

$$A = [a_{ij}] = [b_{ij}] \text{ means } a_{ij} = b_{ij} \text{ for all } i \text{ and } j$$

Example 2.1.1

Given $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, B = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 0 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$ discuss the possibility that $A = B, B = C, A = C$.

Solution:
$A = B$ is impossible because A and B are of different sizes:

A is 2×2 whereas B is 2×3. Similarly, $B = C$ is impossible. But $A = C$ is possible provided that corresponding entries are equal:

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

means $a = 1, b = 0, c = -1$, and $d = 2$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=66#h5p-14

Matrix Addition

Definition 2.1 Matrix Addition

If A and B are matrices of the same size, their **sum** $A + B$ is the matrix formed by adding corresponding entries.

If $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ and $B = \begin{bmatrix} b_{ij} \end{bmatrix}$, this takes the form
$$A + B = \begin{bmatrix} a_{ij} + b_{ij} \end{bmatrix}$$

Note that addition is **not** defined for matrices of different sizes.

Example 2.1.2

If $A = \begin{bmatrix} 2 & 1 & 3 \\ -1 & 2 & 0 \end{bmatrix}$

and $B = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & 6 \end{bmatrix}$,

compute $A + B$.

Solution:

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

Example 2.1.3

Find a, b, and c if

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

Solution:

Add the matrices on the left side to obtain

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

Because corresponding entries must be equal, this gives three equations: $a + c = 3$, $b + a = 2$, and $c + b = -1$. Solving these yields $a = 3, b = -1, c = 0$.

If A, B, and C are any matrices **of the same size**, then

(commutative   law)

$$A + B = B + A$$
$$A + (B + C) = (A + B) + C \qquad \text{(associative law)}$$

In fact, if $A = [a_{ij}]$ and $B = [b_{ij}]$, then the (i, j)-entries of $A + B$ and $B + A$ are, respectively, $a_{ij} + b_{ij}$ and $b_{ij} + a_{ij}$. Since these are equal for all i and j, we get

$$A + B = \begin{bmatrix} a_{ij} + b_{ij} \end{bmatrix} = \begin{bmatrix} b_{ij} + a_{ij} \end{bmatrix} = B + A$$

The associative law is verified similarly.

The $m \times n$ matrix in which every entry is zero is called the $m \times n$ **zero matrix** and is denoted as 0 (or 0_{mn} if it is important to emphasize the size). Hence,

$$0 + X = X$$

holds for all $m \times n$ matrices X. The **negative** of an $m \times n$ matrix A (written $-A$) is defined to be the $m \times n$ matrix obtained by multiplying each entry of A by -1. If $A = [a_{ij}]$, this becomes $-A = [-a_{ij}]$. Hence,

$$A + (-A) = 0$$

holds for all matrices A where, of course, 0 is the zero matrix of the same size as A.

A closely related notion is that of subtracting matrices. If A and B are two $m \times n$ matrices, their **difference** $A - B$ is defined by

$$A - B = A + (-B)$$

Note that if $A = [a_{ij}]$ and $B = [b_{ij}]$, then

$$A - B = [a_{ij}] + [-b_{ij}] = [a_{ij} - b_{ij}]$$

is the $m \times n$ matrix formed by **subtracting** corresponding entries.

Example 2.1.4

Let $A = \begin{bmatrix} 3 & -1 & 0 \\ 1 & 2 & -4 \end{bmatrix}$,

$B = \begin{bmatrix} 1 & -1 & 1 \\ -2 & 0 & 6 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 1 & 1 \end{bmatrix}.$

Compute $-A, A - B,$ and $A + B - C.$

Solution:

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

$$A - B = \begin{bmatrix} 3-1 & -1-(-1) & 0-1 \\ 1-(-2) & 2-0 & -4-6 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -1 \\ 3 & 2 & -10 \end{bmatrix}$$

$$A + B - C = \begin{bmatrix} 3+1-1 & -1-1-0 & 0+1-(-2) \\ 1-2-3 & 2+0-1 & -4+6-1 \end{bmatrix} = \begin{bmatrix} 3 & -2 & 3 \\ -4 & 1 & 1 \end{bmatrix}$$

Example 2.1.5

Solve

$$\begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} + X = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$$

where X is a matrix.

We solve a numerical equation $a + x = b$ by subtracting the number a from both sides to obtain $x = b - a$. This also works for matrices. To solve

$$\begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix} + X = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix}$$

simply subtract the matrix

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

from both sides to get

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

The reader should verify that this matrix X does indeed satisfy the original equation.

The solution in Example 2.1.5 solves the single matrix equation $A + X = B$ directly via matrix subtraction: $X = B - A$. This ability to work with matrices as entities lies at the heart of matrix algebra.

It is important to note that the sizes of matrices involved in some calculations are often determined by the context. For example, if

$$A + C = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 0 & 1 \end{bmatrix}$$

then A and C must be the same size (so that $A + C$ makes sense), and that size must be 2×3 (so that the sum is 2×3). For simplicity we shall often omit reference to such facts when they are clear from the context.

Scalar Multiplication

In gaussian elimination, multiplying a row of a matrix by a number k means multiplying **every** entry of that row by k.

Definition 2.2 Matrix Scalar Multiplication

More generally, if A is any matrix and k is any number, the **scalar multiple** kA is the matrix obtained from A by multiplying each entry of A by k.

The term *scalar* arises here because the set of numbers from which the entries are drawn is usually referred to as the set of scalars. We have been using real numbers as scalars, but we could equally well have been using complex numbers.

Example 2.1.6

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

and $B = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 2 \end{bmatrix}$,

compute $5A, \frac{1}{2}B,$ and $3A - 2B$.

Solution:

$$5A = \begin{bmatrix} 15 & -5 & 20 \\ 10 & 0 & 30 \end{bmatrix}, \quad \frac{1}{2}B = \begin{bmatrix} \frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & \frac{3}{2} & 1 \end{bmatrix}$$

$$3A - 2B = \begin{bmatrix} 9 & -3 & 12 \\ 6 & 0 & 18 \end{bmatrix} - \begin{bmatrix} 2 & 4 & -2 \\ 0 & 6 & 4 \end{bmatrix} = \begin{bmatrix} 7 & -7 & 14 \\ 6 & -6 & 14 \end{bmatrix}$$

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-16

If A is any matrix, note that kA is the same size as A for all scalars k. We also have

$$0A = 0 \quad \text{and} \quad k0 = 0$$

because the zero matrix has every entry zero. In other words, $kA = 0$ if either $k = 0$ or $A = 0$. The converse of this statement is also true, as Example 2.1.7 shows.

Example 2.1.7

If $kA = 0$, show that either $k = 0$ or $A = 0$.

Solution:

Write $A = [a_{ij}]$ so that $kA = 0$ means $ka_{ij} = 0$ for all i and j. If $k = 0$, there is nothing to do. If $k \neq 0$, then $ka_{ij} = 0$ implies that $a_{ij} = 0$ for all i and j; that is, $A = 0$.

For future reference, the basic properties of matrix addition and scalar multiplication are listed in Theorem 2.1.1.

> An interactive H5P element has been excluded from this version of the text. You can view it online here:
>
> https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=66#h5p-29

Theorem 2.1.1

Let A, B, and C denote arbitrary $m \times n$ matrices where m and n are fixed. Let k and p denote arbitrary real numbers. Then

1. $A + B = B + A$.
2. $A + (B + C) = (A + B) + C$.
3. There is an $m \times n$ matrix 0, such that

$0 + A = A$ for each A.

4. For each A there is an $m \times n$ matrix, $-A$, such that $A + (-A) = 0$.
5. $k(A + B) = kA + kB$.
6. $(k + p)A = kA + pA$.
7. $(kp)A = k(pA)$.
8. $1A = A$.

Proof:

Properties 1–4 were given previously. To check Property 5, let $A = [a_{ij}]$ and $B = [b_{ij}]$ denote matrices of the same size. Then $A + B = [a_{ij} + b_{ij}]$, as before, so the (i, j)-entry of $k(A + B)$ is

$$k(a_{ij} + b_{ij}) = ka_{ij} + kb_{ij}$$

But this is just the (i, j)-entry of $kA + kB$, and it follows that $k(A + B) = kA + kB$. The other Properties can be similarly verified; the details are left to the reader.

The Properties in Theorem 2.1.1 enable us to do calculations with matrices in much the same way that

numerical calculations are carried out. To begin, Property 2 implies that the sum

$$(A + B) + C = A + (B + C)$$

is the same no matter how it is formed and so is written as $A + B + C$. Similarly, the sum

$$A + B + C + D$$

is independent of how it is formed; for example, it equals both

$$(A + B) + (C + D) \quad \text{and} \quad A + [B + (C + D)].$$

Furthermore, property 1 ensures that, for example,

$$B + D + A + C = A + B + C + D$$

In other words, the **order** in which the matrices are added does not matter. A similar remark applies to sums of five (or more) matrices.

Properties 5 and 6 in Theorem 2.1.1 are called **distributive laws** for scalar multiplication, and they extend to sums of more than two terms. For example,

$$k(A + B - C) = kA + kB - kC$$
$$(k + p - m)A = kA + pA - mA$$

Similar observations hold for more than three summands. These facts, together with properties 7 and 8, enable us to simplify expressions by collecting like terms, expanding, and taking common factors in exactly the same way that algebraic expressions involving variables and real numbers are manipulated. The following example illustrates these techniques.

Example 2.1.8

Simplify

$$2(A + 3C) - 3(2C - B) - 3\left[2(2A + B - 4C) - 4(A - 2C)\right]$$
where A, B and C are all matrices of the same size.

Solution:

The reduction proceeds as though A, B, and C were variables.

$$2(A + 3C) - 3(2C - B) - 3\left[2(2A + B - 4C) - 4(A - 2C)\right]$$
$$= 2A + 6C - 6C + 3B - 3\left[4A + 2B - 8C - 4A + 8C\right]$$
$$= 2A + 3B - 3\left[2B\right]$$
$$= 2A - 3B$$

Transpose of a Matrix

Many results about a matrix A involve the **rows** of A, and the corresponding result for columns is derived in an analogous way, essentially by replacing the word *row* by the word **column** throughout. The following definition is made with such applications in mind.

> *Definition 2.3 Transpose of a Matrix*
>
> If A is an $m \times n$ matrix, the **transpose** of A, written A^T, is the $n \times m$ matrix whose rows are just the columns of A in the same order.

In other words, the first row of A^T is the first column of A (that is it consists of the entries of column 1 in order). Similarly the second row of A^T is the second column of A, and so on.

Example 2.1.9

Write down the transpose of each of the following matrices.

$$A = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \quad B = \begin{bmatrix} 5 & 2 & 6 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad D = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 3 & 2 \\ -1 & 2 & 1 \end{bmatrix}$$

Solution:

$$A^T = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix}, \ B^T = \begin{bmatrix} 5 \\ 2 \\ 6 \end{bmatrix}, \ C^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}, \ \text{and } D^T = D.$$

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-17

If $A = [a_{ij}]$ is a matrix, write $A^T = [b_{ij}]$. Then b_{ij} is the jth element of the ith row of A^T and so is the jth element of the ith **column** of A. This means $b_{ij} = a_{ji}$, so the definition of A^T can be stated as follows:

(2.1) If $A = [a_{ij}]$, then $A^T = [a_{ji}]$.

This is useful in verifying the following properties of transposition.

Theorem 2.1.2

Let A and B denote matrices of the same size, and let k denote a scalar.

1. If A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix.
2. $(A^T)^T = A$.
3. $(kA)^T = kA^T$.
4. $(A + B)^T = A^T + B^T$.

Proof:

Property 1 is part of the definition of A^T, and Property 2 follows from (2.1). As to Property 3: If $A = a_{ij}$, then $kA = ka_{ij}$, so (2.1) gives

$$(kA)^T = [ka_{ji}] = k[a_{ji}] = kA^T$$

Finally, if $B = [b_{ij}]$, then $A + B = [c_{ij}]$ where $c_{ij} = a_{ij} + b_{ij}$ Then (2.1) gives Property 4:

$$(A+B)^T = [c_{ij}]^T = [c_{ji}] = [a_{ji} + b_{ji}] = [a_{ji}] + [b_{ji}] = A^T + B^T$$

There is another useful way to think of transposition. If $A = [a_{ij}]$ is an $m \times n$ matrix, the elements $a_{11}, a_{22}, a_{33}, \ldots$ are called the **main diagonal** of A. Hence the main diagonal extends down and to the right from the upper left corner of the matrix A; it is shaded in the following examples:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix}$$

Thus forming the transpose of a matrix A can be viewed as "flipping" A about its main diagonal, or as "rotating" A through $180°$ about the line containing the main diagonal. This makes Property 2 in Theorem~?? transparent.

Example 2.1.10

Solve for A if
$$\left(2A^T - 3 \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \right)^T = \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix}.$$

Solution:

Using Theorem 2.1.2, the left side of the equation is

$$\left(2A^T - 3\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}\right)^T = 2\left(A^T\right)^T - 3\begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}^T = 2A - 3\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$$

Hence the equation becomes

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

Thus

$$2A = \begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix} + 3\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 5 & 5 \end{bmatrix}, \qquad \text{so}$$

finally

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

Note that Example 2.1.10 can also be solved by first transposing both sides, then solving for A^T, and so obtaining $A = \left(A^T\right)^T$. The reader should do this.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=66#h5p-39

The matrix $D = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$ in Example 2.1.9 has the property that $D = D^T$. Such matrices are important; a matrix A is called **symmetric** if $A = A^T$. A symmetric matrix A is necessarily square (if A is $m \times n$, then A^T is $n \times m$, so $A = A^T$ forces $n = m$). The name comes from the fact that these matrices exhibit

a symmetry about the main diagonal. That is, entries that are directly across the main diagonal from each other are equal.

For example, $\begin{bmatrix} a & b & c \\ b' & d & e \\ c' & e' & f \end{bmatrix}$ is symmetric when $b = b'$, $c = c'$, and $e = e'$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-141

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-143

Example 2.1.11

If A and B are symmetric $n \times n$ matrices, show that $A + B$ is symmetric.

Solution:

We have $A^T = A$ and $B^T = B$, so, by Theorem 2.1.2, we have $(A + B)^T = A^T + B^T = A + B$. Hence $A + B$ is symmetric.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-18

Example 2.1.12

Suppose a square matrix A satisfies $A = 2A^T$. Show that necessarily $A = 0$.

Solution:

If we iterate the given equation, Theorem 2.1.2 gives

$$A = 2A^T = 2\left[2A^T\right]^T = 2\left[2(A^T)^T\right] = 4A$$

Subtracting A from both sides gives $3A = 0$, so $A = \frac{1}{3}(0) = 0$.

2.2 Matrix-Vector Multiplication

Up to now we have used matrices to solve systems of linear equations by manipulating the rows of the augmented matrix. In this section we introduce a different way of describing linear systems that makes more use of the coefficient matrix of the system and leads to a useful way of "multiplying" matrices.

Vectors

It is a well-known fact in analytic geometry that two points in the plane with coordinates (a_1, a_2) and (b_1, b_2) are equal if and only if $a_1 = b_1$ and $a_2 = b_2$. Moreover, a similar condition applies to points (a_1, a_2, a_3) in space. We extend this idea as follows.

An ordered sequence $(a_1, a_2, \ldots, a_n)$ of real numbers is called an **ordered n-tuple**. The word "ordered" here reflects our insistence that two ordered n-tuples are equal if and only if corresponding entries are the same. In other words,

$$(a_1, a_2, \ldots, a_n) = (b_1, b_2, \ldots, b_n) \quad \text{if and only if} \quad a_1 = b_1, a_2 = b_2, \ldots, \text{ and } a_n = b_n.$$

Thus the ordered 2-tuples and 3-tuples are just the ordered pairs and triples familiar from geometry.

Definition 2.4 The set $\mathbb{R}^n$ of ordered n-tuples of real numbers

Let $\mathbb{R}$ denote the set of all real numbers. The set of all

ordered n-tuples from $\mathbb{R}$ has a special notation:

$\mathbb{R}^n$ denotes the set of all ordered n-tuples of real numbers.

There are two commonly used ways to denote the n-tuples in $\mathbb{R}^n$: As rows $\left(r_1, r_2, \ldots, r_n \right)$ or columns $\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$;
the notation we use depends on the context. In any event they are called **vectors** or n–**vectors** and will be denoted using bold type such as $\mathbf{x}$ or $\mathbf{v}$. For example, an $m \times n$ matrix A will be written as a row of columns:

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \text{ where } \mathbf{a}_j \text{ denotes column } j \text{ of } A \text{ for each } j.$$

If $\mathbf{x}$ and $\mathbf{y}$ are two n-vectors in $\mathbf{R}^n$, it is clear that their matrix sum $\mathbf{x} + \mathbf{y}$ is also in $\mathbf{R}^n$ as is the scalar multiple $k\mathbf{x}$ for any real number k. We express this observation by saying that $\mathbf{R}^n$ is **closed** under addition and scalar multiplication. In particular, all the basic properties in Theorem 2.1.1 are true of these n-vectors. These properties are fundamental and will be used frequently below without comment. As for matrices in general, the $n \times 1$ zero matrix is called the **zero** n–**vector** in $\mathbf{R}^n$ and, if $\mathbf{x}$ is an n-vector, the n-vector $-\mathbf{x}$ is called the **negative** $\mathbf{x}$.

Of course, we have already encountered these n-vectors in Section 1.3 as the solutions to systems of linear equations with n variables. In particular we defined the notion of a linear combination

of vectors and showed that a linear combination of solutions to a homogeneous system is again a solution. Clearly, a linear combination of n-vectors in $\mathbf{R}^n$ is again in $\mathbf{R}^n$, a fact that we will be using.

Matrix-Vector Multiplication

Given a system of linear equations, the left sides of the equations depend only on the coefficient matrix A and the column $\mathbf{x}$ of variables, and not on the constants. This observation leads to a fundamental idea in linear algebra: We view the left sides of the equations as the "product" $A\mathbf{x}$ of the matrix A and the vector $\mathbf{x}$. This simple change of perspective leads to a completely new way of viewing linear systems—one that is very useful and will occupy our attention throughout this book.

To motivate the definition of the "product" $A\mathbf{x}$, consider first the following system of two equations in three variables:

$$\begin{aligned} ax_1 + bx_2 + cx_3 &= b_1 \\ a'x_1 + b'x_2 + c'x_3 &= b_1 \end{aligned}$$

(2.2)

and let $A = \begin{bmatrix} a & b & c \\ a' & b' & c' \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$

denote the coefficient matrix, the variable matrix, and the constant matrix, respectively. The system (2.2) can be expressed as a single vector equation

$$\begin{bmatrix} ax_1 + bx_2 + cx_3 \\ a'x_1 + b'x_2 + c'x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

which in turn can be written as follows:

$$x_1 \begin{bmatrix} a \\ a' \end{bmatrix} + x_2 \begin{bmatrix} b \\ b' \end{bmatrix} + x_3 \begin{bmatrix} c \\ c' \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Now observe that the vectors appearing on the left side are just the columns

$$\mathbf{a}_1 = \begin{bmatrix} a \\ a' \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} b \\ b' \end{bmatrix}, \text{ and } \mathbf{a}_3 = \begin{bmatrix} c \\ c' \end{bmatrix}$$

of the coefficient matrix A. Hence the system (2.2) takes the form

(2.3) $\quad x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{b}$

This shows that the system (2.2) has a solution if and only if the constant matrix $\mathbf{b}$ is a linear combination of the columns of A, and that in this case the entries of the solution are the coefficients x_1, x_2, and x_3 in this linear combination.

Moreover, this holds in general. If A is any $m \times n$ matrix, it is often convenient to view A as a row of columns. That is, if $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are the columns of A, we write

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

and say that $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ is *given in terms of its columns*.

Now consider any system of linear equations with $m \times n$ coefficient matrix A. If $\mathbf{b}$ is the constant matrix of the system, and

if $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

is the matrix of variables then, exactly as above, the system can be written as a single vector equation

(2.4) $\quad x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$

Example 2.2.1

Write the system

$$\begin{cases} 3x_1 + 2x_2 - 4x_3 = 0 \\ x_1 - 3x_2 + x_3 = 3 \\ x_2 - 5x_3 = -1 \end{cases}$$

in the form given in (2.4).

Solution:

$$x_1 \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} -4 \\ 1 \\ -5 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}$$

As mentioned above, we view the left side of (2.4) as the **product** of the matrix A and the vector $\mathbf{x}$. This basic idea is formalized in the following definition:

Definition 2.5 Matrix-Vector Multiplication

Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ be an $m \times n$ matrix, written in terms of its columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. If

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

is any n-vector, the product $A\mathbf{x}$ is defined to be the m-vector given by:

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

In other words, if A is $m \times n$ and $\mathbf{x}$ is an n-vector, the product $A\mathbf{x}$ is the linear combination of the columns of A where the coefficients are the entries of $\mathbf{x}$ (in order).

Note that if A is an $m \times n$ matrix, the product $A\mathbf{x}$ is only defined if $\mathbf{x}$ is an n-vector and then the vector $A\mathbf{x}$ is an m-vector because this is true of each column $\mathbf{a}_j$ of A. But in this case the **system** of linear equations with coefficient matrix A and constant vector $\mathbf{b}$ takes the form of a ***single*** matrix equation

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

The following theorem combines Definition 2.5 and equation (2.4) and summarizes the above discussion. Recall that a system of linear equations is said to be ***consistent*** if it has at least one solution.

Theorem 2.2.1

1. Every system of linear equations has the form $A\mathbf{x} = \mathbf{b}$ where A is the coefficient matrix, $\mathbf{b}$ is the constant matrix, and $\mathbf{x}$ is the matrix of variables.
2. The system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is a linear combination of the columns of A.
3. If $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ are the columns of A and if

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$, then $\mathbf{x}$ is a solution to the linear

system $A\mathbf{x} = \mathbf{b}$ if and only if $x_1, x_2, \ldots, x_n$ are a solution of the vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

A system of linear equations in the form $A\mathbf{x} = \mathbf{b}$ as in (1) of Theorem 2.2.1 is said to be written in **matrix form**. This is a useful way to view linear systems as we shall see.

Theorem 2.2.1 transforms the problem of solving the linear system $A\mathbf{x} = \mathbf{b}$ into the problem of expressing the constant matrix B as a linear combination of the columns of the coefficient matrix A. Such a change in perspective is very useful because one approach or the other may be better in a particular situation; the importance of the theorem is that there is a choice.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=66#h5p-41

Example 2.2.2

If $A = \begin{bmatrix} 2 & -1 & 3 & 5 \\ 0 & 2 & -3 & 1 \\ -3 & 4 & 1 & 2 \end{bmatrix}$ and

$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -2 \end{bmatrix}$, compute $A\mathbf{x}$.

Solution:

By Definition 2.5:

$$A\mathbf{x} = 2\begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix} + 1\begin{bmatrix} -1 \\ 2 \\ 4 \end{bmatrix} + 0\begin{bmatrix} 3 \\ -3 \\ 1 \end{bmatrix} - 2\begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -7 \\ 0 \\ -6 \end{bmatrix}$$

.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-40

Example 2.2.3

Given columns $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3,$ and $\mathbf{a}_4$ in $\mathbf{R}^3$, write $2\mathbf{a}_1 - 3\mathbf{a}_2 + 5\mathbf{a}_3 + \mathbf{a}_4$ in the form $A\mathbf{x}$ where A is a matrix and $\mathbf{x}$ is a vector.

Solution:

Here the column of coefficients is

$$\mathbf{x} = \begin{bmatrix} 2 \\ -3 \\ 5 \\ 1 \end{bmatrix}.$$

Hence Definition 2.5 gives

$$A\mathbf{x} = 2\mathbf{a}_1 - 3\mathbf{a}_2 + 5\mathbf{a}_3 + \mathbf{a}_4$$

where $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{bmatrix}$ is the matrix with $\mathbf{a}_1, \mathbf{a}_2$, $\mathbf{a}_3$, and $\mathbf{a}_4$ as its columns.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-142

Example 2.2.4

Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{bmatrix}$ be the 3×4

matrix given in terms of its columns

$$\mathbf{a}_1 = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 3 \\ -1 \\ -3 \end{bmatrix}, \text{ and}$$

$$\mathbf{a}_4 = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}.$$

In each case below, either express $\mathbf{b}$ as a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, and $\mathbf{a}_4$, or show that it is not such a linear combination. Explain what your answer means for the corresponding system $A\mathbf{x} = \mathbf{b}$ of linear equations.

1. $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

2. $\mathbf{b} = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$

Solution:

By Theorem 2.2.1, $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, and $\mathbf{a}_4$ if and only if the system $A\mathbf{x} = \mathbf{b}$ is consistent (that is, it has a solution). So in each case we carry the augmented matrix $[A|b]$ of the system $A\mathbf{x} = \mathbf{b}$ to reduced form.

1. Here

$$\begin{bmatrix} 2 & 1 & 3 & 3 & | & 1 \\ 0 & 1 & -1 & 1 & | & 2 \\ -1 & 1 & -3 & 0 & | & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 2 & 1 & | & 0 \\ 0 & 1 & -1 & 1 & | & 0 \\ 0 & 0 & 0 & 0 & | & 1 \end{bmatrix}, \text{ so}$$

the system $A\mathbf{x} = \mathbf{b}$ has no solution in this case. Hence $\mathbf{b}$ is
\textit{not} a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, and $\mathbf{a}_4$.

2. Now

$$
\left[\begin{array}{rrrr|r}
2 & 1 & 3 & 3 & 4 \\
0 & 1 & -1 & 1 & 2 \\
-1 & 1 & -3 & 0 & 1
\end{array}\right]
\rightarrow
\left[\begin{array}{rrrr|r}
1 & 0 & 2 & 1 & 1 \\
0 & 1 & -1 & 1 & 2 \\
0 & 0 & 0 & 0 & 0
\end{array}\right], \text{ so}
$$

the system $A\mathbf{x} = \mathbf{b}$ is consistent.

Thus $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, and $\mathbf{a}_4$ in this
case. In fact the general solution is $x_1 = 1 - 2s - t$,
$x_2 = 2 + s - t$, $x_3 = s$, and $x_4 = t$ where s and t are
arbitrary   parameters.   Hence

$$
x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 + x_4\mathbf{a}_4 = \mathbf{b} =
\left[\begin{array}{r}
4 \\
2 \\
1
\end{array}\right]
$$

for **any** choice of s and t. If we take $s = 0$ and $t = 0$, this
becomes $\mathbf{a}_1 + 2\mathbf{a}_2 = \mathbf{b}$, whereas taking $s = 1 = t$ gives
$-2\mathbf{a}_1 + 2\mathbf{a}_2 + \mathbf{a}_3 + \mathbf{a}_4 = \mathbf{b}$.

Example 2.2.5

Taking A to be the zero matrix, we have $0\mathbf{x} = \mathbf{0}$ for all
vectors $\mathbf{x}$ by Definition 2.5 because every column of the
zero matrix is zero. Similarly, $A\mathbf{0} = \mathbf{0}$ for all matrices A
because every entry of the zero vector is zero.

If $I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, show that $I\mathbf{x} = \mathbf{x}$ for any vector $\mathbf{x}$ in $\mathbf{R}^3$.

Solution:

If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, then Definition 2.5 gives

$$I\mathbf{x} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{x}$$

The matrix I in Example 2.2.6 is called the 3×3 **identity matrix**, and we will encounter such matrices again in future. Before proceeding, we develop some algebraic properties of matrix-vector multiplication that are used extensively throughout linear algebra.

Theorem 2.2.2

Let A and B be $m \times n$ matrices, and let x and y be n-vectors in ^{n}. Then:

1. $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$.
2. $A(a\mathbf{x}) = a(A\mathbf{x}) = (aA)\mathbf{x}$ for all scalars a.
3. $(A + B)\mathbf{x} = A\mathbf{x} + B\mathbf{x}$.

Proof:

We prove (3); the other verifications are similar and are left as exercises. Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ and $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$ be given in terms of their columns. Since adding two matrices is the same as adding their columns, we have

$$A + B = \begin{bmatrix} \mathbf{a}_1 + \mathbf{b}_1 & \mathbf{a}_2 + \mathbf{b}_2 & \cdots & \mathbf{a}_n + \mathbf{b}_n \end{bmatrix}$$

If we write $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$.

Definition 2.5 gives

$$\begin{aligned} (A + B)\mathbf{x} &= x_1(\mathbf{a}_1 + \mathbf{b}_1) + x_2(\mathbf{a}_2 + \mathbf{b}_2) + \cdots + x_n(\mathbf{a}_n + \mathbf{b}_n) \\ &= (x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n) + (x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n) \\ &= A\mathbf{x} + B\mathbf{x} \end{aligned}$$

Theorem 2.2.2 allows matrix-vector computations to be carried

out much as in ordinary arithmetic. For example, for any $m \times n$ matrices A and B and any n-vectors $\mathbf{x}$ and $\mathbf{y}$, we have:

$$A(2\mathbf{x} - 5\mathbf{y}) = 2A\mathbf{x} - 5A\mathbf{y} \quad \text{and} \quad (3A - 7B)\mathbf{x} = 3A\mathbf{x} - 7B\mathbf{x}$$

We will use such manipulations throughout the book, often without mention.

Linear Equations

Theorem 2.2.2 also gives a useful way to describe the solutions to a system
$$A\mathbf{x} = \mathbf{b}$$
of linear equations. There is a related system
$$A\mathbf{x} = \mathbf{0}$$
called the **associated homogeneous system**, obtained from the original system $A\mathbf{x} = \mathbf{b}$ by replacing all the constants by zeros. Suppose $\mathbf{x}_1$ is a solution to $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x}_0$ is a solution to $A\mathbf{x} = \mathbf{0}$ (that is $A\mathbf{x}_1 = \mathbf{b}$ and $A\mathbf{x}_0 = \mathbf{0}$). Then $\mathbf{x}_1 + \mathbf{x}_0$ is another solution to $A\mathbf{x} = \mathbf{b}$. Indeed, Theorem 2.2.2 gives
$$A(\mathbf{x}_1 + \mathbf{x}_0) = A\mathbf{x}_1 + A\mathbf{x}_0 = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

This observation has a useful converse.

Theorem 2.2.3

Suppose $\mathbf{x}_1$ is any particular solution to the system $A\mathbf{x} = \mathbf{b}$ of linear equations. Then every solution $\mathbf{x}_2$ to $A\mathbf{x} = \mathbf{b}$ has the form
$$\mathbf{x}_2 = \mathbf{x}_0 + \mathbf{x}_1$$

for some solution $\mathbf{x}_0$ of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$.

Proof:

Suppose $\mathbf{x}_2$ is also a solution to $A\mathbf{x} = \mathbf{b}$, so that $A\mathbf{x}_2 = \mathbf{b}$. Write $\mathbf{x}_0 = \mathbf{x}_2 - \mathbf{x}_1$. Then $\mathbf{x}_2 = \mathbf{x}_0 + \mathbf{x}_1$ and, using Theorem 2.2.2, we compute

$$A\mathbf{x}_0 = A(\mathbf{x}_2 - \mathbf{x}_1) = A\mathbf{x}_2 - A\mathbf{x}_1 = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

Hence $\mathbf{x}_0$ is a solution to the associated homogeneous system $A\mathbf{x} = \mathbf{0}$.

Note that gaussian elimination provides one such representation.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=66#h5p-145

Example 2.2.7

Express every solution to the following system as the sum of a specific solution plus a solution to the associated homogeneous system.

$$\begin{array}{rrrrl}
x_1 - x_2 - & x_3 + 3x_4 &= 2 \\
2x_1 - x_2 - & 3x_3 + 4x_4 &= 6 \\
x_1 - & 2x_3 + x_4 &= 4
\end{array}$$

Solution:

Gaussian elimination gives $x_1 = 4 + 2s - t$, $x_2 = 2 + s + 2t$, $x_3 = s$, and $x_4 = t$ where s and t are arbitrary parameters. Hence the general solution can be written

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 + 2s - t \\ 2 + s + 2t \\ s \\ t \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 0 \end{bmatrix} + \left(s \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right)$$

Thus

$$\mathbf{x}_1 = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 0 \end{bmatrix}$$

is a particular solution (where $s = 0 = t$), and

$$\mathbf{x}_0 = s \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$ gives **all** solutions to the

associated homogeneous system. (To see why this is so, carry out

the gaussian elimination again but with all the constants set equal to zero.)

The following useful result is included with no proof.

Theorem 2.2.4

Let $A\mathbf{x} = \mathbf{b}$ be a system of equations with augmented matrix $\begin{bmatrix} A & | & \mathbf{b} \end{bmatrix}$. Write $\operatorname{rank} A = r$.

1. $\operatorname{rank} \begin{bmatrix} A & | & \mathbf{b} \end{bmatrix}$ is either r or $r + 1$.
2. The system is consistent if and only if $\operatorname{rank} \begin{bmatrix} A & | & \mathbf{b} \end{bmatrix} = r$.
3. The system is inconsistent if and only if $\operatorname{rank} \begin{bmatrix} A & | & \mathbf{b} \end{bmatrix} = r + 1$.

The Dot Product

Definition 2.5 is not always the easiest way to compute a matrix-vector product $A\mathbf{x}$ because it requires that the columns of A be

explicitly identified. There is another way to find such a product which uses the matrix A as a whole with no reference to its columns, and hence is useful in practice. The method depends on the following notion.

Definition 2.6 Dot Product in $\mathbb{R}^n$

If $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ are two ordered n-tuples, their **dot product** is defined to be the number

$$a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

obtained by multiplying corresponding entries and adding the results.

To see how this relates to matrix products, let A denote a 3×4 matrix and let $\mathbf{x}$ be a 4-vector. Writing

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$$

in the notation of Section 2.1, we compute

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} + x_3 \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix} + x_4 \begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 \end{bmatrix}$$

From this we see that each entry of $A\mathbf{x}$ is the dot product of the corresponding row of A with $\mathbf{x}$. This computation goes through in general, and we record the result in Theorem 2.2.5.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=66#h5p-42

Theorem 2.2.5 Dot Product Rule

> Let A be an $m \times n$ matrix and let $\mathbf{x}$ be an n-vector. Then each entry of the vector $A\mathbf{x}$ is the dot product of the corresponding row of A with $\mathbf{x}$.

This result is used extensively throughout linear algebra.

If A is $m \times n$ and $\mathbf{x}$ is an n-vector, the computation of $A\mathbf{x}$ by the dot product rule is simpler than using Definition 2.5 because the computation can be carried out directly with no explicit reference to the columns of A (as in Definition 2.5. The first entry of $A\mathbf{x}$ is the dot product of row 1 of A with $\mathbf{x}$. In hand calculations this is computed by going *across* row one of A, going *down* the column $\mathbf{x}$, multiplying corresponding entries, and adding the results. The other entries of $A\mathbf{x}$ are computed in the same way using the other rows of A with the column $\mathbf{x}$.
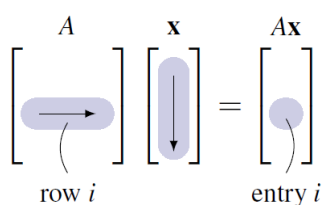
In general, compute entry i of $A\mathbf{x}$ as follows (see the diagram):

Go *across* row i of A and *down* column $\mathbf{x}$, multiply corresponding entries, and add the results.

As an illustration, we rework Example 2.2.2 using the dot product rule instead of Definition 2.5.

Example 2.2.8

$$\text{If } A = \begin{bmatrix} 2 & -1 & 3 & 5 \\ 0 & 2 & -3 & 1 \\ -3 & 4 & 1 & 2 \end{bmatrix}$$

$$\text{and } \mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ -2 \end{bmatrix}, \text{ compute } A\mathbf{x}.$$

Solution:

The entries of $A\mathbf{x}$ are the dot products of the rows of A with $\mathbf{x}$:

$$A\mathbf{x} = \begin{bmatrix} 2 & -1 & 3 & 5 \\ 0 & 2 & -3 & 1 \\ -3 & 4 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} 2\cdot2 & + & (-1)1 & + & 3\cdot0 & + & 5(-2) \\ 0\cdot2 & + & 2\cdot1 & + & (-3)0 & + & 1(-2) \\ (-3)2 & + & 4\cdot1 & + & 1\cdot0 & + & 2(-2) \end{bmatrix} = \begin{bmatrix} -7 \\ 0 \\ -6 \end{bmatrix}$$

Of course, this agrees with the outcome in Example 2.2.2.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=66#h5p-46

Example 2.2.9

Write the following system of linear equations in the form $A\mathbf{x} = \mathbf{b}$.

$$
\begin{aligned}
5x_1 - x_2 + 2x_3 + x_4 - 3x_5 &= 8 \\
x_1 + x_2 + 3x_3 - 5x_4 + 2x_5 &= -2 \\
-x_1 + x_2 - 2x_3 + - 3x_5 &= 0
\end{aligned}
$$

Solution:

Write
$$
A = \begin{bmatrix} 5 & -1 & 2 & 1 & -3 \\ 1 & 1 & 3 & -5 & 2 \\ -1 & 1 & -2 & 0 & -3 \end{bmatrix},
$$

$\mathbf{b} = \begin{bmatrix} 8 \\ -2 \\ 0 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$. Then the dot product rule

gives $A\mathbf{x} = \begin{bmatrix} 5x_1 - x_2 + 2x_3 + x_4 - 3x_5 \\ x_1 + x_2 + 3x_3 - 5x_4 + 2x_5 \\ -x_1 + x_2 - 2x_3 - 3x_5 \end{bmatrix}$, so the

entries of $A\mathbf{x}$ are the left sides of the equations in the linear system. Hence the system becomes $A\mathbf{x} = \mathbf{b}$ because matrices are equal if and only corresponding entries are equal.

An interactive H5P element has been excluded from this

Example 2.2.10

If A is the zero $m \times n$ matrix, then $A\mathbf{x} = \mathbf{0}$ for each n-vector $\mathbf{x}$.

Solution:

For each k, entry k of $A\mathbf{x}$ is the dot product of row k of A with $\mathbf{x}$, and this is zero because row k of A consists of zeros.

Definition 2.7 The Identity Matrix

For each $n > 2$, the **identity matrix** I_n is the $n \times n$ matrix with 1s on the main diagonal (upper left to lower right), and zeros elsewhere.

The first few identity matrices are

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \ldots$$

In Example 2.2.6 we showed that $I_3\mathbf{x} = \mathbf{x}$ for each 3-vector $\mathbf{x}$ using Definition 2.5. The following result shows that this holds in general, and is the reason for the name.

Example 2.2.11

For each $n \geq 2$ we have $I_n\mathbf{x} = \mathbf{x}$ for each n-vector $\mathbf{x}$ in $\mathbf{R}^n$.

Solution:

We verify the case $n = 4$. Given the 4-vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$

the dot product rule gives

$$I_4\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 + 0 + 0 + 0 \\ 0 + x_2 + 0 + 0 \\ 0 + 0 + x_3 + 0 \\ 0 + 0 + 0 + x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x$$

In general, $I_n\mathbf{x} = \mathbf{x}$ because entry k of $I_n\mathbf{x}$ is the dot product of row k of I_n with $\mathbf{x}$, and row k of I_n has 1 in position k and zeros elsewhere.

Example 2.2.12

Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ be any $m \times n$ matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. If $\mathbf{e}_j$ denotes column j of the $n \times n$ identity matrix I_n, then $A\mathbf{e}_j = \mathbf{a}_j$ for each $j = 1, 2, \ldots, n$.

Solution:

Write $\mathbf{e}_j = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}$

where $t_j = 1$, but $t_i = 0$ for all $i \neq j$. Then Theorem 2.2.5 gives

$$A\mathbf{e}_j = t_1\mathbf{a}_1 + \cdots + t_j\mathbf{a}_j + \cdots + t_n\mathbf{a}_n = 0 + \cdots + \mathbf{a}_j + \cdots + 0 = \mathbf{a}_j$$

Example 2.2.12 will be referred to later; for now we use it to prove:

> **Theorem 2.2.6**
>
> Let A and B be $m \times n$ matrices. If $A\mathbf{x} = B\mathbf{x}$ for all $\mathbf{x}$ in $\mathbf{R}^n$, then $A = B$.

Proof:

Write $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ and $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$ and in terms of their columns. It is enough to show that $\mathbf{a}_k = \mathbf{b}_k$ holds for all k. But we are assuming that $A\mathbf{e}_k = B\mathbf{e}_k$, which gives $\mathbf{a}_k = \mathbf{b}_k$ by Example 2.2.12.

We have introduced matrix-vector multiplication as a new way to think about systems of linear equations. But it has several other uses as well. It turns out that many geometric operations can be described using matrix multiplication, and we now investigate how this happens. As a bonus, this description provides a geometric "picture" of a matrix by revealing the effect on a vector when it is multiplied by A. This "geometric view" of matrices is a fundamental tool in understanding them.

2.3 Matrix Multiplication

In Section 2.2 matrix-vector products were introduced. If A is an $m \times n$ matrix, the product $A\mathbf{x}$ was defined for any n-column x in $\mathbf{R}^n$ as follows: If $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ where the

$\mathbf{a}_j$ are the columns of A, and if $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$,

Definition 2.5 reads

(2.5) $\quad A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$

This was motivated as a way of describing systems of linear equations with coefficient matrix A. Indeed every such system has the form $A\mathbf{x} = \mathbf{b}$ where $\mathbf{b}$ is the column of constants.

In this section we extend this matrix-vector multiplication to a way of multiplying matrices in general, and then investigate matrix algebra for its own sake. While it shares several properties of ordinary arithmetic, it will soon become clear that matrix arithmetic is different in a number of ways.

<div style="background:#ccc;padding:1em">

Definition 2.9 Matrix Multiplication

Let A be an $m \times n$ matrix, let B be an $n \times k$ matrix, and write $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_k \end{bmatrix}$ where $\mathbf{b}_j$ is column j of B for each j. The product matrix AB is the $m \times k$ matrix defined as follows:

$$AB = A \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_k \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_k \end{bmatrix}$$

</div>

Thus the product matrix AB is given in terms of its columns $A\mathbf{b}_1, A\mathbf{b}_2, \ldots, A\mathbf{b}_n$: Column j of AB is the matrix-vector product $A\mathbf{b}_j$ of A and the corresponding column $\mathbf{b}_j$ of B. Note that each such product $A\mathbf{b}_j$ makes sense by Definition 2.5 because A is $m \times n$ and each b_j is in $\mathbf{R}^n$ (since B has n rows). Note also

that if B is a column matrix, this definition reduces to Definition 2.5 for matrix-vector multiplication.

Given matrices A and B, Definition 2.9 and the above computation give

$$A(B\vec{x}) = \begin{bmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_n \end{bmatrix} \vec{x} = (AB)\vec{x}$$

for all $\vec{x}$ in $\mathbf{R}^k$. We record this for reference.

Theorem 2.3.1

Let A be an $m \times n$ matrix and let B be an $n \times k$ matrix. Then the product matrix AB is $m \times k$ and satisfies

$$A(B\vec{x}) = (AB)\vec{x} \quad \text{for all } \vec{x} \text{ in } \mathbf{R}^k$$

Here is an example of how to compute the product AB of two matrices using Definition 2.9.

Example 2.3.1

Compute AB if $A = \begin{bmatrix} 2 & 3 & 5 \\ 1 & 4 & 7 \\ 0 & 1 & 8 \end{bmatrix}$

and

$$B = \begin{bmatrix} 8 & 9 \\ 7 & 2 \\ 6 & 1 \end{bmatrix}.$$

Solution:

The columns of B are

$$\vec{b}_1 = \begin{bmatrix} 8 \\ 7 \\ 6 \end{bmatrix} \text{ and } \vec{b}_2 = \begin{bmatrix} 9 \\ 2 \\ 1 \end{bmatrix}, \text{ so Definition 2.5 gives}$$

$$A\vec{b}_1 = \begin{bmatrix} 2 & 3 & 5 \\ 1 & 4 & 7 \\ 0 & 1 & 8 \end{bmatrix} \begin{bmatrix} 8 \\ 7 \\ 6 \end{bmatrix} = \begin{bmatrix} 67 \\ 78 \\ 55 \end{bmatrix} \text{ and } A\vec{b}_2 = \begin{bmatrix} 2 & 3 & 5 \\ 1 & 4 & 7 \\ 0 & 1 & 8 \end{bmatrix} \begin{bmatrix} 9 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 29 \\ 24 \\ 10 \end{bmatrix}$$

Hence Definition 2.9 above gives

$$AB = \begin{bmatrix} A\vec{b}_1 & A\vec{b}_2 \end{bmatrix} = \begin{bmatrix} 67 & 29 \\ 78 & 24 \\ 55 & 10 \end{bmatrix}.$$

While Definition 2.9 is important, there is another way to compute the matrix product AB that gives a way to calculate each individual entry. In Section 2.2 we defined the dot product of two n-tuples to be the sum of the products of corresponding entries. We went on to show (Theorem 2.2.5) that if A is an $m \times n$ matrix and $\vec{x}$ is an n-vector, then entry j of the product $A\vec{x}$ is the dot product of row j of A with $\vec{x}$. This observation was called the "dot product rule" for matrix-vector multiplication, and the next theorem shows that it extends to matrix multiplication in general.

Let A and B be matrices of sizes $m \times n$ and $n \times k$, respectively. Then the (i, j)-entry of AB is the dot product of row i of A with column j of B.

Proof:

Write $B = \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \end{bmatrix}$ in terms of its columns. Then $A\vec{b}_j$ is column j of AB for each j. Hence the (i, j)-entry of AB is entry i of $A\vec{b}_j$, which is the dot product of row i of A with $\vec{b}_j$. This proves the theorem.

Thus to compute the (i, j)-entry of AB, proceed as follows (see the diagram):

Go *across* row i of A, and *down* column j of B, multiply corresponding entries, and add the results.
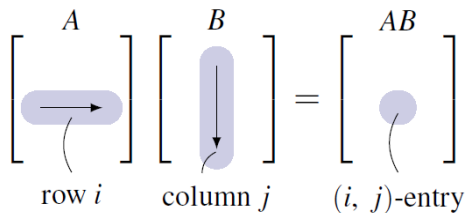
Note that this requires that the rows of A must be the same length as the columns of B. The following rule is useful for remembering this and for deciding the size of the product matrix AB.

Compatibility Rule

Let A and B denote matrices. If A is $m \times n$ and B is $n' \times k$, the product AB can be formed if and only if $n = n'$. In this case the size of the product matrix AB is $m \times k$, and we say that AB is **defined,** or that A and B are **compatible** for multiplication.

$$A \qquad B$$

$$m \times \left(\!\!\begin{array}{cc} n & n' \end{array}\!\!\right) \times k$$

The diagram provides a useful mnemonic for remembering this. We adopt the following convention:

Whenever a product of matrices is written, it is tacitly assumed that the sizes of the factors are such that the product is defined.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=66#h5p-49

To illustrate the dot product rule, we recompute the matrix product in Example 2.3.1.

Example 2.3.3

Compute AB if $A = \begin{bmatrix} 2 & 3 & 5 \\ 1 & 4 & 7 \\ 0 & 1 & 8 \end{bmatrix}$

and $B = \begin{bmatrix} 8 & 9 \\ 7 & 2 \\ 6 & 1 \end{bmatrix}$.

Solution:

Here A is 3×3 and B is 3×2, so the product matrix AB is defined and will be of size 3×2. Theorem 2.3.2 gives each entry of AB as the dot product of the corresponding row of A with the corresponding column of B_j that is,

$$AB = \begin{bmatrix} 2 & 3 & 5 \\ 1 & 4 & 7 \\ 0 & 1 & 8 \end{bmatrix} \begin{bmatrix} 8 & 9 \\ 7 & 2 \\ 6 & 1 \end{bmatrix} = \begin{bmatrix} 2 \cdot 8 + 3 \cdot 7 + 5 \cdot 6 & 2 \cdot 9 + 3 \cdot 2 + 5 \cdot 1 \\ 1 \cdot 8 + 4 \cdot 7 + 7 \cdot 6 & 1 \cdot 9 + 4 \cdot 2 + 7 \cdot 1 \\ 0 \cdot 8 + 1 \cdot 7 + 8 \cdot 6 & 0 \cdot 9 + 1 \cdot 2 + 8 \cdot 1 \end{bmatrix} = \begin{bmatrix} 67 & 29 \\ 78 & 24 \\ 55 & 10 \end{bmatrix}$$

Of course, this agrees with Example 2.3.1.

Example 2.3.4

Compute the $(1, 3)$- and $(2, 4)$-entries of AB where

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 0 & 1 & 4 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 1 & 6 & 0 \\ 0 & 2 & 3 & 4 \\ -1 & 0 & 5 & 8 \end{bmatrix}.$$

Then compute AB.

Solution:

The $(1, 3)$-entry of AB is the dot product of row 1 of A and column 3 of B (highlighted in the following display), computed by multiplying corresponding entries and adding the results.

$$\begin{bmatrix} 3 & -1 & 2 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 & 6 & 0 \\ 0 & 2 & 3 & 4 \\ -1 & 0 & 5 & 8 \end{bmatrix} \qquad (1, 3)\text{-entry} = 3 \cdot 6 + (-1) \cdot 3 + 2 \cdot 5 = 25$$

Similarly, the $(2, 4)$-entry of AB involves row 2 of A and column 4 of B.

$$\begin{bmatrix} 3 & -1 & 2 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 & 6 & 0 \\ 0 & 2 & 3 & 4 \\ -1 & 0 & 5 & 8 \end{bmatrix} \qquad (2, 4)\text{-entry} = 0 \cdot 0 + 1 \cdot 4 + 4 \cdot 8 = 36$$

Since A is 2×3 and B is 3×4, the product is 2×4.

$$AB = \begin{bmatrix} 3 & -1 & 2 \\ 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 & 6 & 0 \\ 0 & 2 & 3 & 4 \\ -1 & 0 & 5 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 1 & 25 & 12 \\ -4 & 2 & 23 & 36 \end{bmatrix}$$

An interactive H5P element has been excluded

Example 2.3.5

If $A = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 5 \\ 6 \\ 4 \end{bmatrix}$, compute A^2, AB, BA, and B^2 when they are defined.

Solution:

Here, A is a 1×3 matrix and B is a 3×1 matrix, so A^2 and B^2 are not defined. However, the compatibility rule reads

$$\begin{array}{cc} A & B \\ 1 \times 3 & 3 \times 1 \end{array} \quad \text{and} \quad \begin{array}{cc} B & A \\ 3 \times 1 & 1 \times 3 \end{array}$$

so both AB and BA can be formed and these are 1×1 and 3×3 matrices, respectively.

$$AB = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 3 \cdot 6 + 2 \cdot 4 \end{bmatrix} = \begin{bmatrix} 31 \end{bmatrix}$$

$$BA = \begin{bmatrix} 5 \\ 6 \\ 4 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 5 \cdot 1 & 5 \cdot 3 & 5 \cdot 2 \\ 6 \cdot 1 & 6 \cdot 3 & 6 \cdot 2 \\ 4 \cdot 1 & 4 \cdot 3 & 4 \cdot 2 \end{bmatrix} = \begin{bmatrix} 5 & 15 & 10 \\ 6 & 18 & 12 \\ 4 & 12 & 8 \end{bmatrix}$$

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=66#h5p-50

Unlike numerical multiplication, matrix products AB and BA *need not be equal.* In fact they need not even be the same size, as Example 2.3.5 shows. It turns out to be rare that $AB = BA$ (although it is by no means impossible), and A and B are said to **commute** when this happens.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=66#h5p-48

Example 2.3.6

Let $A = \begin{bmatrix} 6 & 9 \\ -4 & -6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}$.

Compute A^2, AB, BA.

Solution:

$$A^2 = \begin{bmatrix} 6 & 9 \\ -4 & -6 \end{bmatrix} \begin{bmatrix} 6 & 9 \\ -4 & -6 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \text{so}$$

$A^2 = 0$ can occur even if $A \neq 0$. Next,

$$AB = \begin{bmatrix} 6 & 9 \\ -4 & -6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -3 & 12 \\ 2 & -8 \end{bmatrix}$$

$$BA = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 6 & 9 \\ -4 & -6 \end{bmatrix} = \begin{bmatrix} -2 & -3 \\ -6 & -9 \end{bmatrix}$$

Hence $AB \neq BA$, even though AB and BA are the same size.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-27

Example 2.3.7

If A is any matrix, then $IA = A$ and $AI = A$, and where I denotes an identity matrix of a size so that the multiplications are defined.

Solution:

These both follow from the dot product rule as the reader should verify. For a more formal proof, write $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix}$ where $\vec{a}_j$ is column j of A. Then Definition 2.9 and Example 2.2.1 give

$$IA = \begin{bmatrix} I\vec{a}_1 & I\vec{a}_2 & \cdots & I\vec{a}_n \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = A$$

If $\vec{e}_j$ denotes column j of I, then $A\vec{e}_j = \vec{a}_j$ for each j by Example 2.2.12. Hence Definition 2.9 gives:

$$AI = A \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \end{bmatrix} = \begin{bmatrix} A\vec{e}_1 & A\vec{e}_2 & \cdots & A\vec{e}_n \end{bmatrix} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix} = A$$

The following theorem collects several results about matrix multiplication that are used everywhere in linear algebra.

Assume that a is any scalar, and that A, B, and C are matrices of sizes such that the indicated matrix products are defined. Then:

1. $IA = A$ and $AI = A$ where I denotes an identity matrix.

2. $A(BC) = (AB)C.$

3. $A(B + C) = AB + AC.$

4. $(B + C)A = BA + CA.$

5. $a(AB) = (aA)B = A(aB).$

6. $(AB)^T = B^T A^T.$

Proof:

Condition (1) is Example 2.3.7; we prove (2), (4), and (6) and leave (3) and (5) as exercises.

1. If $C = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_k \end{bmatrix}$ in terms of its columns, then $BC = \begin{bmatrix} B\vec{c}_1 & B\vec{c}_2 & \cdots & B\vec{c}_k \end{bmatrix}$ by Definition 2.9, so

$$
\begin{aligned}
A(BC) &= \begin{bmatrix} A(B\vec{c}_1) & A(B\vec{c}_2) & \cdots & A(B\vec{c}_k) \end{bmatrix} & \text{Definition 2.9} \\
&= \begin{bmatrix} (AB)\vec{c}_1 & (AB)\vec{c}_2 & \cdots & (AB)\vec{c}_k) \end{bmatrix} & \text{Theorem 2.3.1} \\
&= (AB)C & \text{Definition 2.9}
\end{aligned}
$$

4. We know (Theorem 2.2.) that $(B + C)\vec{x} = B\vec{x} + C\vec{x}$ holds for every column $\vec{x}$. If we write $A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix}$ in terms of its columns, we get

$$(B+C)A \quad = \quad \begin{bmatrix} (B+C)\vec{a}_1 & (B+C)\vec{a}_2 & \cdots & (B+C)\vec{a}_n \end{bmatrix} \qquad \text{Definition 2.9}$$

$$= \quad \begin{bmatrix} B\vec{a}_1 + C\vec{a}_1 & B\vec{a}_2 + C\vec{a}_2 & \cdots & B\vec{a}_n + C\vec{a}_n \end{bmatrix} \qquad \text{Theorem 2.2.2}$$

$$= \quad \begin{bmatrix} B\vec{a}_1 & B\vec{a}_2 & \cdots & B\vec{a}_n \end{bmatrix} + \begin{bmatrix} C\vec{a}_1 & C\vec{a}_2 & \cdots & C\vec{a}_n \end{bmatrix} \qquad \text{Adding Columns}$$

$$= \quad BA + CA \qquad \text{Definition 2.9}$$

6. As in Section 2.1, write $A = [a_{ij}]$ and $B = [b_{ij}]$, so that $A^T = [a'_{ij}]$ and $B^T = [b'_{ij}]$ where $a'_{ij} = a_{ji}$ and $b'_{ji} = b_{ij}$ for all i and j. If c_{ij} denotes the (i, j)-entry of $B^T A^T$, then c_{ij} is the dot product of row i of B^T with column j of A^T. Hence

$$c_{ij} = b'_{i1}a'_{1j} + b'_{i2}a'_{2j} + \cdots + b'_{im}a'_{mj} = b_{1i}a_{j1} + b_{2i}a_{j2} + \cdots + b_{mi}a_{jm}$$
$$= a_{j1}b_{1i} + a_{j2}b_{2i} + \cdots + a_{jm}b_{mi}$$

But this is the dot product of row j of A with column i of B; that is, the (j, i)-entry of AB; that is, the (i, j)-entry of $(AB)^T$. This proves (6).

Property 2 in Theorem 2.3.3 is called the **associative law** of matrix multiplication. It asserts that the equation $A(BC) = (AB)C$ holds for all matrices (if the products are defined). Hence this product is the same no matter how it is formed, and so is written simply as ABC. This extends: The product $ABCD$ of four matrices can be formed several ways—for example, $(AB)(CD)$

, $[A(BC)]D$, and $A[B(CD)]$—but the associative law implies that they are all equal and so are written as $ABCD$. A similar remark applies in general: Matrix products can be written unambiguously with no parentheses.

However, a note of caution about matrix multiplication must be taken: The fact that AB and BA need *not* be equal means that the *order* of the factors is important in a product of matrices. For example $ABCD$ and $ADCB$ may *not* be equal.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=66#h5p-24

Warning:
If the order of the factors in a product of matrices is changed, the product matrix may change (or may not be defined). Ignoring this warning is a source of many errors by students of linear algebra!}

Properties 3 and 4 in Theorem 2.3.3 are called **distributive laws**. They assert that $A(B+C) = AB + AC$ and $(B+C)A = BA + CA$ hold whenever the sums and products are defined. These rules extend to more than two terms and, together with Property 5, ensure that many manipulations familiar from ordinary algebra extend to matrices. For example

$$A(2B - 3C + D - 5E) = 2AB - 3AC + AD - 5AE$$
$$(A + 3C - 2D)B = AB + 3CB - 2DB$$

Note again that the warning is in effect: For example

$A(B - C)$ need *not* equal $AB - CA$. These rules make possible a lot of simplification of matrix expressions.

> *Example 2.3.8*
>
> Simplify the expression
> $$A(BC - CD) + A(C - B)D - AB(C - D)$$
> .

Solution:

$$
\begin{aligned}
A(BC - CD) + A(C - B)D - AB(C - D) &= A(BC) - A(CD) + (AC - AB)D - (AB)C + (AB)D \\
&= ABC - ACD + ACD - ABD - ABC + ABD \\
&= 0
\end{aligned}
$$

> 📰 *An interactive H5P element has been excluded from this version of the text. You can view it online here:*
> *https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-25*

Example 2.3.9 and Example 2.3.10 below show how we can use the properties in Theorem 2.3.2 to deduce other facts about matrix

multiplication. Matrices A and B are said to **commute** if $AB = BA$.

Suppose that A, B, and C are $n \times n$ matrices and that both A and B commute with C; that is, $AC = CA$ and $BC = CB$. Show that AB commutes with C.

Solution:

Showing that AB commutes with C means verifying that $(AB)C = C(AB)$. The computation uses the associative law several times, as well as the given facts that $AC = CA$ and $BC = CB$.

$$(AB)C = A(BC) = A(CB) = (AC)B = (CA)B = C(AB)$$

🖹 *An interactive H5P element has been excluded from this version of the text. You can view it online here:*
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-146

Example 2.3.10

Solution:

The following *always* holds:

(2.6)
$$(A - B)(A + B) = A(A + B) - B(A + B) = A^2 + AB - BA - B^2$$

Hence if $AB = BA$, then $(A - B)(A + B) = A^2 - B^2$ follows. Conversely, if this last equation holds, then equation (2.6 becomes

$$A^2 - B^2 = A^2 + AB - BA - B^2$$

This gives $0 = AB - BA$, and $AB = BA$ follows.

In Section 2.2 we saw (in Theorem 2.2.1) that every system of linear equations has the form

$$A\vec{x} = \vec{b}$$

where A is the coefficient matrix, $\vec{x}$ is the column of variables, and $\vec{b}$ is the constant matrix. Thus the **system** of linear equations becomes a single matrix equation. Matrix multiplication can yield information about such a system.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-26

Example 2.3.11

Consider a system $A\vec{x} = \vec{b}$ of linear equations where A is an $m \times n$ matrix. Assume that a matrix C exists such that $CA = I_n$. If the system $A\vec{x} = \vec{b}$ has a solution, show that this solution must be $C\vec{b}$. Give a condition guaranteeing that $C\vec{b}$ is *in fact* a solution.

Solution:

Suppose that $\vec{x}$ is any solution to the system, so that $A\vec{x} = \vec{b}$. Multiply both sides of this matrix equation by C to obtain, successively,

$$C(A\vec{x}) = C\vec{b}, \quad (CA)\vec{x} = C\vec{b}, \quad I_n\vec{x} = C\vec{b}, \quad \vec{x} = C\vec{b}$$

This shows that *if* the system has a solution $\vec{x}$, then that solution must be $\vec{x} = C\vec{b}$, as required. But it does *not* guarantee that the system has a solution. However, if we write $\vec{x}_1 = C\vec{b}$, then

$$A\vec{x}_1 = A(C\vec{b}) = (AC)\vec{b}$$

Thus $\vec{x}_1 = C\vec{b}$ will be a solution if the condition $AC = I_m$ is satisfied.

The ideas in Example 2.3.11 lead to important information about matrices; this will be pursued in the next section.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=66#h5p-38

2.4 Matrix Inverse

Three basic operations on matrices, addition, multiplication, and subtraction, are analogs for matrices of the same operations for numbers. In this section we introduce the matrix analog of numerical division.

To begin, consider how a numerical equation $ax = b$ is solved when a and b are known numbers. If $a = 0$, there is no solution (unless $b = 0$). But if $a \neq 0$, we can multiply both sides by the inverse $a^{-1} = \frac{1}{a}$ to obtain the solution $x = a^{-1}b$. Of course multiplying by a^{-1} is just dividing by a, and the property of a^{-1} that makes this work is that $a^{-1}a = 1$. Moreover, we saw in Section~?? that the role that 1 plays in arithmetic is played in matrix algebra by the identity matrix I. This suggests the following definition.

Definition 2.11 Matrix Inverses

If A is a square matrix, a matrix B is called an **inverse** of A if and only if
$$AB = I \quad \text{and} \quad BA = I$$
A matrix A that has an inverse is called an **invertible matrix**.

Note that only square matrices have inverses. Even though it is plausible that nonsquare matrices A and B could exist such that $AB = I_m$ and $BA = I_n$, where A is $m \times n$ and B is $n \times m$, we claim that this forces $n = m$. Indeed, if $m < n$ there exists a nonzero column $\vec{x}$ such that $A\vec{x} = \vec{0}$ (by Theorem

1.3.1), so $\vec{x} = I_n\vec{x} = (BA)\vec{x} = B(A\vec{x}) = B(\vec{0}) = \vec{0}$, a contradiction. Hence $m \geq n$. Similarly, the condition $AB = I_m$ implies that $n \geq m$. Hence $m = n$ so A is square.}

Example 2.4.1

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

is an inverse of $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$.

Solution:

Compute AB and BA.

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

Hence $AB = I = BA$, so B is indeed an inverse of A.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-147

Show that $A = \begin{bmatrix} 0 & 0 \\ 1 & 3 \end{bmatrix}$

has no inverse.

Solution:

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

denote an arbitrary 2×2 matrix. Then

$$ AB = \begin{bmatrix} 0 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ a + 3c & b + 3d \end{bmatrix} $$

so AB has a row of zeros. Hence AB cannot equal I for any B.

The argument in Example 2.4.2 shows that no zero matrix has an inverse. But Example 2.4.2 also shows that, unlike arithmetic, *it is possible for a nonzero matrix to have no inverse*. However, if a matrix *does* have an inverse, it has only one.

Theorem 2.4.1

If B and C are both inverses of A, then $B = C$.

Proof:

Since B and C are both inverses of A, we have $CA = I = AB$. Hence
$$B = IB = (CA)B = C(AB) = CI = C$$

If A is an invertible matrix, the (unique) inverse of A is denoted A^{-1}. Hence A^{-1} (when it exists) is a square matrix of the same size as A with the property that
$$AA^{-1} = I \quad \text{and} \quad A^{-1}A = I$$

These equations characterize A^{-1} in the following sense:

Inverse Criterion: If somehow a matrix B can be found such that $AB = I$ and $BA = I$, then A is invertible and B is the inverse of A; in symbols, $B = A^{-1}$.}

This is a way to verify that the inverse of a matrix exists. Example 2.3.3 and Example 2.3.4 offer illustrations.

Example 2.4.3

If $A = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$, show that $A^3 = I$ and so find A^{-1}.

Solution:

We have $A^2 = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}$, and so

$$A^3 = A^2 A = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Hence $A^3 = I$, as asserted. This can be written as $A^2 A = I = AA^2$, so it shows that A^2 is the inverse of A. That is, $A^{-1} = A^2 = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}$.

The next example presents a useful formula for the inverse of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ when it exists. To state it, we define the **determinant** $det\,A$ and the **adjugate** $adj\,A$ of the matrix A as follows:

$$det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc, \quad \text{and} \quad adj \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Example 2.4.4

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, show that A has an inverse if and only if $det\,A \neq 0$, and in this case

$$A^{-1} = \frac{1}{det\,A} adj\,A$$

Solution:

For convenience, write $e = det\,A = ad - bc$ and $B = adj\,A = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. Then $AB = eI = BA$ as the reader can verify. So if $e \neq 0$, scalar multiplication by $\frac{1}{e}$ gives

$$A(\frac{1}{e}B) = I = (\frac{1}{e}B)A$$

Hence A is invertible and $A^{-1} = \frac{1}{e}B$. Thus it remains only to show that if A^{-1} exists, then $e \neq 0$.

We prove this by showing that assuming $e = 0$ leads to a contradiction. In fact, if $e = 0$, then $AB = eI = 0$, so left multiplication by A^{-1} gives $A^{-1}AB = A^{-1}0$; that is, $IB = 0$, so $B = 0$. But this implies that a, b, c, and d are *all* zero, so $A = 0$, contrary to the assumption that A^{-1} exists.

As an illustration, if $A = \begin{bmatrix} 2 & 4 \\ -3 & 8 \end{bmatrix}$ then $det A = 2 \cdot 8 - 4 \cdot (-3) = 28 \neq 0$. Hence A is invertible and $A^{-1} = \frac{1}{det A} adj A = \frac{1}{28} \begin{bmatrix} 8 & -4 \\ 3 & 2 \end{bmatrix}$, as the reader is invited to verify.

The determinant and adjugate will be defined in Chapter 3 for any square matrix, and the conclusions in Example 2.4.4 will be proved in full generality.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=66#h5p-51

Inverse and Linear systems

Matrix inverses can be used to solve certain systems of linear equations. Recall that a $system$ of linear equations can be written as a $single$ matrix equation

$$A\vec{x} = \vec{b}$$

where A and $\vec{b}$ are known and $\vec{x}$ is to be determined. If A is invertible, we multiply each side of the equation on the left by A^{-1} to get

$$A^{-1}A\vec{x} = A^{-1}\vec{b}$$
$$I\vec{x} = A^{-1}\vec{b}$$
$$\vec{x} = A^{-1}\vec{b}$$

This gives the solution to the system of equations (the reader should verify that $\vec{x} = A^{-1}\vec{b}$ really does satisfy $A\vec{x} = \vec{b}$). Furthermore, the argument shows that if $\vec{x}$ is any solution, then necessarily $\vec{x} = A^{-1}\vec{b}$, so the solution is unique. Of course the technique works only when the coefficient matrix A has an inverse. This proves Theorem 2.4.2.

Theorem 2.4.2

Suppose a system of n equations in n variables is written in matrix form as

$$A\vec{x} = \vec{b}$$

If the $n \times n$ coefficient matrix A is invertible, the system has the unique solution

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

Example 2.4.5

Use Example 2.4.4 to solve the system
$$\begin{cases} 5x_1 - 3x_2 = -4 \\ 7x_1 + 4x_2 = 8 \end{cases}.$$

Solution:

In matrix form this is $A\vec{x} = \vec{b}$ where $A = \begin{bmatrix} 5 & -3 \\ 7 & 4 \end{bmatrix}$, $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and $\vec{b} = \begin{bmatrix} -4 \\ 8 \end{bmatrix}$. Then $\det A = 5 \cdot 4 - (-3) \cdot 7 = 41$, so A is invertible and $A^{-1} = \frac{1}{41} \begin{bmatrix} 4 & 3 \\ -7 & 5 \end{bmatrix}$ by Example 2.4.4. Thus Theorem 2.4.2 gives

$$\vec{x} = A^{-1}\vec{b} = \frac{1}{41} \begin{bmatrix} 4 & 3 \\ -7 & 5 \end{bmatrix} \begin{bmatrix} -4 \\ 8 \end{bmatrix} = \frac{1}{41} \begin{bmatrix} 8 \\ 68 \end{bmatrix}$$

so the solution is $x_1 = \frac{8}{41}$ and $x_2 = \frac{68}{41}$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-19

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-20

An inversion method

If a matrix A is $n \times n$ and invertible, it is desirable to have an efficient technique for finding the inverse. The following procedure will be justified in Section 2.5.

Matrix Inversion Algorithm

If A is an invertible (square) matrix, there exists a sequence of elementary row operations that carry A to the identity matrix I of the same size, written $A \to I$. This same series of row operations carries I to A^{-1}; that is, $I \to A^{-1}$. The algorithm can be summarized as follows:

$$\begin{bmatrix} A & I \end{bmatrix} \rightarrow \begin{bmatrix} I & A^{-1} \end{bmatrix}$$

where the row operations on A and I are carried out simultaneously.

Example 2.4.6

Use the inversion algorithm to find the inverse of the matrix

$$A = \begin{bmatrix} 2 & 7 & 1 \\ 1 & 4 & -1 \\ 1 & 3 & 0 \end{bmatrix}$$

Solution:

Apply elementary row operations to the double matrix

$$\begin{bmatrix} A & I \end{bmatrix} = \left[\begin{array}{ccc|ccc} 2 & 7 & 1 & 1 & 0 & 0 \\ 1 & 4 & -1 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 & 0 & 1 \end{array} \right]$$

so as to carry A to I. First interchange rows 1 and 2.

$$\left[\begin{array}{rrr|rrr} 1 & 4 & -1 & 0 & 1 & 0 \\ 2 & 7 & 1 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 0 & 1 \end{array}\right]$$

Next subtract 2 times row 1 from row 2, and subtract row 1 from row 3.

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

Continue to reduced row-echelon form.

$$\left[\begin{array}{rrr|rrr} 1 & 0 & 11 & 4 & -7 & 0 \\ 0 & 1 & -3 & -1 & 2 & 0 \\ 0 & 0 & -2 & -1 & 1 & 1 \end{array}\right]$$

$$\left[\begin{array}{rrr|rrr} 1 & 0 & 0 & \frac{-3}{2} & \frac{-3}{2} & \frac{11}{2} \\ 0 & 1 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{-3}{2} \\ 0 & 0 & 1 & \frac{1}{2} & \frac{-1}{2} & \frac{-1}{2} \end{array}\right]$$

Hence $A^{-1} = \frac{1}{2}\left[\begin{array}{rrr} -3 & -3 & 11 \\ 1 & 1 & -3 \\ 1 & -1 & -1 \end{array}\right]$, as is readily verified.

Given any $n \times n$ matrix A, Theorem 1.2.1 shows that A can be carried by elementary row operations to a matrix R in reduced row-echelon form. If $R = I$, the matrix A is invertible (this will be proved in the next section), so the algorithm produces A^{-1}. If $R \neq I$, then R has a row of zeros (it is square), so no system of linear equations $A = b$ can have a unique solution. But then A is

not invertible by Theorem 2.4.2. Hence, the algorithm is effective in the sense conveyed in Theorem 2.4.3.

> **Theorem 2.4.3**
>
> If A is an $n \times n$ matrix, either A can be reduced to I by elementary row operations or it cannot. In the first case, the algorithm produces A^{-1}; in the second case, A^{-1} does not exist.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-21

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-53

Properties of inverses

The following properties of an invertible matrix are used everywhere.

Example 2.4.7: Cancellation Laws

Let A be an invertible matrix. Show that:

1. If $AB = AC$, then $B = C$.
2. If $BA = CA$, then $B = C$.

Solution:

Given the equation $AB = AC$, left multiply both sides by A^{-1} to obtain $A^{-1}AB = A^{-1}AC$. Thus $IB = IC$, that is $B = C$. This proves (1) and the proof of (2) is left to the reader.

Properties (1) and (2) in Example 2.4.7 are described by saying that an invertible matrix can be "left cancelled" and "right cancelled", respectively. Note however that "mixed" cancellation does not hold in general: If A is invertible and $AB = CA$, then B and C may *not* be equal, even if both are 2×2. Here is a specific example:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 2 \end{bmatrix}, C = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Sometimes the inverse of a matrix is given by a formula. Example 2.4.4 is one illustration; Example 2.4.8 and Example 2.4.9 provide two more. The idea is the $Inverse\ Criterion$: If a matrix B can be found such that $AB = I = BA$, then A is invertible and $A^{-1} = B$.

Example 2.4.8

If A is an invertible matrix, show that the transpose A^T is also invertible. Show further that the inverse of A^T is just the transpose of A^{-1}; in symbols,
$$(A^T)^{-1} = (A^{-1})^T.$$

Solution:

A^{-1} exists (by assumption). Its transpose $(A^{-1})^T$ is the candidate proposed for the inverse of A^T. Using the inverse criterion, we test it as follows:
$$A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$$
$$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$$

Hence $(A^{-1})^T$ is indeed the inverse of A^T; that is, $(A^T)^{-1} = (A^{-1})^T.$

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-22

Example 2.4.9

If A and B are invertible $n \times n$ matrices, show that their product AB is also invertible and
$(AB)^{-1} = B^{-1}A^{-1}$.

Solution:

We are given a candidate for the inverse of AB, namely $B^{-1}A^{-1}$. We test it as follows:

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I$$
$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

Hence $B^{-1}A^{-1}$ is the inverse of AB; in symbols, $(AB)^{-1} = B^{-1}A^{-1}$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-52

We now collect several basic properties of matrix inverses for reference.

All the following matrices are square matrices of the same size.

1. I is invertible and $I^{-1} = I$.

2. If A is invertible, so is A^{-1}, and $(A^{-1})^{-1} = A$.

3. If A and B are invertible, so is AB, and $(AB)^{-1} = B^{-1}A^{-1}$.

4. If $A_1, A_2, \ldots, A_k$ are all invertible, so is their product $A_1 A_2 \cdots A_k$, and
$$(A_1 A_2 \cdots A_k)^{-1} = A_k^{-1} \cdots A_2^{-1} A_1^{-1}.$$

5. If A is invertible, so is A^k for any $k \geq 1$, and $(A^k)^{-1} = (A^{-1})^k$.

6. If A is invertible and $a \neq 0$ is a number, then aA is invertible and $(aA)^{-1} = \frac{1}{a}A^{-1}$.

7. If A is invertible, so is its transpose A^T, and $(A^T)^{-1} = (A^{-1})^T$.

Proof:

1. This is an immediate consequence of the fact that $I^2 = I$.

2. The equations $AA^{-1} = I = A^{-1}A$ show that A is the inverse of A^{-1}; in symbols, $(A^{-1})^{-1} = A$.

3. This is Example 2.4.9.

4. Use induction on k. If $k = 1$, there is nothing to prove, and if $k = 2$, the result is property 3. If $k > 2$, assume inductively that

$(A_1 A_2 \cdots A_{k-1})^{-1} = A_{k-1}^{-1} \cdots A_2^{-1} A_1^{-1}$. We apply this fact together with property 3 as follows:

$$[A_1 A_2 \cdots A_{k-1} A_k]^{-1} = [(A_1 A_2 \cdots A_{k-1}) A_k]^{-1}$$
$$= A_k^{-1} (A_1 A_2 \cdots A_{k-1})^{-1}$$
$$= A_k^{-1} \left(A_{k-1}^{-1} \cdots A_2^{-1} A_1^{-1} \right)$$

So the proof by induction is complete.

5. This is property 4 with $A_1 = A_2 = \cdots = A_k = A$.

6. The readers are invited to verify it.

7. This is Example 2.4.8.

The reversal of the order of the inverses in properties 3 and 4 of Theorem 2.4.4 is a consequence of the fact that matrix multiplication is not

commutative. Another manifestation of this comes when matrix equations are dealt with. If a matrix equation $B = C$ is given, it can be *left-multiplied* by a matrix A to yield $AB = AC$. Similarly, *right-multiplication* gives $BA = CA$. However, we cannot mix the two: If $B = C$, it need *not* be the case that $AB = CA$ even if A is invertible, for example, $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = C.$

Part 7 of Theorem 2.4.4 together with the fact that $(A^T)^T = A$ gives

Corollary 2.4.1

A square matrix A is invertible if and only if A^T is invertible.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-23

Example 2.4.10

Find A if $(A^T - 2I)^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$.

Solution:

By Theorem 2.4.2 (2) and Example 2.4.4, we have

$$(A^T - 2I) = \left[(A^T - 2I)^{-1} \right]^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix}$$

Hence $A^T = 2I + \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix}$, so

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

by Theorem 2.4.4(7).

> An interactive H5P element has been excluded from this version of the text. You can view it online here:
> https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-55

The following important theorem collects a number of conditions all equivalent to invertibility. It will be referred to frequently below.

Theorem 2.4.5 Inverse Theorem

The following conditions are equivalent for an $n \times n$ matrix A:

1. A is invertible.

2. The homogeneous system $A\vec{x} = \vec{0}$ has only the trivial solution $\vec{x} = \vec{0}$.

3. A can be carried to the identity matrix I_n by elementary row operations.

4. The system $A\vec{x} = \vec{b}$ has at least one solution $\vec{x}$ for every choice of column $\vec{b}$.

5. There exists an $n \times n$ matrix C such that $AC = I_n$.

Proof:

We show that each of these conditions implies the next, and that (5) implies (1).

(1) $\implies$ (2). If A^{-1} exists, then $A\vec{x} = \vec{0}$ gives $\vec{x} = I_n\vec{x} = A^{-1}A\vec{x} = A^{-1}0 = 0$.

(2) $\implies$ (3). Assume that (2) is true. Certainly $A \to R$ by row operations where R is a reduced, row-echelon matrix. It suffices to show that $R = I_n$. Suppose that this is not the case. Then R has a row of zeros (being square). Now consider the augmented matrix $\left[\ A \mid \vec{0}\ \right]$ of the system $A\vec{x} = \vec{0}$. Then $\left[\ A \mid \vec{0}\ \right] \to \left[\ R \mid \vec{0}\ \right]$ is the reduced form, and $\left[\ R \mid \vec{0}\ \right]$ also has a row of zeros. Since R is square there must be at least one nonleading variable, and hence at least one parameter. Hence the system $A\vec{x} = \vec{0}$ has infinitely many solutions, contrary to (2). So $R = I_n$ after all.

(3) $\implies$ (4). Consider the augmented matrix $\left[\ A \mid \vec{b}\ \right]$ of the system $A\vec{x} = \vec{b}$. Using (3), let $A \to I_n$ by a sequence of row operations. Then these same operations carry $\left[\ A \mid \vec{b}\ \right] \to \left[\ I_n \mid \vec{c}\ \right]$ for some column $\vec{c}$. Hence the system $A\vec{x} = \vec{b}$ has a solution (in fact unique) by gaussian elimination. This proves (4).

(4) $\implies$ (5). Write $I_n = \left[\ \vec{e}_1 \quad \vec{e}_2 \quad \cdots \quad \vec{e}_n\ \right]$ where $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ are the columns of I_n. For each \newline $j = 1, 2, \ldots, n$, the system $A\vec{x} = \vec{e}_j$ has a solution $\vec{c}_j$ by (4), so $A\vec{c}_j = \vec{e}_j$. Now let $C = \left[\ \vec{c}_1 \quad \vec{c}_2 \quad \cdots \quad \vec{c}_n\ \right]$ be the $n \times n$ matrix with these matrices $\vec{c}_j$ as its columns. Then Definition 2.9 gives (5):

$$AC = A \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{bmatrix} = \begin{bmatrix} A\vec{c}_1 & A\vec{c}_2 & \cdots & A\vec{c}_n \end{bmatrix} = \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \end{bmatrix} = I_n$$

(5) $\implies$ (1). Assume that (5) is true so that $AC = I_n$ for some matrix C. Then $C\vec{x} = 0$ implies $\vec{x} = 0$ (because $\vec{x} = I_n\vec{x} = AC\vec{x} = A\vec{0} = \vec{0}$). Thus condition (2) holds for the matrix C rather than A. Hence the argument above that (2) $\implies$ (3) $\implies$ (4) $\implies$ (5) (with A replaced by C) shows that a matrix C' exists such that $CC' = I_n$. But then

$$A = AI_n = A(CC') = (AC)C' = I_nC' = C'$$

Thus $CA = CC' = I_n$ which, together with $AC = I_n$, shows that C is the inverse of A. This proves (1).

The proof of (5) $\implies$ (1) in Theorem 2.4.5 shows that if $AC = I$ for square matrices, then necessarily $CA = I$, and hence that C and A are inverses of each other. We record this important fact for reference.

Corollary 2.4.1

If A and C are square matrices such that $AC = I$, then also $CA = I$. In particular, both A and C are invertible, $C = A^{-1}$, and $A = C^{-1}$.

Here is a quick way to remember Corollary 2.4.1. If A is a square matrix, then

1. If $AC = I$ then $C = A^{-1}$.
2. If $CA = I$ then $C = A^{-1}$.

Observe that Corollary 2.4.1 is false if A and C are not square matrices. For example, we have

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = I_2 \quad \text{but} \quad \begin{bmatrix} -1 & 1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \neq I_3$$

In fact, it can be verified that if $AB = I_m$ and $BA = I_n$, where A is $m \times n$ and B is $n \times m$, then $m = n$ and A and B are (square) inverses of each other.

An $n \times n$ matrix A has $rank\, n$ if and only if (3) of Theorem 2.4.5 holds. Hence

Corollary 2.4.2

An $n \times n$ matrix A is invertible if and only if $rank\, A = n$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66#h5p-54

3. Determinants and Diagonalization

Introduction

With each square matrix we can calculate a number, called the **determinant** of the matrix, which tells us whether or not the matrix is invertible. In fact, determinants can be used to give a formula for the inverse of a matrix. They also arise in calculating certain numbers (called **eigenvalues**) associated with the matrix. These eigenvalues are essential to a technique called diagonalization that is used in many applications where it is desired to predict the future behaviour of a system. For example, we use it to predict whether a species will become extinct.

Determinants were first studied by Leibnitz in 1696, and the term "determinant" was first used in 1801 by Gauss is his *Disquisitiones Arithmeticae*. Determinants are much older than matrices (which were introduced by Cayley in 1878) and were used extensively in the eighteenth and nineteenth centuries, primarily because of their significance in geometry. Although they are somewhat less important today, determinants still play a role in the theory and application of matrix algebra.

3.1 The Cofactor Expansion

In Section 2.4, we defined the determinant of a 2×2 matrix

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

as follows:

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

and showed (in Example 2.4.4) that A has an inverse if and only if det $A \neq 0$. One objective of this chapter is to do this for *any* square matrix A. There is no difficulty for 1×1 matrices: If $A = [a]$, we define $det\, A = det\, [a] = a$ and note that A is invertible if and only if $a \neq 0$.

If A is 3×3 and invertible, we look for a suitable definition of $det\, A$ by trying to carry A to the identity matrix by row operations. The first column is not zero (A is invertible); suppose the (1, 1)-entry a is not zero. Then row operations give

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \rightarrow \begin{bmatrix} a & b & c \\ ad & ae & af \\ ag & ah & ai \end{bmatrix} \rightarrow \begin{bmatrix} a & b & c \\ 0 & ae-bd & af-cd \\ 0 & ah-bg & ai-cg \end{bmatrix} = \begin{bmatrix} a & b & c \\ 0 & u & af-cd \\ 0 & v & ai-cg \end{bmatrix}$$

where $u = ae - bd$ and $v = ah - bg$. Since A is invertible, one of u and v is nonzero (by Example 2.4.11); suppose that $u \neq 0$. Then the reduction proceeds

$$A \rightarrow \begin{bmatrix} a & b & c \\ 0 & u & af-cd \\ 0 & v & ai-cg \end{bmatrix} \rightarrow \begin{bmatrix} a & b & c \\ 0 & u & af-cd \\ 0 & uv & u(ai-cg) \end{bmatrix} \rightarrow \begin{bmatrix} a & b & c \\ 0 & u & af-cd \\ 0 & 0 & w \end{bmatrix}$$

where

$$w = u(ai - cg) - v(af - cd) = a(aei + bfg + cdh - ceg - afh - bdi)$$

. We define

$$det A = aei + bfg + cdh - ceg - afh - bdi$$

and observe that $det A \neq 0$ because $det A = w \neq 0$ (is invertible).

To motivate the definition below, collect the terms in Equation 3.1 involving the entries a, b, and c in row 1 of A:

$$det A = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - afh - bdi$$

$$= a(ei - fh) - b(di - fg) + c(dh - eg)$$

$$= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

This last expression can be described as follows: To compute the determinant of a 3×3 matrix A, multiply each entry in row 1 by a sign times the determinant of the 2×2 matrix obtained by deleting the row and column of that entry, and add the results. The signs alternate down row 1, starting with $+$. It is this observation that we generalize below.

Example 3.1.1

$$det \begin{bmatrix} 2 & 3 & 7 \\ -4 & 0 & 6 \\ 1 & 5 & 0 \end{bmatrix} = 2 \begin{vmatrix} 0 & 6 \\ 5 & 0 \end{vmatrix} - 3 \begin{vmatrix} -4 & 6 \\ 1 & 0 \end{vmatrix} + 7 \begin{vmatrix} -4 & 0 \\ 1 & 5 \end{vmatrix}$$
$$= 2(-30) - 3(-6) + 7(-20)$$
$$= -182$$

This suggests an inductive method of defining the determinant of any square matrix in terms of determinants
of matrices one size smaller. The idea is to define determinants of 3×3 matrices in terms of determinants of 2×2 matrices,

then we do 4×4 matrices in terms of 3×3 matrices, and so on.

To describe this, we need some terminology.

Definition 3.1 Cofactors of a matrix

Assume that determinants of $(n-1) \times (n-1)$ matrices have been defined. Given the $n \times n$ matrix A, let

A_{ij} denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j.

Then the (i, j)-**cofactor** $c_{ij}(A)$ is the scalar defined by

$$c_{ij}(A) = (-1)^{i+j} det(A_{ij})$$

Here $(-1)^{i+j}$ is called the **sign** of the (i, j)-position.

The sign of a position is clearly 1 or -1, and the following diagram is useful for remembering it:

$$\begin{bmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

Note that the signs alternate along each row and column with $+$ in the upper left corner.

Example 3.1.2

Find the cofactors of positions $(1, 2)$, $(3, 1)$, and $(2, 3)$ in the following matrix.

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

Solution:

Here A_{12} is the matrix $\begin{bmatrix} 5 & 7 \\ 8 & 4 \end{bmatrix}$ that remains when row 1 and column 2 are deleted. The sign of position $(1, 2)$ is $(-1)^{1+2} = -1$ (this is also the $(1, 2)$-entry in the sign diagram), so the $(1, 2)$-cofactor is

$$c_{12}(A) = (-1)^{1+2} \begin{vmatrix} 5 & 7 \\ 8 & 4 \end{vmatrix} = (-1)(5 \cdot 4 - 7 \cdot 8) = (-1)(-36) = 36$$

Turning to position $(3, 1)$, we find

$$c_{31}(A) = (-1)^{3+1} A_{31} = (-1)^{3+1} \begin{vmatrix} -1 & 6 \\ 2 & 7 \end{vmatrix} = (+1)(-7 - 12) = -19$$

Finally, the $(2, 3)$-cofactor is

$$c_{23}(A) = (-1)^{2+3} A_{23} = (-1)^{2+3} \begin{vmatrix} 3 & -1 \\ 8 & 9 \end{vmatrix} = (-1)(27 + 8) = -35$$

Clearly other cofactors can be found—there are nine in all, one for each position in the matrix.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-57

We can now define $det A$ for any square matrix A

Definition 3.2 Cofactor expansion of a Matrix

Assume that determinants of $(n-1) \times (n-1)$ matrices have been defined. If $A = [a_{ij}]$ is $n \times n$ define

$$det A = a_{11}c_{11}(A) + a_{12}c_{12}(A) + \cdots + a_{1n}c_{1n}(A)$$

This is called the **cofactor expansion** of $det A$ along row 1.

It asserts that $det A$ can be computed by multiplying the entries of row 1 by the corresponding
cofactors, and adding the results. The astonishing thing is that $det A$ can be computed by taking the cofactor expansion along *any row or column*: Simply multiply each entry of that row or column by the corresponding cofactor and add.

Theorem 3.1.1 Cofactor Expansion Theorem

The determinant of an $n \times n$ matrix A can be computed by using the cofactor expansion along any row or column
of A. That is $det A$ can be computed by multiplying each entry of the row or column by the corresponding cofactor and adding the results.

Example 3.1.3

Compute the determinant of

$$A = \begin{bmatrix} 3 & 4 & 5 \\ 1 & 7 & 2 \\ 9 & 8 & -6 \end{bmatrix}.$$

Solution:

The cofactor expansion along the first row is as follows:

$$\det A = 3c_{11}(A) + 4c_{12}(A) + 5c_{13}(A)$$

$$= 3 \begin{vmatrix} 7 & 2 \\ 8 & -6 \end{vmatrix} - 4 \begin{vmatrix} 1 & 2 \\ 9 & -6 \end{vmatrix} + 3 \begin{vmatrix} 1 & 7 \\ 9 & 8 \end{vmatrix}$$

$$= 3(-58) - 4(-24) + 5(-55)$$

$$= -353$$

Note that the signs alternate along the row (indeed along *any* row or column). Now we compute $\det A$ by expanding along the first column.

$$\det A = 3c_{11}(A) + 1c_{21}(A) + 9c_{31}(A)$$

$$= 3 \begin{vmatrix} 7 & 2 \\ 8 & -6 \end{vmatrix} - \begin{vmatrix} 4 & 5 \\ 8 & -6 \end{vmatrix} + 9 \begin{vmatrix} 4 & 5 \\ 7 & 2 \end{vmatrix}$$

$$= 3(-58) - (-64) + 9(-27)$$

$$= -353$$

The reader is invited to verify that $\det A$ can be computed by expanding along any other row or column.

The fact that the cofactor expansion along

any row or column of a matrix A always gives the same result (the determinant of A) is remarkable, to say the least. The choice of a particular row or column can simplify the calculation.

Compute $det A$ where
$$A = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 5 & 1 & 2 & 0 \\ 2 & 6 & 0 & -1 \\ -6 & 3 & 1 & 0 \end{bmatrix}.$$

Solution:

The first choice we must make is which row or column to use in the
cofactor expansion. The expansion involves multiplying entries by cofactors, so the work is minimized when the row or column contains as
many zero entries as possible. Row 1 is a best choice in this matrix (column 4 would do as well), and the expansion is

$$det A = 3c_{11}(A) + 0c_{12}(A) + 0c_{13}(A) + 0c_{14}(A)$$

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

This is the first stage of the calculation, and we have succeeded in expressing the determinant of the 4×4 matrix A
in terms of the determinant of a 3×3 matrix. The next stage involves
this 3×3 matrix. Again, we can use any row or column for the cofactor
expansion. The third column is preferred (with two zeros), so

$$det A = 3 \left(0 \begin{vmatrix} 6 & 0 \\ 3 & 1 \end{vmatrix} - (-1) \begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} + 0 \begin{vmatrix} 1 & 2 \\ 6 & 0 \end{vmatrix} \right)$$

$$= 3[0 + 1(-5) + 0]$$

$$= -15$$

This completes the calculation.

This example shows us that calculating a determinant is simplified a great deal when a row or column consists mostly of zeros. (In fact, when a row or column consists $entirely$ of zeros, the determinant is zero—simply expand along that row or column.) We did learn that one method of $creating$ zeros in a matrix is to apply elementary row operations to it. Hence, a natural question to ask is what effect such a row operation has on the determinant of the matrix. It turns out that the effect is easy to determine and that elementary $column$ operations can be used in the same way. These observations lead to a technique for evaluating determinants that greatly reduces the labour involved. The necessary information is given in Theorem 3.1.2.

Theorem 3.1.2

Let A denote an $n \times n$ matrix.

1. If A has a row or column of zeros, $det A = 0$.
2. If two distinct rows (or columns) of A are interchanged, the determinant of the resulting matrix is $-det A$.

3. If a row (or column) of A is multiplied by a constant u, the determinant of the resulting matrix is $u(det\,A)$.

4. If two distinct rows (or columns) of A are identical, $det\,A = 0$.

5. If a multiple of one row of A is added to a different row (or if a multiple of a column is added to a different column), the determinant of the resulting matrix is $det\,A$.

The following four examples illustrate how Theorem 3.1.2 is used to evaluate determinants.

Example 3.1.5

Evaluate $det\,A$ when

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

Solution:

The matrix does have zero entries, so expansion along (say) the second row would involve somewhat less work. However, a column operation can be

used to get a zero in position $(2, 3)$—namely, add column 1 to column 3. Because this does not change the value of the determinant, we obtain

$$det\,A = \begin{vmatrix} 1 & -1 & 3 \\ 1 & 0 & -1 \\ 2 & 1 & 6 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 4 \\ 1 & 0 & 0 \\ 2 & 1 & 8 \end{vmatrix} = -\begin{vmatrix} -1 & 4 \\ 1 & 8 \end{vmatrix} = 12$$

where we expanded the second 3×3 matrix along row 2.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-58

Example 3.1.6

If $det \begin{bmatrix} a & b & c \\ p & q & r \\ x & y & z \end{bmatrix} = 6,$

evaluate $det\,A$ where

$$A = \begin{bmatrix} a+x & b+y & c+z \\ 3x & 3y & 3z \\ -p & -q & -r \end{bmatrix}.$$

Solution:

First take common factors out of rows 2 and 3.

$$det A = 3(-1)det \begin{bmatrix} a+x & b+y & c+z \\ x & y & z \\ p & q & r \end{bmatrix}$$

Now subtract the second row from the first and interchange the last two rows.

$$det A = -3det \begin{bmatrix} a & b & c \\ x & y & z \\ p & q & r \end{bmatrix} = 3det \begin{bmatrix} a & b & c \\ p & q & r \\ x & y & z \end{bmatrix} = 3 \cdot 6 = 18$$

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=70#h5p-59

The determinant of a matrix is a sum of products of its entries. In particular, if these entries are polynomials in x, then the determinant itself is a polynomial in x. It is often of interest to determine which values of x make the determinant zero, so it is very useful if the determinant is given in factored form. Theorem 3.1.2 can help.

Example 3.1.7

Find the values of x for which $\det A = 0$, where
$$A = \begin{bmatrix} 1 & x & x \\ x & 1 & x \\ x & x & 1 \end{bmatrix}.$$

Solution:

To evaluate $\det A$, first subtract x times row 1 from rows 2 and 3.

$$\det A = \begin{vmatrix} 1 & x & x \\ x & 1 & x \\ x & x & 1 \end{vmatrix} = \begin{vmatrix} 1 & x & x \\ 0 & 1-x^2 & x-x^2 \\ 0 & x-x^2 & 1-x^2 \end{vmatrix} = \begin{vmatrix} 1-x^2 & x-x^2 \\ x-x^2 & 1-x^2 \end{vmatrix}$$

At this stage we could simply evaluate the determinant (the result is $2x^3 - 3x^2 + 1$). But then we would have to factor this polynomial to find the values of x that make it zero. However, this factorization can be obtained directly by first factoring each entry in the determinant and taking a common
factor of $(1-x)$ from each row.

$$\det A = \begin{vmatrix} (1-x)(1+x) & x(1-x) \\ x(1-x) & (1-x)(1+x) \end{vmatrix} = (1-x)^2 \begin{vmatrix} 1+x & x \\ x & 1+x \end{vmatrix}$$
$$= (1-x)^2(2x+1)$$

Hence, $\det A = 0$ means $(1-x)^2(2x+1) = 0$, that is $x = 1$ or $x = -\frac{1}{2}$.

Example 3.1.8

If a_1, a_2, and a_3 are given show that

$$det \begin{bmatrix} 1 & a_1 & a_1^2 \\ 1 & a_2 & a_2^2 \\ 1 & a_3 & a_3^2 \end{bmatrix} = (a_3 - a_1)(a_3 - a_2)(a_2 - a_1)$$

Solution:

Begin by subtracting row 1 from rows 2 and 3, and then expand along column 1:

$$det \begin{bmatrix} 1 & a_1 & a_1^2 \\ 1 & a_2 & a_2^2 \\ 1 & a_3 & a_3^2 \end{bmatrix} = det \begin{bmatrix} 1 & a_1 & a_1^2 \\ 0 & a_2 - a_1 & a_2^2 - a_1^2 \\ 0 & a_3 - a_1 & a_3^2 - a_1^2 \end{bmatrix} = \begin{bmatrix} a_2 - a_1 & a_2^2 - a_1^2 \\ a_3 - a_1 & a_3^2 - a_1^2 \end{bmatrix}$$

Now $\left(a_2 - a_1\right)$ and $\left(a_3 - a_1\right)$ are common factors in rows 1 and 2, respectively, so

$$det \begin{bmatrix} 1 & a_1 & a_1^2 \\ 1 & a_2 & a_2^2 \\ 1 & a_3 & a_3^2 \end{bmatrix} = (a_2 - a_1)(a_3 - a_1)det \begin{bmatrix} 1 & a_2 + a_1 \\ 1 & a_3 + a_1 \end{bmatrix}$$

$$= (a_2 - a_1)(a_3 - a_1)(a_3 - a_2)$$

The matrix in Example 3.1.8 is called a **Vandermonde matrix**, and the formula for its determinant can be generalized to the $n \times n$ case.

If A is an $n \times n$ matrix, forming uA means multiplying *every* row of A by u. Applying property 3 of Theorem 3.1.2, we can take the common factor u out of each row and so obtain the following useful result.

Theoerem 3.1.3

If A is an $n \times n$ matrix, then $det(uA) = u^n det A$

for any number u.

The next example displays a type of matrix whose determinant is easy to compute.

Example 3.1.9

Evaluate $\det A$ if
$$A = \begin{bmatrix} a & 0 & 0 & 0 \\ u & b & 0 & 0 \\ v & w & c & 0 \\ x & y & z & d \end{bmatrix}.$$

Solution:

Expand along row 1 to get $\det A = a \begin{vmatrix} b & 0 & 0 \\ w & c & 0 \\ y & z & d \end{vmatrix}$. Now expand this along the top row to get

$$det A = ab \begin{vmatrix} c & 0 \\ z & d \end{vmatrix} = abcd,$$ the product of the main diagonal entries.

A square matrix is called a **lower triangular matrix** if all entries above the main diagonal are zero (as in Example 3.1.9). Similarly, an **upper triangular matrix** is one for which all entries below the main diagonal are zero. A **triangular matrix** is one that is either upper or lower triangular. Theorem 3.1.4 gives an easy rule for calculating the determinant of any triangular matrix.

Theorem 3.1.4

If A is a square triangular matrix, then det A is the product of the entries on the main diagonal.

Theorem 3.1.4 is useful in computer calculations because it is a routine matter to carry a matrix to triangular form using row operations.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-60

3.2 Determinants and Matrix Inverses

In this section, several theorems about determinants are derived. One consequence of these theorems is that a square matrix A is invertible if and only if $det\,A \neq 0$. Moreover, determinants are used to give a formula for A^{-1} which, in turn, yields a formula (called Cramer's rule) for the
solution of any system of linear equations with an invertible coefficient matrix.

We begin with a remarkable theorem (due to Cauchy in 1812) about the determinant of a product of matrices.

Theorem 3.2.1 Product Theorem

If A and B are $n \times n$ matrices, then
$$det(AB) = det\,A\,det\,B.$$

The complexity of matrix multiplication makes the product theorem quite unexpected. Here is an example where it reveals an important numerical identity.

Example 3.2.1

If $A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ and $B = \begin{bmatrix} c & d \\ -d & c \end{bmatrix}$

then $AB = \begin{bmatrix} ac - bd & ad + bc \\ -(ad + bc) & ac - bd \end{bmatrix}$.

Hence $det A det B = det(AB)$ gives the identity

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2$$

Theorem 3.2.1 extends easily to $det(ABC) = det A det B det C$. In fact, induction gives

$$det(A_1 A_2 \cdots A_{k-1} A_k) = det A_1 det A_2 \cdots det A_{k-1} det A_k$$

for any square matrices $A_1, \ldots, A_k$ of the same size. In particular, if each $A_i = A$, we obtain

$$det(A^k) = (det A)^k, \text{ for any } k \geq 1$$

We can now give the invertibility condition.

Theorem 3.2.2

An $n \times n$ matrix A is invertible if and only if $det A \neq 0$. When this is the case,
$det(A^{-1}) = \frac{1}{det A}$

Proof:
If A is invertible, then $AA^{-1} = I$; so the product theorem gives

$$1 = det\,I = det(AA^{-1}) = det\,A\,det\,A^{-1}$$

Hence, $det\,A \neq 0$ and also $det\,A^{-1} = \frac{1}{det\,A}$.

Conversely, if $det\,A \neq 0$, we show that A can be carried to I by elementary row operations (and invoke Theorem 2.4.5). Certainly, A can be carried to its reduced row-echelon form R, so $R = E_k \cdots E_2 E_1 A$ where the E_i are elementary matrices (Theorem 2.5.1). Hence the product theorem gives

$$det\,R = det\,E_k \cdots det\,E_2\,det\,E_1\,det\,A$$

Since $det\,E \neq 0$ for all elementary matrices E, this shows $det\,R \neq 0$. In particular, R has no row of zeros, so $R = I$ because R is square and reduced row-echelon. This is what we wanted.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-61

Example 3.2.2

For which values of c does $A = \begin{bmatrix} 1 & 0 & -c \\ -1 & 3 & 1 \\ 0 & 2c & -4 \end{bmatrix}$ have an inverse?

Solution:

Compute $det A$ by first adding c times column 1 to column 3 and then expanding along row 1.

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

Hence, $det A = 0$ if $c = -2$ or $c = 3$, and A has an inverse if $c \neq -2$ and $c \neq 3$.

Example 3.2.3

If a product $A_1 A_2 \cdots A_k$ of square matrices is invertible, show that each A_i is invertible.

Solution:

We have $det A_1 det A_2 \cdots det A_k = det(A_1 A_2 \cdots A_k)$ by the product theorem, and $det(A_1 A_2 \cdots A_k) \neq 0$ by Theorem 3.2.2 because $A_1 A_2 \cdots A_k$ is invertible. Hence

$det A_1 det A_2 \cdots det A_k \neq 0$

so $det A_i \neq 0$ for each i. This shows that each A_i is invertible, again by Theorem 3.2.2.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

Theorem 3.2.3

If A is any square matrix, $det A^T = det A$.

Proof:

Consider first the case of an elementary matrix E. If E is of type I or II, then $E^T = E$; so certainly $det E^T = det E$. If E is of type III, then E^T is also of type III; so $det E^T = 1 = det E$ by Theorem 3.1.2. Hence, $det E^T = det E$ for every elementary matrix E.

Now let A be any square matrix. If A is not invertible, then neither is A^T; so $det A^T = 0 = det A$ by Theorem 3.1.2. On the other hand, if A is invertible, then $A = E_k \cdots E_2 E_1$, where the E_i are elementary matrices (Theorem 2.5.2). Hence, $A^T = E_1^T E_2^T \cdots E_k^T$ so the product theorem gives

$$det A^T = det E_1^T det E_2^T \cdots det E_k^T = det E_1 det E_2 \cdots det E_k$$
$$= det E_k \cdots det E_2 det E_1$$
$$= det A$$

This completes the proof.

Example 3.2.4

If $det\,A = 2$ and $det\,B = 5$, calculate $det(A^3 B^{-1} A^T B^2)$.

Solution:

We use several of the facts just derived.

$$
\begin{aligned}
det(A^3 B^{-1} A^T B^2) &= det(A^3)det(B^{-1})det(A^T)det(B^2) \\
&= (det\,A)^3 \frac{1}{det\,B} det\,A (det\,B)^2 \\
&= 2^3 \cdot \frac{1}{5} \cdot 2 \cdot 5^2 \\
&= 80
\end{aligned}
$$

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-62

A square matrix is called **orthogonal** if $A^{-1} = A^T$. What are the possible values of $det A$ if A is orthogonal?

Solution:

If A is orthogonal, we have $I = AA^T$. Take determinants to obtain

$$1 = det I = det(AA^T) = det A \, det A^T = (det A)^2$$

Since $det A$ is a number, this means $det A = \pm 1$.

Adjugates

In Section 2.4 we defined the adjugate of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$
to be $adj(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.
Then we verified that $A(adj A) = (det A)I = (adj A)A$ and hence that, if $det A \neq 0$, $A^{-1} = \frac{1}{det A} adj A$. We are now able to define the adjugate of an arbitrary square matrix and to show that this formula for the inverse remains valid (when the inverse exists).

Recall that the (i, j)-cofactor $c_{ij}(A)$ of a square matrix A is a number defined for each position (i, j) in the matrix. If A is a square matrix, the **cofactor matrix of** A is defined to be

the matrix $[c_{ij}(A)]$ whose (i, j)-entry is the (i, j)-cofactor of A.

Definition 3.3 Adjugate of a Matrix

The **adjugate** of A, denoted $adj(A)$, is the transpose of this cofactor matrix; in symbols,

$$adj(A) = [c_{ij}(A)]^T$$

Example 3.2.6

Compute the adjugate of $A = \begin{bmatrix} 1 & 3 & -2 \\ 0 & 1 & 5 \\ -2 & -6 & 7 \end{bmatrix}$ and calculate $A(adj\,A)$ and $(adj\,A)A$.

Solution:

We first find the cofactor matrix.

$$\begin{bmatrix} c_{11}(A) & c_{12}(A) & c_{13}(A) \\ c_{21}(A) & c_{22}(A) & c_{23}(A) \\ c_{31}(A) & c_{32}(A) & c_{33}(A) \end{bmatrix} = \begin{bmatrix} \begin{vmatrix} 1 & 5 \\ -6 & 7 \end{vmatrix} & -\begin{vmatrix} 0 & 5 \\ -2 & 7 \end{vmatrix} & \begin{vmatrix} 0 & 1 \\ -2 & -6 \end{vmatrix} \\ -\begin{vmatrix} 3 & -2 \\ -6 & 7 \end{vmatrix} & \begin{vmatrix} 1 & -2 \\ -2 & 7 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ -2 & -6 \end{vmatrix} \\ \begin{vmatrix} 3 & -2 \\ 1 & 5 \end{vmatrix} & -\begin{vmatrix} 1 & -2 \\ 0 & 5 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 0 & 1 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} 37 & -10 & 2 \\ -9 & 3 & 0 \\ 17 & -5 & 1 \end{bmatrix}$$

Then the adjugate of A is the transpose of this cofactor matrix.

$$adj\,A = \begin{bmatrix} 37 & -10 & 2 \\ -9 & 3 & 0 \\ 17 & -5 & 1 \end{bmatrix}^{T} = \begin{bmatrix} 37 & -9 & 17 \\ -10 & 3 & -5 \\ 2 & 0 & 1 \end{bmatrix}$$

The computation of $A(adj\,A)$ gives

$$A(adj\,A) = \begin{bmatrix} 1 & 3 & -2 \\ 0 & 1 & 5 \\ -2 & -6 & 7 \end{bmatrix} \begin{bmatrix} 37 & -9 & 17 \\ -10 & 3 & -5 \\ 2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} = 3I$$

and the reader can verify that also $(adj\,A)A = 3I$. Hence, analogy with the 2×2 case would indicate that $det\,A = 3$; this is, in fact, the case.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=70#h5p-64

The relationship $A(adj\,A) = (det\,A)I$ holds for any square matrix A.

If A is any square matrix, then

$$A(adj\,A) = (det\,A)I = (adj\,A)A$$

In particular, if det A $\neq$ 0, the inverse of A is given by

$$A^{-1} = \frac{1}{det\,A}adj\,A$$

It is important to note that this theorem is not an efficient way to find the inverse of the matrix A. For example, if A were 10×10, the calculation of $adj\,A$ would require computing $10^2 = 100$ determinants of 9×9 matrices! On the other hand, the matrix inversion algorithm would find A^{-1} with about the same effort as finding $det\,A$. Clearly, Theorem 3.2.4 is not a $practical$ result: its virtue is that it gives a formula for A^{-1} that is useful for $theoretical$ purposes.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-65

Example 3.2.7

Find the $(2, 3)$-entry of A^{-1} if

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 5 & -7 & 1 \\ 3 & 0 & -6 \end{bmatrix}.$$

Solution:

First compute

$$det\,A = \begin{vmatrix} 2 & 1 & 3 \\ 5 & -7 & 1 \\ 3 & 0 & -6 \end{vmatrix} = \begin{vmatrix} 2 & 1 & 7 \\ 5 & -7 & 11 \\ 3 & 0 & 0 \end{vmatrix} = 3 \begin{vmatrix} 1 & 7 \\ -7 & 11 \end{vmatrix} = 180$$

Since $A^{-1} = \frac{1}{det\,A}adj\,A = \frac{1}{180}\left[c_{ij}(A)\right]^{T}$,
the $(2, 3)$-entry of A^{-1} is the $(3, 2)$-entry of the matrix
$\frac{1}{180}c_{ij}(A)$; that is, it equals

$$\frac{1}{180}c_{32}(A) = \frac{1}{180}\left(-\begin{vmatrix} 2 & 3 \\ 5 & 1 \end{vmatrix}\right) = \frac{13}{180}.$$

Example 3.2.8

If A is $n \times n, n \geq 2$, show that
$$det(adj\,A) = (det\,A)^{n-1}.$$

Solution:

Write $d = \det A$; we must show that $\det(adj A) = d^{n-1}$. We have $A(adj A) = dI$ by Theorem 3.2.4, so taking determinants gives $d \det(adj A) = d^n$. Hence we are done if $d \neq 0$. Assume $d = 0$; we must show that $\det(adj A) = 0$, that is, $adj A$ is not invertible. If $A \neq 0$, this follows from $A(adj A) = dI = 0$; if $A = 0$, it follows because then $adj A = 0$.

Cramer's Rule

Theorem 3.2.4 has a nice application to linear equations. Suppose

$$A\vec{x} = \vec{b}$$

is a system of n equations in n variables $x_1, x_2, \ldots, x_n$. Here A is the $n \times n$ coefficient matrix and $\vec{x}$ and $\vec{b}$ are the columns

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ and } \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

of variables and constants, respectively. If $\det A \neq 0$, we left multiply by A^{-1} to obtain the solution $\vec{x} = A^{-1}\vec{b}$. When we use the adjugate formula, this becomes

$$
\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{det A}(adj A)\vec{b}
$$

$$
= \frac{1}{det A} \begin{bmatrix} c_{11}(A) & c_{21}(A) & \cdots & c_{n1}(A) \\ c_{12}(A) & c_{22}(A) & \cdots & c_{n2}(A) \\ \vdots & \vdots & & \vdots \\ c_{1n}(A) & c_{2n}(A) & \cdots & c_{nn}(A) \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}
$$

Hence, the variables $x_1, x_2, \ldots, x_n$ are given by

$$x_1 = \frac{1}{det A} \left[b_1 c_{11}(A) + b_2 c_{21}(A) + \cdots + b_n c_{n1}(A) \right]$$

$$x_2 = \frac{1}{det A} \left[b_1 c_{12}(A) + b_2 c_{22}(A) + \cdots + b_n c_{n2}(A) \right]$$

$$\vdots \qquad\qquad \vdots$$

$$x_n = \frac{1}{det A} \left[b_1 c_{1n}(A) + b_2 c_{2n}(A) + \cdots + b_n c_{nn}(A) \right]$$

Now the quantity $b_1 c_{11}(A) + b_2 c_{21}(A) + \cdots + b_n c_{n1}(A)$ occurring in the formula for x_1 looks like the cofactor expansion of the determinant of a matrix. The cofactors involved are $c_{11}(A), c_{21}(A), \ldots, c_{n1}(A)$, corresponding to the first column of A. If A_1 is obtained from A by replacing the first column of A by b, then $c_{i1}(A_1) = c_{i1}(A)$ for each i because column 1 is deleted when computing them. Hence, expanding $det(A_1)$ by the first column gives

$$det A_1 = b_1 c_{11}(A_1) + b_2 c_{21}(A_1) + \cdots + b_n c_{n1}(A_1)$$
$$= b_1 c_{11}(A) + b_2 c_{21}(A) + \cdots + b_n c_{n1}(A)$$
$$= (det A) x_1$$

Hence, $x_1 = \frac{det A_1}{det A}$ and similar results hold for the other variables.

Theorem 3.2.5 Cramer's Rule

If A is an invertible $n \times n$ matrix, the solution to the system
$$A\vec{x} = \vec{b}$$
of n equations in the variables $x_1, x_2, \ldots, x_n$ is given by

$$x_1 = \frac{det A_1}{det A}, \quad x_2 = \frac{det A_2}{det A}, \quad \cdots, \quad x_n = \frac{det A_n}{det A}$$

where, for each k, A_k is the matrix obtained from A by replacing column k by $\vec{b}$.

Example 3.2.9

Find x_1, given the following system of equations.
$$5x_1 + x_2 - x_3 = 4$$
$$9x_1 + x_2 - x_3 = 1$$
$$x_1 - x_2 + 5x_3 = 2$$

Solution:

Compute the determinants of the coefficient matrix A and the matrix A_1 obtained from it by replacing the first column by the column of constants.

$$det\,A = det \begin{bmatrix} 5 & 1 & -1 \\ 9 & 1 & -1 \\ 1 & -1 & 5 \end{bmatrix} = -16$$

$$det\,A_1 = det \begin{bmatrix} 4 & 1 & -1 \\ 1 & 1 & -1 \\ 2 & -1 & 5 \end{bmatrix} = 12$$

Hence, $x_1 = \frac{det\,A_1}{det\,A} = -\frac{3}{4}$ by Cramer's rule.

Cramer's rule is *not* an efficient way to solve linear systems or invert matrices. True, it enabled us to calculate x_1 here without computing x_2 or x_3. Although this might seem an advantage, the truth of the matter is that, for large systems of equations, the number of computations needed to find *all* the variables by the gaussian algorithm is comparable to the number required to find

one of the determinants involved in Cramer's rule. Furthermore, the algorithm works when the matrix of the system is not invertible and even when the coefficient matrix is not square. Like the adjugate formula, then, Cramer's rule is *not* a practical numerical technique; its virtue is theoretical.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-66

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-152

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-153

3.3 Diagonalization and Eigenvalues

The world is filled with examples of systems that evolve in time—the weather in a region, the economy of a nation, the diversity of an ecosystem, etc. Describing such systems is difficult in general and various methods have been developed in special cases. In this section we describe one such method, called *diagonalization,* which is one of the most important techniques in linear algebra. A very fertile example of this procedure is in modelling the growth of the population of an animal species. This has attracted more attention in recent years with the ever increasing awareness that many species are endangered. To motivate the technique, we begin by setting up a simple model of a bird population in which we make assumptions about survival and reproduction rates.

Example 3.3.1

Consider the evolution of the population of a species of birds. Because the number of males and females are nearly equal, we count only females. We assume that each female remains a juvenile for one year and then becomes an adult, and that only adults have offspring. We make three assumptions about reproduction and survival rates:

1. The number of juvenile females hatched in any year is twice the number of adult females alive the year before (we say the **reproduction rate** is 2).
2. Half of the adult females in any year survive to the next year (the **adult survival rate** is $\frac{1}{2}$).

Solution:

Let a_k and j_k denote, respectively, the number of adult and juvenile females after k years, so that the total female population is the sum $a_k + j_k$. Assumption 1 shows that $j_{k+1} = 2a_k$, while assumptions 2 and 3 show that $a_{k+1} = \frac{1}{2}a_k + \frac{1}{4}j_k$. Hence the numbers a_k and j_k in successive years are related by the following equations:

$$a_{k+1} = \frac{1}{2}a_k + \frac{1}{4}j_k$$

$$j_{k+1} = 2a_k$$

If we write $\vec{v}_k = \begin{bmatrix} a_k \\ j_k \end{bmatrix}$

and $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 2 & 0 \end{bmatrix}$

these equations take the matrix form

$$\vec{v}_{k+1} = A\vec{v}_k, \ \text{ for each } k = 0, 1, 2, \ldots$$

Taking $k = 0$ gives $\vec{v}_1 = A\vec{v}_0$, then taking $k = 1$ gives $\vec{v}_2 = A\vec{v}_1 = A^2\vec{v}_0$, and taking $k = 2$ gives $\vec{v}_3 = A\vec{v}_2 = A^3\vec{v}_0$. Continuing in this way, we get

$$\vec{v}_k = A^k v_0, \ \text{ for each } k = 0, 1, 2, \ldots$$

Since $\vec{v}_0 = \begin{bmatrix} a_0 \\ j_0 \end{bmatrix} = \begin{bmatrix} 100 \\ 40 \end{bmatrix}$

is known, finding the population profile $\vec{v}_k$ amounts to computing A^k for all $k \geq 0$. We will complete this calculation in Example 3.3.12 after some new techniques have been developed.

Let A be a fixed $n \times n$ matrix. A sequence $\vec{v}_0, \vec{v}_1, \vec{v}_2, \ldots$ of column vectors in n is called a **linear dynamical system**. Many models regard $\vec{v}_t$ as a continuous function of the time t, and replace our condition between $\vec{b}_{k+1}$ and $A\vec{v}_k$ with a differential relationship viewed as functions of time if $\vec{v}_0$ is known and the other $\vec{v}_k$ are determined (as in Example 3.3.1) by the conditions

$$\vec{v}_{k+1} = A\vec{v}_k \text{ for each } k = 0, 1, 2, \ldots$$

These conditions are called a **matrix recurrence** for the vectors $\vec{v}_k$. As in Example 3.3.1, they imply that

$$\vec{v}_k = A^k \vec{v}_0 \text{ for all } k \geq 0$$

so finding the columns $\vec{v}_k$ amounts to calculating A^k for $k \geq 0$
.

Direct computation of the powers A^k of a square matrix A can be time-consuming, so we adopt an indirect method that is commonly used. The idea is to first **diagonalize** the matrix A, that is, to find an invertible matrix P such that

(3.8) $\quad P^{-1}AP = D$ is a diagonal matrix

This works because the powers D^k of the diagonal matrix D are easy to compute, and Equation (3.8) enables us to compute powers A^k of the matrix A in terms of powers D^k of D. Indeed, we can solve Equation (3.8) for A to get $A = PDP^{-1}$. Squaring this gives

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD^2P^{-1}$$

Using this we can compute A^3 as follows:
$$A^3 = AA^2 = (PDP^{-1})(PD^2P^{-1}) = PD^3P^{-1}$$

Continuing in this way we obtain Theorem 3.3.1 (even if D is not diagonal).

If $A = PDP^{-1}$ then $A^k = PD^kP^{-1}$ for each $k = 1, 2, \ldots$.

Hence computing A^k comes down to finding an invertible matrix P as in equation Equation (3.8). To do this it is necessary to first compute certain numbers (called eigenvalues) associated with the matrix A.

Eigenvalue and Eigenvectors

Definition 3.4 Eigenvalues and Eigenvectors of a Matrix

If A is an $n \times n$ matrix, a number λ is called an **eigenvalue** of A if

$$A\vec{x} = \lambda\vec{x} \text{ for some column } \vec{x} \neq \vec{0} \text{ in } {}^n$$

In this case, $\vec{x}$ is called an **eigenvector** of A corresponding to the eigenvalue λ, or a λ–**eigenvector** for short.

Example 3.3.2

If $A = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix}$ and $\vec{x} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ then $A\vec{x} = 4\vec{x}$ so $\lambda = 4$ is an eigenvalue of A with corresponding eigenvector $\vec{x}$.

The matrix A in Example 3.3.2 has another eigenvalue in addition to $\lambda = 4$. To find it, we develop a general procedure for *any* $n \times n$ matrix A.

By definition a number λ is an eigenvalue of the $n \times n$ matrix A if and only if $A\vec{x} = \lambda\vec{x}$ for some column $\vec{x} \neq \vec{0}$. This is equivalent to asking that the homogeneous system

$$(\lambda I - A)\vec{x} = \vec{0}$$

of linear equations has a nontrivial solution $\vec{x} \neq \vec{0}$. By Theorem 2.4.5 this happens if and only if the matrix $\lambda I - A$ is not invertible and this, in turn, holds if and only if the determinant of the coefficient matrix is zero:

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

This last condition prompts the following definition:

Definition 3.5 Characteristic Polynomial of a Matrix

If A is an $n \times n$ matrix, the

characteristic polynomial $c_A(x)$ of A is defined by
$$c_A(x) = det(xI - A)$$

Note that $c_A(x)$ is indeed a polynomial in the variable x, and it has degree n when A is an $n \times n$ matrix (this is illustrated in the examples below). The above discussion shows that a number λ is an eigenvalue of A if and only if $c_A(\lambda) = 0$, that is if and only if λ is a **root** of the characteristic polynomial $c_A(x)$. We record these observations in

Theorem 3.3.2

Let A be an $n \times n$ matrix.

1. The eigenvalues λ of A are the roots of the characteristic polynomial $c_A(x)$ of A.
2. The λ-eigenvectors $\vec{x}$ are the nonzero solutions to the homogeneous system

$$(\lambda I - A)\vec{x} = \vec{0}$$

of linear equations with $\lambda I - A$ as coefficient matrix.

In practice, solving the equations in part 2 of Theorem 3.3.2 is a routine application of gaussian elimination, but finding the eigenvalues can be difficult, often requiring computers. For now, the examples and exercises will be constructed so that the roots of the characteristic polynomials are relatively easy to find

(usually integers). However, the reader should not be misled by this into thinking that eigenvalues are so easily obtained for the matrices that occur in practical applications!

Example 3.3.3

Find the characteristic polynomial of the matrix
$$A = \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix}$$
discussed in Example 3.3.2, and then find all the eigenvalues and their eigenvectors.

Solution:

Since
$$xI - A = \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} - \begin{bmatrix} 3 & 5 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} x-3 & -5 \\ -1 & x+1 \end{bmatrix}$$

we get

$$c_A(x) = det \begin{bmatrix} x-3 & -5 \\ -1 & x+1 \end{bmatrix} = x^2 - 2x - 8 = (x-4)(x+2)$$

Hence, the roots of $c_A(x)$ are $\lambda_1 = 4$ and $\lambda_2 = -2$, so these are the eigenvalues of A. Note that $\lambda_1 = 4$ was the eigenvalue mentioned in Example 3.3.2, but we have found a new one: $\lambda_2 = -2$.

To find the eigenvectors corresponding to $\lambda_2 = -2$, observe that in this case

$$(\lambda_2 I - A)\vec{x} = \begin{bmatrix} \lambda_2 - 3 & -5 \\ -1 & \lambda_2 + 1 \end{bmatrix} = \begin{bmatrix} -5 & -5 \\ -1 & -1 \end{bmatrix}$$

so the general solution to $(\lambda_2 I - A)\vec{x} = \vec{0}$ is $\vec{x} = t \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ where t is an arbitrary real number. Hence, the eigenvectors $\vec{x}$ corresponding to λ_2 are $\vec{x} = t \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ where $t \neq 0$ is arbitrary. Similarly, $\lambda_1 = 4$ gives rise to the eigenvectors $\vec{x} = t \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, $t \neq 0$ which includes the observation in Example 3.3.2.

⌨ *An interactive H5P element has been excluded from this version of the text. You can view it online here:*
https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=70#h5p-68

Note that a square matrix A has *many* eigenvectors associated with any given eigenvalue λ. In fact *every* nonzero solution $\vec{x}$ of $(\lambda I - A)\vec{x} = \vec{0}$ is an eigenvector. Recall that these solutions are all linear combinations of certain basic solutions determined by the gaussian algorithm (see Theorem 1.3.2). Observe that any nonzero multiple of an eigenvector is again an eigenvector, and such multiples are often more convenient. Any set of nonzero multiples of the basic solutions of $(\lambda I - A)\vec{x} = \vec{0}$ will be called a set of **basic eigenvectors** corresponding to λ.

https://www.geogebra.org/m/DJXTtm2k

Please answer these questions after you open the webpage:

1. Set the matrix to be

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

2. Drag the point u until you see the vector $\vec{u}$ and $M\vec{u}$ are on the same line. Record the value of λ. How many times do you see $\vec{u}$ and $M\vec{u}$ lying on the same line when $\vec{u}$ travel through the whole circle? Why?

3. Based on your observation, what can we say about the eigenvalue and eigenvector of M?

4. Set the matrix to be

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

and repeat what you did above.

5. Check your lecture notes about the eigenvalues and eigenvectors of this matrix. Are the results consistent with what you observe?

Example 3.3.4:

Find the characteristic polynomial, eigenvalues, and basic eigenvectors for

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

Solution:

Here the characteristic polynomial is given by

$$c_A(x) = det \begin{bmatrix} x-2 & 0 & 0 \\ -1 & x-2 & 1 \\ -1 & -3 & x+2 \end{bmatrix} = (x-2)(x-1)(x+1)$$

so the eigenvalues are $\lambda_1 = 2$, $\lambda_2 = 1$, and $\lambda_3 = -1$. To find all eigenvectors for $\lambda_1 = 2$, compute

$$\lambda_1 I - A = \begin{bmatrix} \lambda_1 - 2 & 0 & 0 \\ -1 & \lambda_1 - 2 & 1 \\ -1 & -3 & \lambda_1 + 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ -1 & -3 & 4 \end{bmatrix}$$

We want the (nonzero) solutions to $(\lambda_1 I - A)\vec{x} = \vec{0}$. The augmented matrix becomes

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

using row operations. Hence, the general solution $\vec{x}$ to

$(\lambda_1 I - A)\vec{x} = \vec{0}$ is $\vec{x} = t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

where t is arbitrary, so we can use $\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

as the basic eigenvector corresponding to $\lambda_1 = 2$. As the reader can verify, the gaussian algorithm gives basic eigenvectors $\vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

and $\vec{x}_3 = \begin{bmatrix} 0 \\ \frac{1}{3} \\ 1 \end{bmatrix}$

corresponding to $\lambda_2 = 1$ and $\lambda_3 = -1$, respectively. Note that to eliminate fractions, we could instead use $3\vec{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$

as the basic λ_3-eigenvector.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-155

Example 3.3.5

If A is a square matrix, show that A and A^T have the same characteristic polynomial, and hence the same eigenvalues.

Solution:

We use the fact that $xI - A^T = (xI - A)^T$. Then

$$c_{A^T}(x) = det\left(xI - A^T\right) = det\left[(xI - A)^T\right] = det(xI - A) = c_A(x)$$

by Theorem 3.2.3. Hence $c_{A^T}(x)$ and $c_A(x)$ have the same roots, and so A^T and A have the same eigenvalues (by Theorem 3.3.2).

The eigenvalues of a matrix need not be distinct. For example, if

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

the characteristic polynomial is $(x - 1)^2$ so the eigenvalue 1 occurs twice. Furthermore, eigenvalues are usually not computed as the roots of the characteristic polynomial. There are iterative, numerical methods that are much more efficient for large matrices.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70#h5p-67

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=70#h5p-156

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=70#h5p-157

4. Vector Geometry

4.1 Vectors and Lines

In this chapter we study the geometry of 3-dimensional space. We view a point in 3-space as an arrow from the origin to that point. Doing so provides a "picture" of the point that is truly worth a thousand words.

Vectors in $\mathbb{R}^3$

Introduce a coordinate system in 3-dimensional space in the usual way. First, choose a point O called the $origin$, then choose three mutually perpendicular lines through O, called the x, y, and $z\,axes$, and establish a number scale on each axis with zero at the origin. Given a point P in 3-space we associate three numbers x, y, and z with P, as described in Figure 4.1.1.
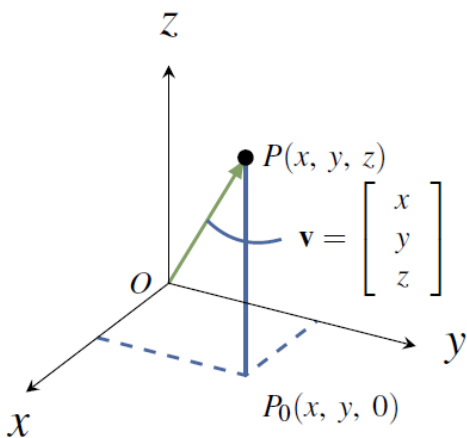
Figure 4.1.1

These numbers are called the *coordinates* of P, and we denote the point as (x, y, z), or $P(x, y, z)$ to emphasize the label P. The result is called a *cartesian* coordinate system for 3-space, and the resulting description of 3-space is called *cartesian geometry.*

As in the plane, we introduce vectors by identifying each point $P(x, y, z)$ with the vector

$$\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$ in $\mathbb{R}^3$, represented by the **arrow** from the origin

to P as in Figure 4.1.1. Informally, we say that the point P has vector $\vec{v}$, and that vector $\vec{v}$ has point P. In this way 3-space is identified with $\mathbb{R}^3$, and this identification will be made throughout this chapter, often without comment. In particular, the terms "vector" and "point" are interchangeable. The resulting description

of 3-space is called **vector geometry**. Note that the origin

is $\vec{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Length and direction

We are going to discuss two fundamental geometric properties of vectors in $\mathbb{R}^3$: length and direction. First, if $\vec{v}$ is a vector with point P, the **length** of vector $\vec{v}$ is defined to be the distance from the origin to P, that is the length of the arrow representing $\vec{v}$. The following properties of length will be used frequently.

Theorem 4.1.1

Let $\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ be a vector.

1. $||\vec{v}|| = \sqrt{x^2 + y^2 + z^2}$.
2. $\vec{v} = \vec{0}$ if and only if $||\vec{v}|| = 0$
3. $||a\vec{v}|| = |a|\,||\vec{v}||$ for all scalars a.

Proof:
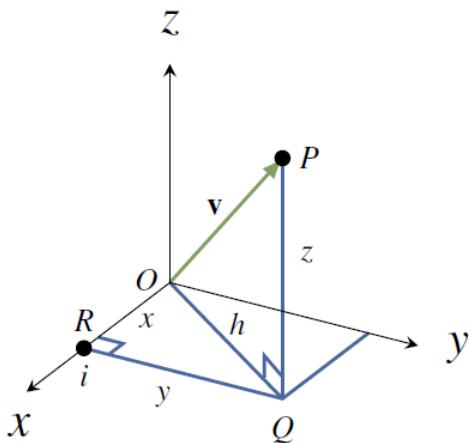
Figure 4.1.2

Let $\vec{v}$ have point $P(x, y, z)$.

1. In Figure 4.1.2, $||\vec{v}||$ is the hypotenuse of the right triangle OQP, and so $||\vec{v}||^2 = h^2 + z^2$ by Pythagoras' theorem. But h is the hypotenuse of the right triangle ORQ, so $h^2 = x^2 + y^2$. Now (1) follows by eliminating h^2 and taking positive square roots.

2. If $||\vec{v}|| = 0$, then $x^2 + y^2 + z^2 = 0$ by (1). Because squares of real numbers are nonnegative, it follows that $x = y = z = 0$, and hence that $\vec{v} = \vec{0}$. The converse is because $||\vec{0}|| = 0$.

3. We have $a\vec{v} = \begin{bmatrix} ax & ay & az \end{bmatrix}^T$ so (1) gives

$$||a\vec{v}||^2 = (ax)^2 + (ay)^2 + (az)^2 = a^2||\vec{v}||^2$$

Hence $||a\vec{v}|| = \sqrt{a^2}||\vec{v}||$, and we are done because $\sqrt{a^2} = |a|$ for any real number a.

Example 4.1.1

If $\vec{v} = \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}$

then $\|\vec{v}\| = \sqrt{4 + 9 + 9} = \sqrt{22}$. Similarly if

$\vec{v} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$

in 2-space then $\|\vec{v}\| = \sqrt{4 + 9} = \sqrt{13}$.

When we view two nonzero vectors as arrows emanating from the origin, it is clear geometrically what we mean by saying that they have the same or opposite **direction**. This leads to a fundamental new description of vectors.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-161

An interactive H5P element has been excluded from this

Theorem 4.1.2

Let $\vec{v} \neq \vec{0}$ and $\vec{w} \neq \vec{0}$ be vectors in $\mathbb{R}^3$. Then $\vec{v} = \vec{w}$ as matrices if and only if $\vec{v}$ and $\vec{w}$ have the same direction and the same length.

Proof:

If $\vec{v} = \vec{w}$, they clearly have the same direction and length. Conversely, let $\vec{v}$ and $\vec{w}$ be vectors with points $P(x, y, z)$ and $Q(x_1, y_1, z_1)$ respectively. If $\vec{v}$ and $\vec{w}$ have the same length and direction then, geometrically, P and Q must be the same point.

Hence $x = x_1$, $y = y_1$, and $z = z_1$, that is
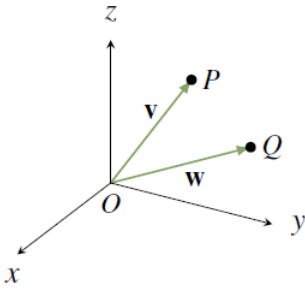
Figure 4.1.3

$$\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \vec{w}.$$

Note that a vector's length and direction do *not* depend on the choice of coordinate system in $\mathbb{R}^3$. Such descriptions are important in applications because physical laws are often stated in terms of vectors, and these laws cannot depend on the particular coordinate system used to describe the situation.

Geometric Vectors

If A and B are distinct points in space, the arrow from A to B has length and direction.
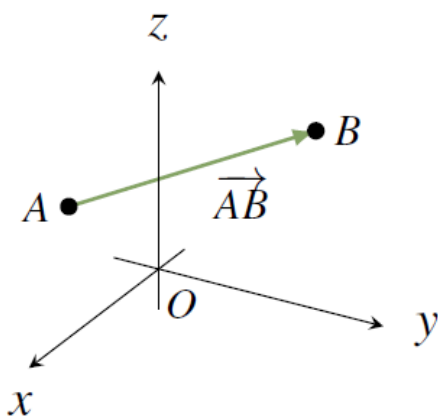
Figure 4.1.4

Hence,

Suppose that A and B are any two points in $\mathbb{R}^3$. In Figure 4.1.4 the line segment from A to B is denoted $\vec{AB}$ and is called the **geometric vector** from A to B. Point A is called the **tail** of AB, B is called the **tip** and the **length** is denoted $||\vec{AB}||$.

Note that if $\vec{v}$ is any vector in $\mathbb{R}^3$ with point P then $\vec{v} = \vec{OP}$ is itself a geometric vector where O is the origin. Referring to $\vec{AB}$ as a "vector" seems justified by Theorem 4.1.2 because it has a direction

(from A to B) and a length $||\vec{AB}||$. However there appears to be a problem because two geometric vectors can have the same length and direction even if the tips and tails are different.
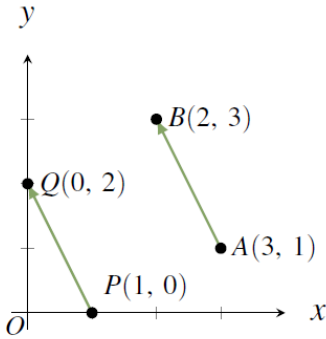
Figure 4.1.5

For example $\vec{AB}$ and $\vec{PQ}$ in Figure 4.1.5 have the same length $\sqrt{5}$ and the same direction (1 unit left and 2 units up) so, by Theorem 4.1.2, they are the same vector! The best way to understand this apparent paradox is to see $\vec{AB}$ and $\vec{PQ}$ as different *representations* of the same underlying vector $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$. Once it is clarified, this phenomenon is a great benefit because, thanks to Theorem 4.1.2, it means that the same geometric vector can be positioned anywhere in space; what is important is the length and direction, not the location of the tip and tail. This ability to move geometric vectors about is very useful.

The Parallelogram Law
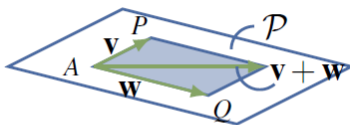
Figure 4.1.6

We now give an intrinsic description of the sum of two vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^3$, that is a description that depends only on the lengths and directions of $\vec{v}$ and $\vec{w}$ and not on the choice of coordinate system. Using Theorem 4.1.2 we can think of these vectors as having a common tail A. If their tips are P and Q

respectively, then they both lie in a plane $\mathcal{P}$ containing A, P, and Q, as shown in Figure 4.1.6. The vectors $\vec{v}$ and $\vec{w}$ create a parallelogram in $\mathcal{P}$, shaded in Figure 4.1.6, called the parallelogram **determined** by $\vec{v}$ and $\vec{w}$.

If we now choose a coordinate system in the plane $\mathcal{P}$ with A as origin, then the parallelogram law in the plane shows that their sum $\vec{v} + \vec{w}$ is the diagonal of the parallelogram they determine with tail A. This is an intrinsic description of the sum $\vec{v} + \vec{w}$ because it makes no reference to coordinates. This discussion proves:

The Parallelogram Law

In the parallelogram determined by two vectors $\vec{v}$ and $\vec{w}$, the vector $\vec{v} + \vec{w}$ is the diagonal with the same tail as $\vec{v}$ and $\vec{w}$.

Figure 4.1.7

Because a vector can be positioned with its tail at any point, the parallelogram law leads to another way to view vector addition. In Figure 4.1.7 (a) the sum $\vec{v} + \vec{w}$ of two vectors $\vec{v}$ and $\vec{w}$ is shown as given by the parallelogram law. If $\vec{w}$ is moved so its tail coincides with the tip of $\vec{v}$ (shown in (b)) then the sum $\vec{v} + \vec{w}$ is seen as "first $\vec{v}$ and then $\vec{w}$. Similarly, moving the tail of $\vec{v}$ to the tip of $\vec{w}$ shows in (c) that $\vec{v} + \vec{w}$ is "first $\vec{w}$ and then $\vec{v}$." This will be referred to as the **tip-to-tail rule**, and it gives a graphic illustration of why $\vec{v} + \vec{w} = \vec{w} + \vec{v}$.

Since $\vec{AB}$ denotes the vector from a point A to a point B, the tip-to-tail rule takes the easily remembered form

$$\vec{AB} + \vec{BC} = \vec{AC}$$

for any points $A, B,$ and C.
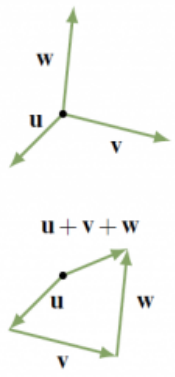
Figure 4.1.8

One reason for the importance of the tip-to-tail rule is that it means two or more vectors can be added by placing them tip-to-tail in sequence. This gives a useful "picture" of the sum of several vectors, and is illustrated for three vectors in Figure 4.1.8 where $\vec{u} + \vec{v} + \vec{w}$ is viewed as first $\vec{u}$, then $\vec{v}$, then $\vec{w}$.
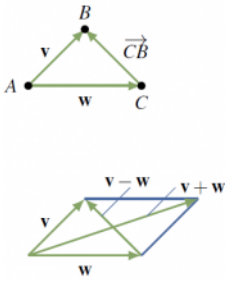
Figure 4.1.9

There is a simple geometrical way to visualize the (matrix) **difference** $\vec{v} - \vec{w}$ of two vectors. If $\vec{v}$ and $\vec{w}$ are positioned so that they have a common tail A , and if B and C are their respective tips, then the tip-to-tail rule gives $\vec{w} + \vec{CB} = \vec{v}$. Hence $\vec{v} - \vec{w} = \vec{CB}$ is the vector from the tip of $\vec{w}$ to the tip of $\vec{v}$. Thus both $\vec{v} - \vec{w}$ and $\vec{v} + \vec{w}$ appear as diagonals in the parallelogram determined by $\vec{v}$ and $\vec{w}$ (see Figure 4.1.9.

Theorem 4.1.3

If $\vec{v}$ and $\vec{w}$ have a common tail, then $\vec{v} - \vec{w}$ is the vector from the tip of $\vec{w}$ to the tip of $\vec{v}$.

One of the most useful applications of vector subtraction is that it gives a simple formula for the vector from one point to another, and for the distance between the points.

Theorem 4.1.4

Let $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ be two points. Then:

1. $\vec{P_1P_2} = \begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{bmatrix}.$

2. The distance between P_1 and P_2 is
$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

Can you prove these results?

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-162

Example 4.1.3

The distance between $P_1(2, -1, 3)$ and $P_2(1, 1, 4)$ is $\sqrt{(-1)^2 + (2)^2 + (1)^2} = \sqrt{6}$, and the vector

from P_1 to P_2 is

$$\vec{P_1P_2} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}.$$

The next theorem tells us what happens to the length and direction of a scalar multiple of a given vector.

Scalar Multiple Law

If a is a real number and $\vec{v} \neq \vec{0}$ is a vector then:

- The length of $a\vec{v}$ is $||a\vec{v}|| = |a|||\vec{v}||$.
- If $a\vec{v} \neq \vec{0}$, the direction of $a\vec{v}$ is the same as $\vec{v}$ if $a > 0$; opposite to $\vec{v}$ if $a < 0$.

Proof:

The first statement is true due to Theorem 4.1.1.

To prove the second statement, let O denote the origin in $\mathbb{R}^3$. Let $\vec{v}$ have point P, and choose any plane containing O and P. If we set up a coordinate system in this plane with O as origin, then $\vec{v} = \vec{OP}$ so the result follows from the scalar multiple law in the plane.

A vector $\vec{u}$ is called a **unit vector** if $||u|| = 1$. Then

$$\vec{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; \vec{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \text{ and } \vec{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

are unit vectors, called the **coordinate** vectors.

Example 4.1.4

If $\vec{v} \neq \vec{0}$ show that $\frac{1}{\|\vec{v}\|}\vec{v}$ is the unique unit vector in the same direction as $\vec{v}$.

Solution:

The vectors in the same direction as $\vec{v}$ are the scalar multiples $a\vec{v}$ where $a > 0$. But $\|a\vec{v}\| = |a|\|\vec{v}\| = a\|\vec{v}\|$ when $a > 0$, so $a\vec{v}$ is a unit vector if and only if $a = \frac{1}{\|v\|}$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=73#h5p-158

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-160

Given points $P(2, -1, 4), Q(3, -1, 3), A(0, 2, 1)$, and $B(1, 3, 0)$, determine if $\vec{PQ}$ and $\vec{AB}$ are parallel.

Solution:

By Theorem 4.1.3, $\vec{PQ} = (1, 0, -1)$ and $\vec{AB} = (1, 1, -1)$. If $\vec{PQ} = t\vec{AB}$ then $(1, 0, -1) = (t, t, -t)$, so $1 = t$ and $0 = t$, which is impossible. Hence $\vec{PQ}$ is *not* a scalar multiple of $\vec{AB}$, so these vectors are not parallel by Theorem 4.1.5.

Lines in Space

These vector techniques can be used to give a very simple way of describing straight lines in space. In order to do this, we first need a way to
specify the orientation of such a line.

Definition 4.3 Direction Vector of a Line

We call a nonzero vector $\vec{d} \neq \vec{0}$ a **direction vector** for the line if it is parallel to $\vec{AB}$ for some pair of distinct points A and B on the line.

Note that any nonzero scalar multiple of $\vec{d}$ would also serve as a direction vector of the line.

We use the fact that there is exactly one line that passes through a particular point $P_0(x_0, y_0, z_0)$ and has a given direction vector $\vec{d} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. We want to describe this line by giving a condition on x, y, and z that the point $P(x, y, z)$ lies on this line. Let $\vec{p_0} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$ and $\vec{p} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ denote the vectors of P_0 and P, respectively.
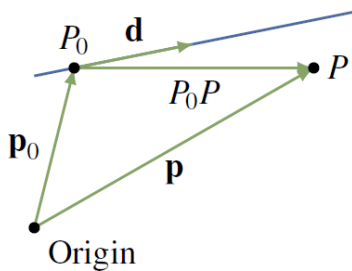
Figure 4.1.10

Then

$$\vec{p} = \vec{p_0} + \vec{P_0P}$$

Hence P lies on the line if and only if $\vec{P_0P}$ is parallel to $\vec{d}$ —that is, if and only if $\vec{P_0P} = t\vec{d}$ for some scalar t by Theorem 4.1.5. Thus $\vec{p}$ is the vector of a point on the line if and only if $\vec{p} = \vec{p_0} + t\vec{d}$ for some scalar t.

Vector Equation of a line

The line parallel to $\vec{d} \neq \vec{0}$ through the point with vector $\vec{p_0}$ is given by

$$\vec{p} = \vec{p_0} + t\vec{d} \quad t \text{ any scalar}$$

In other words, the point P with vector $\vec{p}$ is on this line if and only if a real number t exists such that $\vec{p} = \vec{p_0} + t\vec{d}$.

In component form the vector equation becomes

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + t \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Equating components gives a different description of the line.

Parametric Equations of a line

The line through $P_0(x_0, y_0, z_0)$ with direction vector

$$\vec{d} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \neq \vec{0} \text{ is given by}$$

$$x = x_0 + ta$$
$$y = y_0 + tb \quad t \text{ any scalar}$$
$$z = z_0 + tc$$

In other words, the point $P(x, y, z)$ is on this line if and only if a real number t exists such that $x = x_0 + ta$, $y = y_0 + tb$, and $z = z_0 + tc$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-163

Example 4.1.6

Find the equations of the line through the points $P_0(2, 0, 1)$ and $P_1(4, -1, 1)$.

Solution:

Let

$$\vec{d} = \vec{P_0 P_1} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

denote the vector from P_0 to P_1. Then $\vec{d}$ is parallel to the line (P_0 and P_1 are *on* the line), so $\vec{d}$ serves as a direction vector for the line. Using P_0 as the point on the line leads to the parametric equations

$$\begin{aligned} x &= 2 + 2t \\ y &= -t \qquad t \text{ a parameter} \\ z &= 1 \end{aligned}$$

Note that if P_1 is used (rather than P_0), the equations are

$$\begin{aligned} x &= 4 + 2s \\ y &= -1 - s \qquad s \text{ a parameter} \\ z &= 1 \end{aligned}$$

These are different from the preceding equations, but this is merely the result of a change of parameter. In fact, $s = t - 1$.

Example 4.1.7

Determine whether the following lines intersect and, if so, find the point of intersection.

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

Solution:

Suppose $P(x, y, z)$ with vector $\vec{p}$ lies on both lines. Then

$$\begin{bmatrix} 1 - 3t \\ 2 + 5t \\ 1 + t \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -1 + s \\ 3 - 4s \\ 1 - s \end{bmatrix} \quad \text{for some } t \text{ and } s,$$

where the first (second) equation is because P lies on the first (second) line. Hence the lines intersect if and only if the three equations

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

have a solution. In this case, $t = 1$ and $s = -1$ satisfy all three equations, so the lines *do* intersect and the point of intersection is

$$\vec{p} = \begin{bmatrix} 1 - 3t \\ 2 + 5t \\ 1 + t \end{bmatrix} = \begin{bmatrix} -2 \\ 7 \\ 2 \end{bmatrix}$$

using $t = 1$. Of course, this point can also be found from

$$\vec{p} = \begin{bmatrix} -1 + s \\ 3 - 4s \\ 1 - s \end{bmatrix} \quad \text{using } s = -1.$$

An interactive H5P element has been excluded from this

4.2 Projections and Planes

Suppose a point P and a plane are given and it is desired to find the point Q that lies in the plane and is closest to P, as shown in Figure 4.2.1.
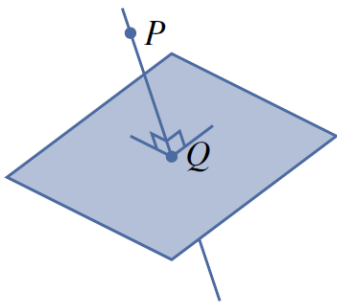
Figure 4.2.1

Clearly, what is required is to find the line through P that is perpendicular to the plane and then to obtain Q as the point of intersection of this line with the plane. Finding the line *perpendicular* to the plane requires a way to determine when two vectors are perpendicular. This can be done using the idea of the dot product of two vectors.

The Dot Product and Angles

Definition 4.4 Dot Product in $\mathbb{R}^3$

Given vectors
$$\vec{v} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \text{ and }$$
$$\vec{w} = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}, \text{ their } \textbf{dot product } \vec{v} \cdot \vec{w} \text{ is a number}$$
defined

$$\vec{v}\vec{w} = x_1 x_2 + y_1 y_2 + z_1 z_2 = \vec{v}^T \vec{w}$$

Because $\vec{v} \cdot \vec{w}$ is a number, it is sometimes called the **scalar product** of $\vec{v}$ and $\vec{w}$.

Example 4.2.1

If $\vec{v} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$

and $\vec{w} = \begin{bmatrix} 1 \\ 4 \\ -1 \end{bmatrix}$, then

$\vec{v} \cdot \vec{w} = 2 \cdot 1 + (-1) \cdot 4 + 3 \cdot (-1) = -5$.

Theorem 4.2.1

Let $\vec{u}, \vec{v}$, and $\vec{w}$ denote vectors in $\mathbb{R}^3$ (or $\mathbb{R}^2$).

1. $\vec{v} \cdot \vec{w}$ is a real number.
2. $\vec{v} \cdot \vec{w} = \vec{w} \cdot \vec{v}$.
3. $\vec{v} \cdot \vec{0} = 0 = \vec{0} \cdot \vec{v}$.

4. $\vec{v} \cdot v = ||\vec{v}||^2$.
5. $(k\vec{v}) \cdot \vec{w} = k(\vec{w} \cdot \vec{v}) = \vec{v} \cdot (k\vec{w})$ for all scalars k.
6. $\vec{u} \cdot (\vec{v} \pm \vec{w}) = \vec{u} \cdot \vec{v} \pm \vec{u} \cdot \vec{w}$.

The readers are invited to prove these properties using the definition of dot products.

Example 4.2.2

Verify that $||\vec{v} - 3\vec{w}||^2 = 1$ when $||\vec{v}|| = 2$, $||\vec{w}|| = 1$, and $\vec{v} \cdot \vec{w} = 2$.

Solution:

We apply Theorem 4.2.1 several times:

$$\|\vec{v} - 3\vec{w}\|^2 = (\vec{v} - 3\vec{w}) \cdot (\vec{v} - 3\vec{w})$$
$$= \vec{v} \cdot (\vec{v} - 3\vec{w}) - 3\vec{w} \cdot (\vec{v} - 3\vec{w})$$
$$= \vec{v} \cdot \vec{v} - 3(\vec{v} \cdot \vec{w}) - 3(\vec{w} \cdot \vec{v}) + 9(\vec{w} \cdot \vec{w})$$
$$= \|\vec{v}\|^2 - 6(\vec{v} \cdot \vec{w}) + 9\|\vec{w}\|^2$$
$$= 4 - 12 + 9 = 1$$

There is an intrinsic description of the dot product of two nonzero vectors in ^{3}. To understand it we require the following result from trigonometry.

Laws of Cosine

If a triangle has sides a, b, and c, and if θ is the interior angle opposite c then

$$c^2 = a^2 + b^2 - 2ab\cos\theta$$
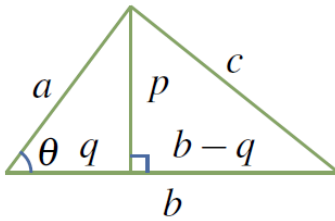
Proof:

We prove it when is θ acute, that is $0 \le \theta < \frac{\pi}{2}$; the obtuse case is similar. In Figure 4.2.2 we have $p = a \sin \theta$ and $q = a \cos \theta$.

Hence Pythagoras' theorem gives

Figure 4.2.2

$$c^2 = p^2 + (b-q)^2 = a^2 \sin^2 \theta + (b - a \cos \theta)^2$$
$$= a^2 (\sin^2 \theta + \cos^2 \theta) + b^2 - 2ab \cos \theta$$

The law of cosines follows because $\sin^2 \theta + \cos^2 \theta = 1$ for any angle θ.

Note that the law of cosines reduces to Pythagoras' theorem if θ is a right angle (because $\cos \frac{\pi}{2} = 0$).

Now let $\vec{v}$ and $\vec{w}$ be nonzero vectors positioned with a common tail. Then they determine a unique angle θ in the range

$$0 \le \theta \le \pi$$

This angle θ will be called the **angle between** $\vec{v}$ and $\vec{w}$. Clearly v and w are parallel if θ is either 0 or π. Note that we do not define the angle between $\vec{v}$ and $\vec{w}$ if one of these vectors is $\vec{0}$.

The next result gives an easy way to compute the angle between two nonzero vectors using the dot product.

Theorem 4.2.2

Let $\vec{v}$ and $\vec{w}$ be nonzero vectors. If θ is the angle between $\vec{v}$ and $\vec{w}$, then

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

Figure 4.2.4

Proof:
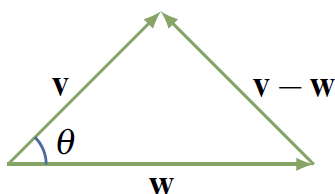We calculate $||\vec{v} - \vec{w}||^2$ in two ways. First apply the law of cosines to the triangle in Figure 4.2.4 to obtain:

$$||\vec{v} - \vec{w}||^2 = ||\vec{v}||^2 + ||\vec{w}||^2 - 2||\vec{v}||\,||\vec{w}|| \cos \theta$$

On the other hand, we use Theorem 4.2.1:

$$||\vec{v} - \vec{w}||^2 = (\vec{v} - \vec{w}) \cdot (\vec{v} - \vec{w})$$
$$= \vec{v} \cdot \vec{v} - \vec{v} \cdot w - \vec{w} \cdot \vec{v} + \vec{w} \cdot \vec{w}$$
$$= ||\vec{v}||^2 - 2(\vec{v} \cdot \vec{w}) + ||\vec{w}||^2$$

Comparing these we see that $-2||\vec{v}||\,||\vec{w}|| \cos \theta = -2(\vec{v} \cdot \vec{w})$, and the result follows.

If $\vec{v}$ and $\vec{w}$ are nonzero vectors, Theorem 4.2.2 gives an intrinsic description of $\vec{v} \cdot \vec{w}$ because $||\vec{v}||$, $||\vec{w}||$, and the angle θ between $\vec{v}$ and $\vec{w}$ do not depend on the choice of coordinate system. Moreover, since $||\vec{v}||$ and $||\vec{w}||$ are nonzero ($\vec{v}$ and $\vec{w}$ are nonzero vectors), it gives a formula for the cosine of the angle θ:

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

Since $0 \leq \theta \leq \pi$, this can be used to find θ.

Compute the angle between
$$\vec{u} = \begin{bmatrix} -1 \\ 1 \\ 2 \\ 2 \end{bmatrix} \text{ and}$$

$$\vec{v} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}.$$

Solution:

Compute $\cos \theta = \frac{\vec{v} \cdot \vec{w}}{||\vec{v}||||\vec{w}||} = \frac{-2+1-2}{\sqrt{6}\sqrt{6}} = -\frac{1}{2}$. Now recall that $\cos \theta$ and $\sin \theta$ are defined so that $(\cos \theta, \sin \theta)$ is the point on the unit circle determined by the angle θ (drawn counterclockwise, starting from the positive x axis). In the present case, we know that $\cos \theta = -\frac{1}{2}$ and that $0 \leq \theta \leq \pi$. Because $\cos \frac{\pi}{3} = \frac{1}{2}$, it follows that $\theta = \frac{2\pi}{3}$.

If $\vec{v}$ and $\vec{w}$ are nonzero, the previous example shows that $\cos \theta$ has the same sign as $v \cdot \vec{w}$, so

$$\vec{v} \cdot \vec{w} > 0 \quad \text{if and only if} \quad \theta \text{ is acute } (0 \leq \theta < \tfrac{\pi}{2})$$
$$\vec{v} \cdot \vec{w} < 0 \quad \text{if and only if} \quad \theta \text{ is obtuse } (\tfrac{\pi}{2} < \theta \leq 0)$$
$$\vec{v} \cdot \vec{w} = 0 \quad \text{if and only if} \quad \theta = \tfrac{\pi}{2}$$

In this last case, the (nonzero) vectors are perpendicular. The following terminology is used in linear algebra:

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-70

Definition 4.5 Orthogonal Vectors in $\mathbb{R}^3$

Two vectors $\vec{v}$ and $\vec{w}$ are said to be \textbf{orthogonal}\index{orthogonal vectors}\index{vectors!orthogonal vectors} if $v = 0$ or $w = 0$ or the angle between them is $\tfrac{\pi}{2}$.

Since $\vec{v} \cdot \vec{w} = 0$ if either $\vec{v} = \vec{0}$ or $\vec{w} = \vec{0}$, we have the following theorem:

Two vectors $\vec{v}$ and $\vec{w}$ are orthogonal if and only if $\vec{v} \cdot \vec{w} = 0$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-72

Example 4.2.4

Show that the points $P(3, -1, 1), Q(4, 1, 4),$ and $R(6, 0, 4)$ are the vertices of a right triangle.

Solution:

The vectors along the sides of the triangle are

$$\vec{PQ} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \ \vec{PR} = \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix}, \ \text{and} \ \vec{QR} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$$

Evidently $\vec{PQ} \cdot \vec{QR} = 2 - 2 + 0 = 0$, so $\vec{PQ}$ and $\vec{QR}$ are orthogonal vectors. This means sides PQ and QR are perpendicular—that is, the angle at Q is a right angle.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-34

Projections

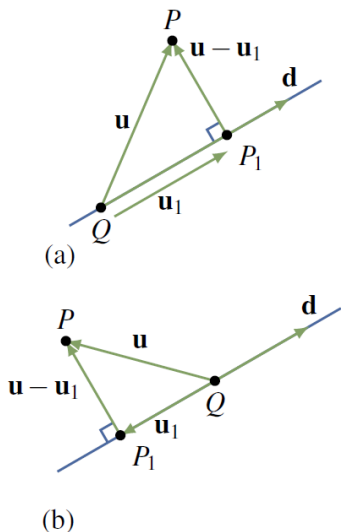In applications of vectors, it is frequently useful to write a vector as the sum of two orthogonal vectors.

If a nonzero vector $\vec{d}$ is specified, the key idea is to be able to write an arbitrary vector $\vec{u}$ as a sum of two vectors,

$$\vec{u} = \vec{u}_1 + \vec{u}_2$$

where $\vec{u}_1$ is parallel to $\vec{d}$ and $\vec{u}_2 = \vec{u} - \vec{u}_1$ is orthogonal to $\vec{d}$. Suppose that $\vec{u}$ and $\vec{d} \neq \vec{0}$ emanate from a common tail Q (see Figure 4.2.5). Let P be the tip of $\vec{u}$, and let P_1 denote the foot of the perpendicular from P to the line through Q parallel to $\vec{d}$.

Then $\vec{u}_1 = \vec{QP_1}$ has the required properties:

1. $\vec{u}_1$ is parallel to $\vec{d}$.
2. $\vec{u}_2 = \vec{u} - \vec{u}_1$ is orthogonal to $\vec{d}$.
3. $\vec{u} = \vec{u}_1 + \vec{u}_2$.

(a)

(b)

Figure 4.2.5

Definition 4.6 Projection in $\mathbb{R}^3$

The vector $\vec{u}_1 = \vec{QP_1}$ in Figure 4.2.6 is called *the projection of $\vec{u}$ on $\vec{d}$.*

It is denoted

$$\vec{u}_1 = proj_{\vec{d}}\vec{u}$$

In Figure 4.2.5 (a) the vector $\vec{u}_1 = proj_{\vec{d}}\vec{u}$ has the same direction as $\vec{d}$; however, $\vec{u}_1$ and $\vec{d}$ have opposite directions if the angle between $\vec{u}$ and $\vec{d}$ is greater than $\frac{\pi}{2}$ (see Figure 4.2.5 (b)). Note that the projection $\vec{u}_1 = proj_{\vec{d}}\vec{u}$ is zero if and only if $\vec{u}$ and $\vec{d}$ are orthogonal.

Calculating the projection of $\vec{u}$ on $\vec{d} \neq \vec{0}$ is remarkably easy.

Theorem 4.2.4

Let $\vec{u}$ and $\vec{d} \neq \vec{0}$ be vectors.

1. The projection of $\vec{u}$ on $\vec{d}$ is given by
    $$proj_{\vec{d}}\vec{u} = \frac{\vec{u}\vec{d}}{||\vec{d}||^2}\vec{d}.$$
2. The vector $\vec{u} - proj_{\vec{d}}\vec{u}$ is orthogonal to $\vec{d}$.

Proof:

The vector $\vec{u}_1 = proj_{\vec{d}}\vec{u}$ is parallel to $\vec{d}$ and so has the form $\vec{u}_1 = t\vec{d}$ for some scalar t. The requirement that $\vec{u} - \vec{u}_1$ and $\vec{d}$ are orthogonal determines t. In fact, it means that $(\vec{u} - \vec{u}_1) \cdot \vec{d} = 0$ by Theorem 4.2.3. If $\vec{u}_1 = t\vec{d}$ is substituted here, the condition is

$$0 = (\vec{u} - t\vec{d}) \cdot \vec{d} = \vec{u} \cdot \vec{d} - t(\vec{d} \cdot \vec{d}) = \vec{u} \cdot \vec{d} - t||\vec{d}||^2$$

It follows that $t = \frac{\vec{u} \cdot \vec{d}}{||\vec{d}||^2}$, where the assumption that $\vec{d} \neq \vec{0}$ guarantees that $||\vec{d}||^2 \neq 0$.

Example 4.2.5

Find the projection of
$$\vec{u} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$$
on $\vec{d} = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$
and express $\vec{u} = \vec{u}_1 + \vec{u}_2$ where $\vec{u}_1$ is parallel to $\vec{d}$ and $\vec{u}_2$ is orthogonal to $\vec{d}$.

Solution:

The projection $\vec{u}_1$ of $\vec{u}$ on $\vec{d}$ is

$$\vec{u}_1 = proj_{\vec{d}}\vec{u} = \frac{\vec{u} \cdot \vec{d}}{||\vec{d}||^2}\vec{d} = \frac{2+3+3}{1^2+(-1)^2+3^2}\begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix} = \frac{8}{11}\begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$$

Hence $\quad \vec{u}_2 = \vec{u} - \vec{u}_1 = \frac{1}{11}\begin{bmatrix} 14 \\ -25 \\ -13 \end{bmatrix}$, and this is

orthogonal to $\vec{d}$ by Theorem 4.2.4 (alternatively, observe that $\vec{d} \cdot \vec{u}_2 = 0$). Since $\vec{u} = \vec{u}_1 + \vec{u}_2$, we are done.

Note that the idea of projections can be used to find the shortest distance from a point to a straight line in $\mathbb{R}^3$ which is $||\vec{u_1}||$, the length of the vector that's orthogonal to the direction vector of the line.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-36

Planes

A nonzero vector $\vec{n}$ is called a **normal** for a plane if it is orthogonal to every vector in the plane.

For example, the unit vector $\vec{k} = (0, 0, 1)$ is a normal vector for $x - y$ plane.
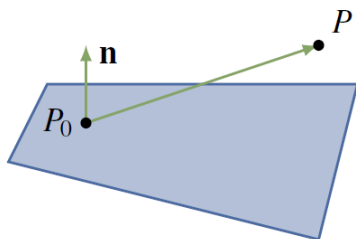
Figure 4.2.6

Given a point $P_0 = P_0(x_0, y_0, z_0)$ and a nonzero vector $\vec{n}$, there is a unique plane through P_0 with normal $\vec{n}$, shaded in Figure 4.2.6. A point $P = P(x, y, z)$ lies on this plane if and only if the vector $\vec{P_0P}$ is orthogonal to $\vec{n}$−that is, if and only if $\vec{n} \cdot \vec{P_0P} = 0$

. Because $\vec{P_0P} = \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix}$ this gives the following result:

Scalar equation of a plane

The plane through $P_0(x_0, y_0, z_0)$ with normal

$$\vec{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \neq \vec{0}$$

as a normal vector is given by

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

In other words, a point $P(x, y, z)$ is on this plane if and only if x, y, and z satisfy this equation.

Example 4.2.8

Find an equation of the plane through $P_0(1, -1, 3)$ with $\vec{n} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$ as normal.

Solution:

Here the general scalar equation becomes

$$3(x - 1) - (y + 1) + 2(z - 3) = 0$$

This simplifies to $3x - y + 2z = 10$.

If we write $d = ax_0 + by_0 + cz_0$, the scalar equation shows that every plane with normal $\vec{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$

has a linear equation of the form

$$(4.2) \quad ax + by + cz = d$$

for some constant d. Conversely, the graph of this equation is a plane with $\vec{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ as a normal vector (assuming that a, b, and c are not all zero).

Example 4.2.9

Find an equation of the plane through $P_0(3, -1, 2)$ that is parallel to the plane with equation $2x - 3y = 6$.

Solution:

The plane with equation $2x - 3y = 6$ has normal $\vec{n} = \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix}$. Because the two planes are parallel, $\vec{n}$ serves as a normal for the plane we seek, so the equation is $2x - 3y = d$ for some d according to (4.2). Insisting that $P_0(3, -1, 2)$ lies on the plane determines d; that is, $d = 2 \cdot 3 - 3(-1) = 9$. Hence, the equation is $2x - 3y = 9$.

Consider points $P_0(x_0, y_0, z_0)$ and $P(x, y, z)$ with vectors

$$\vec{p_0} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$$

and

$$\vec{p} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Given a nonzero vector $\vec{n}$, the scalar equation of the plane through $P_0(x_0, y_0, z_0)$ with normal $\vec{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ takes the vector form:

Vector Equation of a Plane

The plane with normal $\vec{n} \neq \vec{0}$ through the point with vector $\vec{p_0}$ is given by

$$\vec{n} \cdot (\vec{p} - \vec{p_0}) = 0$$

In other words, the point with vector $\vec{p}$ is on the plane if and only if $\vec{p}$ satisfies this condition.
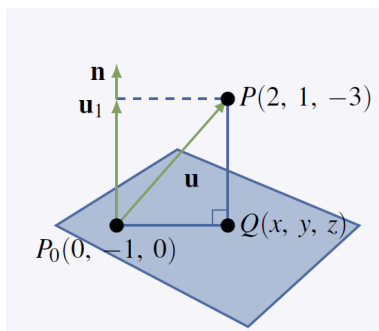
Moreover, Equation (4.2) translates as follows:

Every plane with normal $\vec{n}$ has vector equation $\vec{n} \cdot \vec{p} = d$ for some number d.

Find the shortest distance from the point $P(2, 1, -3)$ to the plane with equation $3x - y + 4z = 1$. Also find the point Q on this plane closest to P.

Solution:

The plane in question has normal $\vec{n} = \begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix}$.

Choose any point P_0 on the plane—say $P_0(0, -1, 0)$ —and let $Q(x, y, z)$ be the point on the plane closest to P (see the diagram). The vector

from P_0 to P is $\vec{u} = \begin{bmatrix} 2 \\ 2 \\ -3 \end{bmatrix}$. Now erect $\vec{n}$ with its tail at P_0.

Then $\vec{QP} = \vec{u}_1$ and $\vec{u}_1$ is the projection of $\vec{u}$ on $\vec{n}$:

$$\vec{u}_1 = \frac{\vec{n} \cdot \vec{u}}{||n||^2}\vec{n} = \frac{-8}{26}\begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix} = \frac{-4}{13}\begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix}$$

Hence the distance is $||\vec{QP}|| = ||\vec{u}_1|| = \frac{4\sqrt{26}}{13}$. To calculate the point Q, let $\vec{q} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

and

$$\vec{p}_0 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

be the vectors of Q and P_0. Then

$$\vec{q} = \vec{p}_0 + \vec{u} - \vec{u}_1 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \\ -3 \end{bmatrix} + \frac{4}{13}\begin{bmatrix} 3 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{38}{13} \\ \frac{9}{13} \\ \frac{-23}{13} \end{bmatrix}$$

This gives the coordinates of $Q\left(\frac{38}{13}, \frac{9}{13}, \frac{-23}{13}\right)$.

The Cross Product

If P, Q, and R are three distinct points in $\mathbb{R}^3$ that are not all on some line, it is clear geometrically that there is a unique plane containing all three. The vectors PQ and PR both lie in this plane, so finding a normal amounts to finding a nonzero vector orthogonal to both PQ and PR. The cross product provides a systematic way to do this.

Definition 4.8 Cross Product

Given vectors $\vec{v}_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$,

define the **cross product** $\vec{v}_1 \times \vec{v}_2$ by

$$\vec{v}_1 \times \vec{v}_2 = \begin{bmatrix} y_1 z_2 - z_1 y_2 \\ -(x_1 z_2 - z_1 x_2) \\ x_1 y_2 - y_1 x_2 \end{bmatrix}$$

Because it is a vector, $\vec{v}_1 \times \vec{v}_2$ is often called the **vector product.** There is an easy way to remember this definition using the **coordinate vectors:**

$$\vec{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \vec{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \ \text{ and } \vec{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

They are vectors of length 1 pointing along the positive x, y, and z axes. The reason for the name is that any vector can be written as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = x\vec{i} + y\vec{j} + z\vec{k}$$

With this, the cross product can be described as follows:

Determinant form of the cross product

If $\vec{v}_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$ are two vectors,

then

$$\vec{v}_1 \times \vec{v}_2 = det \begin{bmatrix} \vec{i} & x_1 & x_2 \\ \vec{j} & y_1 & y_2 \\ \vec{k} & z_1 & z_2 \end{bmatrix} = \begin{vmatrix} y_1 & y_2 \\ z_1 & z_2 \end{vmatrix} \vec{i} - \begin{vmatrix} x_1 & x_2 \\ z_1 & z_2 \end{vmatrix} \vec{j} + \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \vec{k}$$

where the determinant is expanded along the first column.

Example 4.2.11

If $\vec{v} = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix}$, then

$$\vec{v}_1 \times \vec{v}_2 = det \begin{bmatrix} \vec{i} & 2 & 1 \\ \vec{j} & -1 & 3 \\ \vec{k} & 4 & 7 \end{bmatrix} = \begin{vmatrix} -1 & 3 \\ 4 & 7 \end{vmatrix} \vec{i} - \begin{vmatrix} 2 & 1 \\ 4 & 7 \end{vmatrix} \vec{j} + \begin{vmatrix} 2 & 1 \\ -1 & 3 \end{vmatrix} \vec{k}$$

$$= -19\vec{i} - 10\vec{j} + 7\vec{k}$$

$$= \begin{bmatrix} -19 \\ -10 \\ 7 \end{bmatrix}$$

Observe that $\vec{v} \times \vec{w}$ is orthogonal to both $\vec{v}$ and $\vec{w}$ in Example 4.2.11. This holds in general as can be verified directly by computing $\vec{v} \cdot (\vec{v} \times \vec{w})$ and $\vec{w} \cdot (\vec{v} \times \vec{w})$, and is recorded as the first part of the following theorem. It will follow from a more general result which, together with the second part, will be proved later on.

Theorem 4.2.5

Let $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^3$:

1. $\vec{v} \times \vec{w}$ is a vector orthogonal to both $\vec{v}$ and $\vec{w}$.
2.   If $\vec{v}$ and $\vec{w}$ are nonzero, then $\vec{v} \times \vec{w} = \vec{0}$ if and only if $\vec{v}$ and $\vec{w}$ are parallel.

Recall that

$$\vec{v} \cdot \vec{w} = 0 \quad \text{if and only if } \vec{v} \text{ and } \vec{w} \text{ are orthogonal.}$$

Example 4.2.12

Find the equation of the plane through $P(1, 3, -2)$, $Q(1, 1, 5)$, and $R(2, -2, 3)$.

Solution:

The vectors

$$\vec{PQ} = \begin{bmatrix} 0 \\ -2 \\ 7 \end{bmatrix} \text{ and}$$

$$\vec{PR} = \begin{bmatrix} 1 \\ -5 \\ 5 \end{bmatrix}$$

lie in the plane, so

$$\vec{PQ} \times \vec{PR} = det \begin{bmatrix} \vec{i} & 0 & 1 \\ \vec{j} & -2 & -5 \\ \vec{k} & 7 & 5 \end{bmatrix} = 25\vec{i} + 7\vec{j} + 2\vec{k} = \begin{bmatrix} 25 \\ 7 \\ 2 \end{bmatrix}$$

is a normal for the plane (being orthogonal to both $\vec{PQ}$ and $\vec{PR}$). Hence the plane has equation

$$25x + 7y + 2z = d \quad \text{for some number } d.$$

Since $P(1, 3, -2)$ lies in the plane we have $25 \cdot 1 + 7 \cdot 3 + 2(-2) = d$. Hence $d = 42$ and the equation is $25x + 7y + 2z = 42$. Can you verify that he same equation can be obtained if $\vec{QP}$ and $\vec{QR}$, or $\vec{RP}$ and $\vec{RQ}$, are used as the vectors in the plane?

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-35

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-69

4.3 More on the Cross Product

The cross product $\vec{v} \times \vec{w}$ of two $\mathbb{R}^3$ -vectors $\vec{v} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$ was defined in Section 4.2 where we observed that it can be best remembered using a determinant:

(4.3)

$$\vec{v} \times \vec{w} = det \begin{bmatrix} \vec{i} & x_1 & x_2 \\ \vec{j} & y_1 & y_2 \\ \vec{k} & z_1 & z_2 \end{bmatrix} = \begin{vmatrix} y_1 & y_2 \\ z_1 & z_2 \end{vmatrix} \vec{i} - \begin{vmatrix} x_1 & x_2 \\ z_1 & z_2 \end{vmatrix} \vec{j} + \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \vec{k}$$

Here $\vec{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\vec{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and

$\vec{k} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ are the coordinate vectors, and the determinant is expanded along the first column. We observed (but did not prove) in Theorem 4.2.5 that $\vec{v} \times \vec{w}$ is orthogonal to both $\vec{v}$ and $\vec{w}$. This follows easily from the next result.

Theorem 4.3.1

If $\vec{u} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$, and $\vec{w} = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$,

then $\vec{u} \cdot (\vec{v} \times \vec{w}) = det \begin{bmatrix} x_0 & x_1 & x_2 \\ y_0 & y_1 & y_2 \\ z_0 & z_1 & z_2 \end{bmatrix}$.

Proof:
Recall that $\vec{u} \cdot (\vec{v} \times \vec{w})$ is computed by multiplying

corresponding components of $\vec{u}$ and $\vec{v} \times \vec{w}$ and then adding. Using equation (4.3), the result is:

$$\vec{u}\cdot(\vec{v}\times\vec{w}) = x_0 \left(\begin{vmatrix} y_1 & y_2 \\ z_1 & z_2 \end{vmatrix} \right) + y_0 \left(- \begin{vmatrix} x_1 & x_2 \\ z_1 & z_2 \end{vmatrix} \right) + z_0 \left(\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \right) = det \begin{bmatrix} x_0 & x_1 & x_2 \\ y_0 & y_1 & y_2 \\ z_0 & z_1 & z_2 \end{bmatrix}$$

where the last determinant is expanded along column 1.

The result in Theorem 4.3.1 can be succinctly stated as follows: If $\vec{u}$, $\vec{v}$, and $\vec{w}$ are three vectors in $\mathbb{R}^3$, then

$$\vec{u} \cdot (\vec{v} \times \vec{w}) = det \begin{bmatrix} \vec{u} & \vec{v} & \vec{w} \end{bmatrix}$$

where $\begin{bmatrix} \vec{u} & \vec{v} & \vec{w} \end{bmatrix}$ denotes the matrix with $\vec{u}$, $\vec{v}$, and $\vec{w}$ as its columns. Now it is clear that $\vec{v} \times \vec{w}$ is orthogonal to both $\vec{v}$ and $\vec{w}$ because the determinant of a matrix is zero if two columns are identical.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-81

An interactive H5P element has been excluded from this version of the text. You can view it online here:

Because of (4.3) and Theorem 4.3.1, several of the following properties of the cross product follow from
properties of determinants (they can also be verified directly).

Theorem 4.3.2

Let $\vec{u}, \vec{v}$, and $\vec{w}$ denote arbitrary vectors in $\mathbb{R}^3$.

1. $\vec{u} \times \vec{v}$ is a vector.
2. $\vec{u} \times \vec{v}$ is orthogonal to both $\vec{u}$ and $\vec{v}$.
3. $\vec{u} \times \vec{0} = \vec{0} = \vec{0} \times \vec{u}$.
4. $\vec{u} \times \vec{u} = \vec{0}$.
5. $\vec{u} \times \vec{v} = -(\vec{v} \times \vec{u})$.
6. $(k\vec{u}) \times \vec{v} = k(\vec{u} \times \vec{v}) = \vec{u} \times (k\vec{v})$ for any scalar k.
7. $\vec{u} \times (\vec{v} + \vec{w}) = (\vec{u} \times \vec{v}) + (\vec{u} \times \vec{w})$.
8. $(\vec{v} + \vec{w}) \times \vec{u} = (\vec{v} \times \vec{u}) + (\vec{w} \times \vec{u})$.

We have seen some of these results in the past; can you prove 6,7, and 8?

We now come to a fundamental relationship between the dot and cross products.

Theorem 4.3.3 Lagrange Identity

If $\vec{u}$ and $\vec{v}$ are any two vectors in $\mathbb{R}^3$, then

$$||\vec{u} \times \vec{v}|| = ||\vec{u}||^2 ||\vec{v}||^2 - (\vec{u} \cdot \vec{v})^2$$

Proof:

Given $\vec{u}$ and $\vec{v}$, introduce a coordinate system and write

$\vec{u} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$ and

$\vec{v} = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$ in component form. Then all the terms in the identity can be computed in terms of the components.

An expression for the magnitude of the vector $\vec{u} \times \vec{v}$ can be easily obtained from the Lagrange identity. If θ is the angle between

$\vec{u}$ and $\vec{v}$, substituting $\vec{u} \cdot \vec{v} = ||\vec{u}||\,||\vec{v}|| \cos\theta$ into the Lagrange identity gives

$$||\vec{u} \times \vec{v}||^2 = ||\vec{u}||^2||\vec{v}||^2 - ||\vec{u}||^2||\vec{v}||^2 \cos^2\theta = ||\vec{u}||^2||\vec{v}||^2 \sin^2\theta$$

using the fact that $1 - \cos^2\theta = \sin^2\theta$. But $\sin\theta$ is nonnegative on the range $0 \leq \theta \leq \pi$, so taking the positive square root of both sides gives

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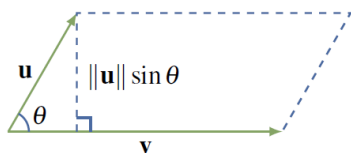

Figure 4.3.1

This expression for $||\vec{u} \times \vec{v}||$ makes no reference to a coordinate system and, moreover, it has a nice geometrical interpretation. The parallelogram determined by the vectors $\vec{u}$ and $\vec{v}$ has base length $||\vec{v}||$ and altitude $||\vec{u}|| \sin\theta$. Hence the area of the parallelogram formed by $\vec{u}$ and $\vec{v}$ is

$$(||\vec{u}|| \sin\theta)||\vec{v}|| = ||\vec{u} \times \vec{v}||$$

Theorem 4.3.4

If $\vec{u}$ and $\vec{v}$ are two nonzero vectors and θ is the angle between $\vec{u}$ and $\vec{v}$, then:

1. $||\vec{u} \times \vec{v}|| = ||\vec{u}||\,||\vec{v}|| \sin\theta =$ the area of the parallelogram determined by $\vec{u}$ and $\vec{v}$.

2. $\vec{u}$ and $\vec{v}$ are parallel if and only if $\vec{u} \times \vec{v} = \vec{0}$.

Proof of 2:

By (1), $\vec{u} \times \vec{v} = \vec{0}$ if and only if the area of the parallelogram is zero. The area vanishes if and only if $\vec{u}$ and $\vec{v}$ have the same or opposite direction—that is, if and only if they are parallel.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-80

Example 4.3.1

Find the area of the triangle with vertices $P(2, 1, 0)$, $Q(3, -1, 1)$, and $R(1, 0, 1)$.

Solution:

We have

$$\vec{RP} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \quad \text{and} \quad \vec{RQ} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}.$$ The area of the triangle is half the area of the parallelogram formed by these vectors, and so equals $\frac{1}{2}||\vec{RP} \times \vec{RQ}||$. We have

$$\vec{RP} \times \vec{RQ} = det \begin{bmatrix} i & 1 & 2 \\ j & 1 & -1 \\ k & -1 & 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix}$$

so the area of the triangle is $\frac{1}{2}||\vec{RP} \times \vec{RQ}|| = \frac{1}{2}\sqrt{1+4+9} = \frac{1}{2}\sqrt{14}$.
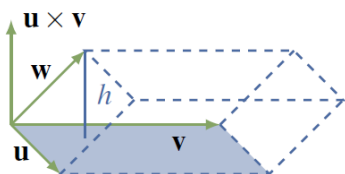

Figure 4.3.2

If three vectors $\vec{u}$, $\vec{v}$, and $\vec{w}$ are given, they determine a "squashed" rectangular solid called a **parallelepiped** (Figure 4.3.2), and it is often useful to be able to find the volume of such a solid. The base of the solid is the parallelogram determined by $\vec{u}$ and $\vec{v}$, so it has area $A = ||\vec{u} \times \vec{v}||$. The height of the solid is the length h of the projection of $\vec{w}$ on $\vec{u} \times \vec{v}$. Hence

$$h = \left| \frac{\vec{w} \cdot (\vec{u} \times \vec{v})}{||\vec{u} \times \vec{v}||^2} \right| ||\vec{u} \times \vec{v}|| = \frac{|\vec{w} \cdot (\vec{u} \times \vec{v})|}{||\vec{u} \times \vec{v}||} = \frac{|\vec{w} \cdot (\vec{u} \times \vec{v})|}{A}$$

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-77

Thus the volume of the parallelepiped is $hA = |\vec{w} \cdot (\vec{u} \times \vec{v})|$. This proves

Theorem 4.3.5

The volume of the parallelepiped determined by three vectors $\vec{w}$, $\vec{u}$, and $\vec{v}$ is given by $|\vec{w} \cdot (\vec{u} \times \vec{v})|$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-76

Example 4.3.2

Find the volume of the parallelepiped determined by the vectors

$$\vec{w} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \vec{u} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \vec{v} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

Solution:

By Theorem 4.3.1,
$$\vec{w} \cdot (\vec{u} \times \vec{v}) = det \begin{bmatrix} 1 & 1 & -2 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} = -3.$$
Hence the volume is $|\vec{w} \cdot (\vec{u} \times \vec{v})| = |-3| = 3$ by Theorem 4.3.5.

We can now give an intrinsic description of the cross product $\vec{u} \times \vec{v}$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=73#h5p-78

Right-hand Rule

If the vector $\vec{u} \times \vec{v}$ is grasped in the right hand and the fingers curl around from $\vec{u}$ to $\vec{v}$ through the angle θ, the thumb points in the direction for $\vec{u} \times \vec{v}$.

To indicate why this is true, introduce coordinates in $\mathbb{R}^3$ as follows: Let $\vec{u}$ and $\vec{v}$ have a common tail O, choose the origin at O, choose the x axis so that $\vec{u}$ points in the positive x direction, and then choose the y axis so that $\vec{v}$ is in the x-y plane and the positive y axis is on the same side of the x axis as $\vec{v}$. Then, in this system, $\vec{u}$ and $\vec{v}$ have component form

$$\vec{u} = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} \text{ and } \vec{v} = \begin{bmatrix} b \\ c \\ 0 \end{bmatrix}$$

where $a > 0$ and $c > 0$. Can you draw a graph based on the description here?

The right-hand rule asserts that $\vec{u} \times \vec{v}$ should point in the positive z direction. But our definition of $\vec{u} \times \vec{v}$ gives

$$\vec{u} \times \vec{v} = det \begin{bmatrix} i & a & b \\ j & 0 & c \\ k & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ ac \end{bmatrix} = (ac)k$$

and $(ac)\vec{k}$ has the positive z direction because $ac > 0$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=73#h5p-79

An interactive H5P element has been excluded from this
version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=73#h5p-74

5. Vector Space [latex size ="40"]\mathbb{R}^n[/latex]

5.1 Subspaces and Spanning

In Section 2.2 we introduced the set $\mathbb{R}^n$ of all n-tuples (called \textit{vectors}), and began our investigation of the matrix transformations $\mathbb{R}^n \rightarrow \mathbb{R}^m$ given by matrix multiplication by an $m \times n$ matrix. Particular attention was paid to the euclidean plane $\mathbb{R}^2$ where certain simple geometric transformations were seen to be matrix transformations.

In this chapter we investigate $\mathbb{R}^n$ in full generality, and introduce some of the most important concepts and methods in linear algebra. The n-tuples in $\mathbb{R}^n$ will continue to be denoted $\vec{x}$, $\vec{y}$, and so on, and will be written as rows or columns depending on the context.

Subspaces of $\mathbb{R}^n$

Definition 5.1 Subspaces of $\mathbb{R}^n$

A set U of vectors in $\mathbb{R}^n$ is called a **subspace** of $\mathbb{R}^n$ if it satisfies the following properties:

S1. The zero vector $\vec{0} \in U$.

S2. If $\vec{x} \in U$ and $\vec{y} \in U$, then $\vec{x} + \vec{y} \in U$.

S3. If $\vec{x} \in U$, then $a\vec{x} \in U$ for every real number a.

We say that the subset U is **closed under addition** if S2 holds, and that U is **closed under scalar multiplication** if S3 holds.

Clearly $\mathbb{R}^n$ is a subspace of itself, and this chapter is about these subspaces and their properties. The set $U = \{\vec{0}\}$, consisting of only the zero vector, is also a subspace because $\vec{0} + \vec{0} = \vec{0}$ and $a\vec{0} = \vec{0}$ for each a in $\mathbb{R}$; it is called the **zero subspace**. Any subspace of $\mathbb{R}^n$ other than $\{\vec{0}\}$ or $\mathbb{R}^n$ is called a **proper** subspace.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=76#h5p-85
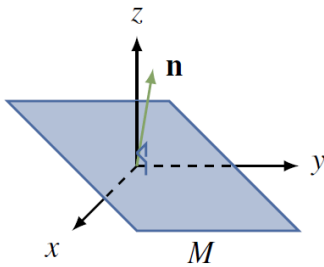
We saw in Section 4.2 that every plane M through the origin in $\mathbb{R}^3$ has equation $ax + by + cz = 0$ where a, b, and c are not all zero.

Here $\vec{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ is a normal for the plane and

$$M = \{\vec{v} \text{ in } \mathbb{R}^3 \mid \vec{n} \cdot \vec{v} = 0\}$$

where

$\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and $\vec{n} \cdot \vec{v}$ denotes the dot product introduced in Section 2.2 (see the diagram). Then M is a subspace of ^{3}. Indeed we show that M satisfies S1, S2, and S3 as follows:

S1. $\vec{0} \in M$ *because* $\vec{n} \cdot \vec{0} = 0$;

S2. If $\vec{v} \in M$ and $\vec{v}_1 \in M$, then $\vec{n} \cdot (\vec{v} + \vec{v}_1) = \vec{n}\vec{v} + \vec{n} \cdot \vec{v}_1 = 0 + 0 = 0$, so $\vec{v} + \vec{v}_1 \in M$;

S3. If $\vec{v} \in M$, then $\vec{n} \cdot (a\vec{v}) = a(\vec{n} \cdot \vec{v}) = a(0) = 0$, so $a\vec{v} \in M$.

Example 5.1.1

Planes and lines through the origin in $\mathbb{R}^3$ are all subspaces of $\mathbb{R}^3$.

Solution:

We proved the statement for planes above. If L is a line through the origin with direction vector $\vec{d}$, then $L = \{t\vec{d} \mid t \in \mathbb{R}\}$. Can you verify that L satisfies S1, S2, and S3?

Example 5.1.1 shows that lines through the origin in $\mathbb{R}^2$ are subspaces; in fact, they are the only proper subspaces of $\mathbb{R}^2$. Indeed, we will prove that lines and planes through the origin in $\mathbb{R}^3$ are the only proper subspaces of $\mathbb{R}^3$. Thus the geometry of lines and planes through the origin is captured by the subspace concept. (Note that every line or plane is just a translation of one of these.)

Subspaces can also be used to describe important features of an $m \times n$ matrix A. The **null space** of A, denoted $null\, A$, and the **image space** of A, denoted $im\, A$, are defined by

$$null\, A = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}\} \quad \text{and} \quad im\, A = \{A\vec{x} \mid \vec{x} \in \mathbb{R}^n\}$$

In the language of Chapter 2, $null\, A$ consists of all solutions $\vec{x}$ in $\mathbb{R}^n$ of the homogeneous system $A\vec{x} = \vec{0}$, and $im\, A$ is the set of all vectors $\vec{y}$ in $\mathbb{R}^m$ such that $A\vec{x} = \vec{y}$ has a solution $\vec{x}$. Note that $\vec{x}$ is in $null\, A$ if it satisfies the *condition* $A\vec{x} = \vec{0}$, while $im\, A$ consists of vectors of the form $A\vec{x}$ for some $\vec{x}$ in $\mathbb{R}^n$. These two ways to describe subsets occur frequently.

Example 5.1.2

If A is an $m \times n$ matrix, then:

1. $null\,A$ is a subspace of $\mathbb{R}^n$.
2. $im\,A$ is a subspace of $\mathbb{R}^m$.

Solution:

1. The zero vector $\vec{0} \in \mathbb{R}^n$ lies in $null\,A$ because $A\vec{0} = \vec{0}$. If $\vec{x}$ and $\vec{x}_1$ are in $null\,A$, then $\vec{x} + \vec{x}_1$ and $a\vec{x}$ are in $null\,A$ because they satisfy the required condition:

$$A(\vec{x} + \vec{x}_1) = A\vec{x} + A\vec{x}_1 = \vec{0} + \vec{0} = \vec{0} \quad \text{and} \quad A(a\vec{x}) = a(A\vec{x}) = a\vec{0} = \vec{0}$$

   Hence $null\,A$ satisfies S1, S2, and S3, and so is a subspace of $\mathbb{R}^n$.

2. The zero vector $\vec{0} \in \mathbb{R}^m$ lies in $im\,A$ because $\vec{0} = A\vec{0}$. Suppose that $\vec{y}$ and $\vec{y}_1$ are in $im\,A$, say $\vec{y} = A\vec{x}$ and $\vec{y}_1 = A\vec{x}_1$ where $\vec{x}$ and $\vec{x}_1$ are in $\mathbb{R}^n$. Then

$$\vec{y} + \vec{y}_1 = A\vec{x} + A\vec{x}_1 = A(\vec{x} + \vec{x}_1) \quad \text{and} \quad a\vec{y} = a(A\vec{x}) = A(a\vec{x})$$

   show that $\vec{y} + \vec{y}_1$ and $a\vec{y}$ are both in $im\,A$ (they have the required form). Hence $im\,A$ is a subspace of $\mathbb{R}^m$.

There are other important subspaces associated with a matrix A that clarify basic properties of A. If A is an $n \times n$ matrix and λ is any number, let

$$E_\lambda(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \lambda\vec{x}\}$$

A vector $\vec{x}$ is in $E_\lambda(A)$ if and only if $(\lambda I - A)\vec{x} = \vec{0}$, so Example 5.1.2 gives:

$E_\lambda(A) = null(\lambda I - A)$ is a subspace of $\mathbb{R}^n$ for each $n \times n$ matrix A and number λ.

$E_\lambda(A)$ is called the **eigenspace** of A corresponding to λ. The reason for the name is that, in the terminology of Section 3.3, λ is an **eigenvalue** of A if $E_\lambda(A) \neq \{\vec{0}\}$. In this case the nonzero vectors in $E_\lambda(A)$ are called the **eigenvectors** of A corresponding to λ.

The reader should not get the impression that every subset of $\mathbb{R}^n$ is a subspace. For example:

$$U_1 = \left\{ \left[\begin{array}{c} x \\ y \end{array} \right] \,\middle|\, x \geq 0 \right\} \quad \text{satisfies S1 and S2, but not S3;}$$

$$U_2 = \left\{ \left[\begin{array}{c} x \\ y \end{array} \right] \,\middle|\, x^2 = y^2 \right\} \quad \text{satisfies S1 and S3, but not S2;}$$

Hence neither U_1 nor U_2 is a subspace of ^{2}.

Spanning sets

Let $\vec{v}$ and $\vec{w}$ be two nonzero, nonparallel vectors in $\mathbb{R}^3$ with their tails at the origin. The plane M through the origin containing these vectors is described in Section 4.2 by saying that $\vec{n} = \vec{v} \times \vec{w}$ is a *normal* for M, and that M consists of all vectors $\vec{p}$ such that $\vec{n} \cdot \vec{p} = 0$.

While this is a very useful way to look at planes, there is another approach that is at least as useful in $\mathbb{R}^3$ and, more importantly, works for all subspaces of $\mathbb{R}^n$ for any $n \geq 1$.
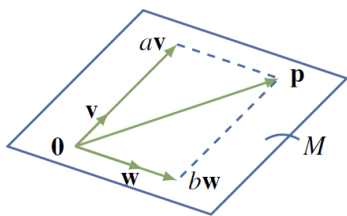
The idea is as follows: Observe that, by the diagram, a vector $\vec{p}$ is in M if and only if it has the form

$$\vec{p} = a\vec{v} + b\vec{w}$$

for certain real numbers a and b (we say that $\vec{p}$ is a *linear combination* of $\vec{v}$ and $\vec{w}$).

Hence we can describe M as

$$M = \{a\vec{x} + b\vec{w} \mid a, b \in \mathbb{R}\}$$

and we say that $\{\vec{v}, \vec{w}\}$ is a *spanning set* for M. It is this notion

of a spanning set that provides a way to describe all subspaces of $\mathbb{R}^n$.

As in Section 1.3, given vectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$ in $\mathbb{R}^n$, a vector of the form

$$t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k \quad \text{where the } t_i \text{ are scalars}$$

is called a **linear combination** of the $\vec{x}_i$, and t_i is called the **coefficient** of $\vec{x}_i$ in the linear combination.

Definition 5.2 Linear Combinations and Span in $\mathbb{R}^n$

The set of all such linear combinations is called the **span** of the $\vec{x}_i$ and is denoted

$$span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\} = \{t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k \mid t_i \text{ in } \mathbb{R}\}$$

If $V = span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$, we say that V is **spanned** by the vectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$, and that the vectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$ **span** the space V.

Here are two examples:
$$span\{\vec{x}\} = \{t\vec{x} \mid t \in \mathbb{R}\}$$
which we write as $span\{\vec{x}\} = \mathbb{R}\vec{x}$ for simplicity.
$$span\{\vec{x}, \vec{y}\} = \{r\vec{x} + s\vec{y} \mid r, s \in \mathbb{R}\}$$

In particular, the above discussion shows that, if $\vec{v}$ and $\vec{w}$ are two nonzero, nonparallel vectors in $\mathbb{R}^3$, then

$$M = span\{\vec{v}, \vec{w}\}$$

is the plane in $\mathbb{R}^3$ containing $\vec{v}$ and $\vec{w}$. Moreover, if $\vec{d}$ is any nonzero vector in $\mathbb{R}^3$ (or $\mathbb{R}^2$), then

$$L = span\{\vec{v}\} = \{t\vec{d} \mid t \in \mathbb{R}\} = \mathbb{R}\vec{d}$$

is the line with direction vector $\vec{d}$. Hence lines and planes can both be described in terms of spanning sets.

Let $\vec{x} = (2, -1, 2, 1)$ and $\vec{y} = (3, 4, -1, 1)$ in $\mathbb{R}^4$. Determine whether $\vec{p} = (0, -11, 8, 1)$ or $\vec{q} = (2, 3, 1, 2)$ are in $U = span\{\vec{x}, \vec{y}\}$.

Solution:

The vector $\vec{p}$ is in U if and only if $\vec{p} = s\vec{x} + t\vec{y}$ for scalars s and t. Equating components gives equations

$$2s + 3t = 0, \quad -s + 4t = -11, \quad 2s - t = 8, \quad \text{and} \quad s + t = 1$$

This linear system has solution $s = 3$ and $t = -2$, so $\vec{p}$ is in U. On the other hand, asking that $\vec{q} = s\vec{x} + t\vec{y}$ leads to equations

$$2s + 3t = 2, \quad -s + 4t = 3, \quad 2s - t = 1, \quad \text{and} \quad s + t = 2$$

and this system has no solution. So $\vec{q}$ does not lie in U.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-87

Let $U = span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ in $\mathbb{R}^n$. Then:

1. U is a subspace of $\mathbb{R}^n$ containing each $\vec{x}_i$.
2. If W is a subspace of $\mathbb{R}^n$ and each $\vec{x}_i \in W$, then $U \subseteq W$.

Proof:

1. The zero vector $\vec{0}$ is in U because $\vec{0} = 0\vec{x}_1 + 0\vec{x}_2 + \cdots + 0\vec{x}_k$ is a linear combination of the $\vec{x}_i$. If $\vec{x} = t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k$ and $\vec{y} = s_1\vec{x}_1 + s_2\vec{x}_2 + \cdots + s_k\vec{x}_k$ are in U, then $\vec{x} + \vec{y}$ and $a\vec{x}$ are in U because

$$\begin{aligned} \vec{x} + \vec{y} &= (t_1 + s_1)\vec{x}_1 + (t_2 + s_2)\vec{x}_2 + \cdots + (t_k + s_k)\vec{x}_k, \text{ and} \\ a\vec{x} &= (at_1)\vec{x}_1 + (at_2)\vec{x}_2 + \cdots + (at_k)\vec{x}_k \end{aligned}$$

   Finally each $\vec{x}_i$ is in U (for example, $\vec{x}_2 = 0\vec{x}_1 + 1\vec{x}_2 + \cdots + 0\vec{x}_k$) so S1, S2, and S3 are satisfied for U, proving (1).
2. Let $\vec{x} = t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k$ where the t_i are scalars and each $\vec{x}_i \in W$. Then each $t_i\vec{x}_i \in W$ because W satisfies S3. But then $\vec{x} \in W$ because W satisfies S2 (verify). This proves (2).

Condition (2) in Theorem 5.1.1 can be expressed by saying that $span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ is the *smallest* subspace of $\mathbb{R}^n$ that contains each $\vec{x}_i$. This is useful for showing that two subspaces U and W are equal, since this amounts to showing that both $U \subseteq W$ and $W \subseteq U$. Here is an example of how it is used.

Example 5.1.5

If $\vec{x}$ and $\vec{y}$ are in $\mathbb{R}^n$, show that
$$span\{\vec{x}, \vec{y}\} = span\{\vec{x} + \vec{y}, \vec{x} - \vec{y}\}.$$

Solution:

Since both $\vec{x} + \vec{y}$ and $\vec{x} - \vec{y}$ are in $span\{\vec{x}, \vec{y}\}$, Theorem 5.1.1 gives
$$span\{\vec{x} + \vec{y}, \vec{x} - \vec{y}\} \subseteq span\{\vec{x}, \vec{y}\}$$
But $\qquad \vec{x} = \frac{1}{2}(\vec{x} + \vec{y}) + \frac{1}{2}(\vec{x} - \vec{y}) \qquad$ and $\vec{y} = \frac{1}{2}(\vec{x} + \vec{y}) - \frac{1}{2}(\vec{x} - \vec{y}) \qquad$ are both in $span\{\vec{x} + \vec{y}, \vec{x} - \vec{y}\}$, so
$$span\{\vec{x}, \vec{y}\} \subseteq span\{\vec{x} + \vec{y}, \vec{x} - \vec{y}\}$$

again   by   Theorem   5.1.1.   Thus
$span\{\vec{x}, \vec{y}\} = span\{\vec{x} + \vec{y}, \vec{x} - \vec{y}\}$, as desired.

It turns out that many important subspaces are best described by giving a spanning set. Here are three examples, beginning with an important spanning set for $\mathbb{R}^n$ itself. Column j of the $n \times n$ identity matrix I_n is denoted $\vec{e}_j$ and called the jth **coordinate vector** in $\mathbb{R}^n$, and the set $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ is called the **standard basis** of $\mathbb{R}^n$. If

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{is any vector in} \quad \mathbb{R}^n, \quad \text{then}$$

$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n$, as the reader can verify. This proves:

Example 5.1.6

$\mathbb{R}^n = span\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ where $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ are the columns of I_n.

If A is an $m \times n$ matrix A, the next two examples show that it is a routine matter to find spanning sets for $null\,A$ and $im\,A$.

Example 5.1.7

Given an $m \times n$ matrix A, let $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$ denote the basic solutions to the system $A\vec{x} = \vec{0}$ given by the gaussian algorithm. Then

$$null\,A = span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$$

Solution:

If $\vec{x} \in null\,A$, then $A\vec{x} = \vec{0}$ so Theorem 1.3.2 shows that $\vec{x}$ is a linear combination of the basic solutions; that is, $null\,A \subseteq span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$. On the other hand, if $\vec{x}$ is in $span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$, then $\vec{x} = t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k$ for scalars t_i, so

$$A\vec{x} = t_1 A\vec{x}_1 + t_2 A\vec{x}_2 + \cdots + t_k A\vec{x}_k = t_1\vec{0} + t_2\vec{0} + \cdots + t_k\vec{0} = \vec{0}$$

This shows that $\vec{x} \in null\,A$, and hence that $span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\} \subseteq null\,A$. Thus we have equality.

Example 5.1.8

Let $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n$ denote the columns of the $m \times n$ matrix A. Then

$$im\,A = span\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$$

Solution:

If $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ is the standard basis of $\mathbb{R}^n$, observe that

$$\begin{bmatrix} A\vec{e}_1 & A\vec{e}_2 & \cdots & A\vec{e}_n \end{bmatrix} = A \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \end{bmatrix} = AI_n = A = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots \vec{c}_n \end{bmatrix}.$$

Hence $\vec{c}_i = A\vec{e}_i$ is in $im\,A$ for each i, so $span\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\} \subseteq im\,A$.

Conversely, let $\vec{y}$ be in $im\,A$, say $\vec{y} = A\vec{x}$ for some $\vec{x}$ in $\mathbb{R}^n$. If

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \text{ then Definition 2.5 gives}$$

$\vec{y} = A\vec{x} = x_1\vec{c}_1 + x_2\vec{c}_2 + \cdots + x_n\vec{c}_n$ is in $span\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$

This shows that $im\,A \subseteq span\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$, and the result follows.

https://www.geogebra.org/m/Q4hT3V5N

Please answer these questions after you open the webpage:

1. Set $\vec{OA} = (1, 0)$ and $\vec{OB} = (0, 1)$. Set $P = (4, 5)$.

2. Click "start animation" to see which linear combination of $\vec{OA}$ and $\vec{OB}$ will produce the vector $\vec{OP}$.

3. Change $\vec{OA} = (-1, 2)$ and $\vec{OB} = (2, 1)$. Randomly choose a point P.

4. Click "start animation" to see which linear combination of $\vec{OA}$ and $\vec{OB}$ will produce the vector $\vec{OP}$. Write it down.

5. What if we set $\vec{OA} = (-1, 2)$ and $\vec{OB} = (2, -1)$? Can you explain what's happening now? Would you still be able to find a linear combination of $\vec{OA}$ and $\vec{OB}$ to produce a vector $\vec{OP}$?

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/
linearalgebrautm/?p=76#h5p-89

5.2 Independence and Dimension

Some spanning sets are better than others. If $U = span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ is a subspace of $\mathbb{R}^n$, then every vector in U can be written as a linear combination of the $\vec{x}_i$ in at least one way. Our interest here is in spanning sets where each vector in U has *exactly one* representation as a linear combination of these vectors.

Linear Independence

Given $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$ in $\mathbb{R}^n$, suppose that two linear combinations are equal:

$$r_1\vec{x}_1 + r_2\vec{x}_2 + \cdots + r_k\vec{x}_k = s_1\vec{x}_1 + s_2\vec{x}_2 + \cdots + s_k\vec{x}_k$$

We are looking for a condition on the set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ of vectors that guarantees that this representation is *unique*; that is, $r_i = s_i$ for each i. Taking all terms to the left side gives

$$(r_1 - s_1)\vec{x}_1 + (r_2 - s_2)\vec{x}_2 + \cdots + (r_k - s_k)\vec{x}_k = \vec{0}$$

so the required condition is that this equation forces all the coefficients $r_i - s_i$ to be zero.

Definition 5.3 Linear Independence in $\mathbb{R}^n$

We call a set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ of vectors **linearly independent** if it satisfies the following condition:

If $t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k = \vec{0}$ then $t_1 = t_2 = \cdots = t_k = 0$

Theorem 5.2.1

If $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ is an independent set of vectors in $\mathbb{R}^n$, then every vector in $span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ has a **unique** representation as a linear combination of the $\vec{x}_i$.

It is useful to state the definition of independence in different language. Let us say that a linear combination **vanishes** if it equals the zero vector, and
call a linear combination **trivial** if every coefficient is zero. Then the definition of independence can be compactly stated as follows:

A set of vectors is independent if and only if the only linear combination that vanishes is the trivial one.

Hence we have a procedure for checking that a set of vectors is independent:

To verify that a set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ of vectors in $\mathbb{R}^n$ is independent, proceed as follows:

1. Set a linear combination equal to zero:
   $$t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k = \vec{0}.$$
2. Show that $t_i = 0$ for each i (that is, the linear combination is trivial).

Of course, if some nontrivial linear combination vanishes, the vectors are not independent.

Example 5.2.1

Determine whether
$$\{(1,0,-2,5), (2,1,0,-1), (1,1,2,1)\} \text{ is}$$
independent in $\mathbb{R}^4$.

Solution:

Suppose a linear combination vanishes:

$$r(1,0,-2,5) + s(2,1,0,-1) + t(1,1,2,1) = (0,0,0,0)$$

Equating corresponding entries gives a system of four equations:

$$r + 2s + t = 0, s + t = 0, -2r + 2t = 0, \text{ and } 5r - s + t = 0$$

The only solution is the trivial one $r = s = t = 0$ (please verify), so these vectors are independent by the independence test.

Solution:

The components of $t_1\vec{e}_1 + t_2\vec{e}_2 + \cdots + t_n\vec{e}_n$ are $t_1, t_2, \ldots, t_n$. So the linear combination vanishes if and only if each $t_i = 0$. Hence the independence test applies.

Solution:

If $s(2\vec{x} + 3\vec{y}) + t(\vec{x} - 5\vec{y}) = \vec{0}$, collect terms to get $(2s + t)\vec{x} + (3s - 5t)\vec{y} = \vec{0}$. Since $\{\vec{x}, \vec{y}\}$ is independent this combination must be trivial; that is, $2s + t = 0$ and $3s - 5t = 0$. These equations have only the trivial solution $s = t = 0$, as required.

Example 5.2.4

Show that the zero vector in $\mathbb{R}^n$ does not belong to any independent set.

Solution:

No set $\left\{\vec{0}, \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\right\}$ of vectors is independent because we have a vanishing, nontrivial linear combination $1 \cdot \vec{0} + 0\vec{x}_1 + 0\vec{x}_2 + \cdots + 0\vec{x}_k = \vec{0}$.

Example 5.2.5

Given $\vec{x}$ in $\mathbb{R}^n$, show that $\{\vec{x}\}$ is independent if and only if $\vec{x} \neq \vec{0}$.

Solution:

A vanishing linear combination from $\{\vec{x}\}$ takes the form $t\vec{x} = \vec{0}$, t in $\mathbb{R}$. This implies that $t = 0$ because $\vec{x} \neq \vec{0}$.

Show that the nonzero rows of a row-echelon matrix R are independent.

Solution:

We illustrate the case with 3 leading 1s; the general case is analogous. Suppose R has the form

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

where $*$ indicates a nonspecified number. Let R_1, R_2, and R_3 denote the nonzero rows of R. If $t_1 R_1 + t_2 R_2 + t_3 R_3 = 0$ we show that $t_1 = 0$, then $t_2 = 0$, and finally $t_3 = 0$. The condition $t_1 R_1 + t_2 R_2 + t_3 R_3 = 0$ becomes

$$(0, t_1, *, *, *, *) + (0, 0, 0, t_2, *, *) + (0, 0, 0, 0, t_3, *) = (0, 0, 0, 0, 0, 0)$$

Equating second entries show that $t_1 = 0$, so the condition becomes $t_2 R_2 + t_3 R_3 = 0$. Now the same argument shows that $t_2 = 0$. Finally, this gives $t_3 R_3 = 0$ and we obtain $t_3 = 0$.

A set of vectors in $\mathbb{R}^n$ is called **linearly dependent** (or simply **dependent**) if it is *not* linearly independent, equivalently if some nontrivial linear combination vanishes.

Example 5.2.7

If $\vec{v}$ and $\vec{w}$ are nonzero vectors in $\mathbb{R}^3$, show that $\{\vec{v}, \vec{w}\}$ is dependent if and only if $\vec{v}$ and $\vec{w}$ are parallel.
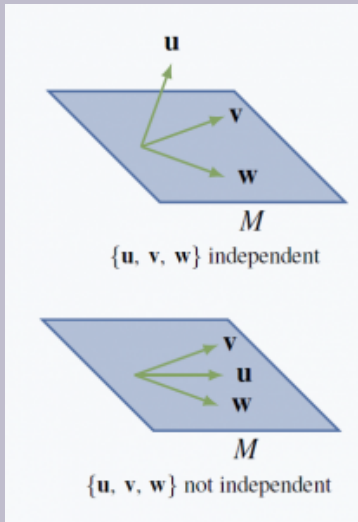
Solution:

If $\vec{v}$ and $\vec{w}$ are parallel, then one is a scalar multiple of the other, say $\vec{v} = a\vec{w}$ for some scalar a. Then the nontrivial linear combination $\vec{v} - a\vec{w} = \vec{0}$ vanishes, so $\{\vec{v}, \vec{w}\}$ is dependent.

Conversely, if $\{\vec{v}, \vec{w}\}$ is dependent, let $s\vec{v} + t\vec{w} = \vec{0}$ be nontrivial, say $s \neq 0$. Then $\vec{v} = -\frac{t}{s}\vec{w}$ so $\vec{v}$ and $\vec{w}$ are parallel. A similar argument works if $t \neq 0$.

With this we can give a geometric description of what it means for a set $\{\vec{u}, \vec{v}, \vec{w}\}$ in $\mathbb{R}^3$ to be independent. Note that this requirement means that $\{\vec{v}, \vec{w}\}$ is also independent ($a\vec{v} + b\vec{w} = \vec{0}$ means that $0\vec{u} + a\vec{v} + b\vec{w} = \vec{0}$), so $M = span\{\vec{v}, \vec{w}\}$ is the plane containing $\vec{v}, \vec{w}$, and $\vec{0}$ (see the discussion preceding Example 5.1.4). So we assume that $\{\vec{v}, \vec{w}\}$ is independent in the following example.

Let $\vec{u}$, $\vec{v}$, and $\vec{w}$ be nonzero vectors in $\mathbb{R}^3$ where $\{\vec{v}, \vec{w}\}$ independent. Show that $\{\vec{u}, \vec{v}, \vec{w}\}$ is independent if and only if $\vec{u}$ is not in the plane $M = span\{\vec{v}, \vec{w}\}$. This is illustrated in the diagrams.

Solution:

If $\{\vec{u}, \vec{v}, \vec{w}\}$ is independent, suppose $\vec{u}$ is in the plane $M = span\{\vec{v}, \vec{w}\}$, say $\vec{u} = a\vec{v} + b\vec{w}$, where a and b are in $\mathbb{R}$. Then $1\vec{u} - a\vec{v} - b\vec{w} = \vec{0}$, contradicting the independence of $\{\vec{u}, \vec{v}, \vec{w}\}$.

On the other hand, suppose that $\vec{u}$ is not in M; we must show that $\{\vec{u}, \vec{v}, \vec{w}\}$ is independent. If $r\vec{u} + s\vec{v} + t\vec{w} = \vec{0}$ where r, s, and t are in $\mathbb{R}$, then $r = 0$ since otherwise $\vec{u} = -\frac{s}{r}\vec{v} + \frac{-t}{r}\vec{w}$ is in M. But then $s\vec{v} + t\vec{w} = \vec{0}$, so $s = t = 0$ by our assumption. This shows that $\{\vec{u}, \vec{v}, \vec{w}\}$ is independent, as required.

By Theorem 2.4.5, the following conditions are equivalent for an $n \times n$ matrix A:

1. A is invertible.
2. If $A\vec{x} = \vec{0}$ where $\vec{x}$ is in $\mathbb{R}^n$, then $\vec{x} = \vec{0}$.
3. $A\vec{x} = \vec{b}$ has a solution $\vec{x}$ for every vector $\vec{b}$ in $\mathbb{R}^n$.

While condition 1 makes no sense if A is not square, conditions 2 and 3 are meaningful for any matrix A and, in fact, are related to independence and spanning. Indeed, if $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n$ are the columns of A, and if we write

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \text{ then}$$

$$A\vec{x} = x_1\vec{c}_1 + x_2\vec{c}_2 + \cdots + x_n\vec{c}_n$$

by Definition 2.5. Hence the definitions of independence and spanning show, respectively, that condition 2 is equivalent to the independence of $\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$ and condition 3 is equivalent to the requirement that $span\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\} = \mathbb{R}^m$. This discussion is summarized in the following theorem:

Theorem 5.2.2

If A is an $m \times n$ matrix, let $\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$ denote the columns of A.

1. $\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$ is independent in $\mathbb{R}^m$ if and only if $A\vec{x} = \vec{0}, \vec{x}$ in $\mathbb{R}^n$, implies $\vec{x} = \vec{0}$.

2. $\mathbb{R}^m = span\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$ if and only if $A\vec{x} = \vec{b}$ has a solution $\vec{x}$ for every vector $\vec{b}$ in $\mathbb{R}^m$
.

For a square matrix A, Theorem 5.2.2 characterizes the invertibility of A in terms of the spanning and independence of its columns (see the discussion preceding Theorem 5.2.2). It is important to be able to discuss these notions for rows. If $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$ are $1 \times n$ rows, we define $span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ to be the set of all linear combinations of the $\vec{x}_i$ (as matrices), and we say that $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ is linearly independent if the only vanishing linear combination is the trivial one (that is, if $\{\vec{x}_1^T, \vec{x}_2^T, \ldots, \vec{x}_k^T\}$ is independent in $\mathbb{R}^n$, as the reader can verify).

Theorem 5.2.3

The following are equivalent for an $n \times n$ matrix A:

1. A is invertible.
2. The columns of A are linearly independent.
3. The columns of A span $\mathbb{R}^n$.
4. The rows of A are linearly independent.
5. The rows of A span the set of all $1 \times n$ rows.

Proof:

Let $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n$ denote the columns of A.

(1) $\Longleftrightarrow$ (2). By Theorem 2.4.5, A is invertible if and only if $A\vec{x} = \vec{0}$ implies $\vec{x} = \vec{0}$; this holds if and only if $\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$ is independent by Theorem 5.2.2.

(1) $\Longleftrightarrow$ (3). Again by Theorem 2.4.5, A is invertible if and only if $A\vec{x} = \vec{b}$ has a solution for every column B in $\mathbb{R}^n$; this holds if and only if $span\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\} = \mathbb{R}^n$ by Theorem 5.2.2.

(1) $\Longleftrightarrow$ (4). The matrix A is invertible if and only if A^T is invertible (by Corollary 2.4.1 to Theorem 2.2.4); this in turn holds if and only if A^T has independent columns (by (1) $\Longleftrightarrow$ (2)); finally, this last statement holds if and only if A has independent rows (because the rows of A are the transposes of the columns of A^T).

(1) $\Longleftrightarrow$ (5). The proof is similar to (1) $\Longleftrightarrow$ (4).

Example 5.2.9

Show that
$S = \{(2, -2, 5), (-3, 1, 1), (2, 7, -4)\}$ is independent in $\mathbb{R}^3$.

Solution:

Consider the matrix $A = \begin{bmatrix} 2 & -2 & 5 \\ -3 & 1 & 1 \\ 2 & 7 & -4 \end{bmatrix}$ with the vectors in S as its rows. A routine computation shows that $det A = -117 \neq 0$, so A is invertible. Hence S is independent by Theorem 5.2.3. Note that Theorem 5.2.3 also shows that $\mathbb{R}^3 = span S$.

Dimension

It is common geometrical language to say that $\mathbb{R}^3$ is 3-dimensional, that planes are 2-dimensional and that lines are 1-dimensional. The next theorem is a basic tool for clarifying this idea of "dimension".

Theorem 5.2.4 Fundamental Theorem

Let U be a subspace of $\mathbb{R}^n$. If U is spanned by m vectors, and if U contains k linearly independent vectors, then $k \leq m$

Definition 5.4 Basis of $\mathbb{R}^n$

If U is a subspace of $\mathbb{R}^n$, a set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ of vectors in U is called a **basis** of U if it satisfies the following two conditions:

1. $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ is linearly independent.
2. $U = span\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$.

Theorem 5.2.5 Invariance Theorem

If $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ and $\{\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_k\}$ are bases of a subspace U of $\mathbb{R}^n$, then $m = k$.

Proof:

We have $k \leq m$ by the fundamental theorem because $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ spans U, and $\{\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_k\}$ is

independent. Similarly, by interchanging $\vec{x}$'s and $\vec{y}$'s we get $m \leq k$. Hence $m = k$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-98

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-99

Definition 5.5 Dimension of a Subspace of $\mathbb{R}^n$

If U is a subspace of $\mathbb{R}^n$ and $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ is any basis
of U, the number, m, of vectors in the basis is called the **dimension** of U, denoted

$$dim\, U = m$$

The importance of the invariance theorem is that the dimension of U can be determined by counting the number of vectors in any basis.

Let $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ denote the standard basis of $\mathbb{R}^n$, that is the set of columns of the identity matrix. Then $\mathbb{R}^n = span\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ by Example 5.1.6, and $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ is independent by Example 5.2.2. Hence it is indeed a basis of $\mathbb{R}^n$ in the present terminology, and we have

Example 5.2.10

$$dim(\mathbb{R}^n) = n \text{ and } \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\} \text{ is a basis.}$$

This agrees with our geometric sense that $\mathbb{R}^2$ is two-dimensional and $\mathbb{R}^3$ is three-dimensional. It also says that $\mathbb{R}^1 = \mathbb{R}$ is one-dimensional, and $\{1\}$ is a basis. Returning to subspaces of $\mathbb{R}^n$, we define

$$dim\{\vec{0}\} = 0$$

This amounts to saying $\{\vec{0}\}$ has a basis containing no vectors. This makes sense because $\vec{0}$ cannot belong to any independent set.

Example 5.2.11

Let $U = \left\{ \begin{bmatrix} r \\ s \\ r \end{bmatrix} \mid r, s \text{ in } \mathbb{R} \right\}$. Show that U is a subspace of $\mathbb{R}^3$, find a basis, and calculate $dimU$.

Solution:

Clearly,

$$\begin{bmatrix} r \\ s \\ r \end{bmatrix} = r\vec{u} + s\vec{v} \text{ where}$$

$$\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \vec{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}. \quad \text{It} \quad \text{follows} \quad \text{that}$$

$U = span\{\vec{u}, \vec{v}\}$, and hence that U is a subspace of $\mathbb{R}^3$. Moreover, if $r\vec{u} + s\vec{v} = \vec{0}$, then

$$\begin{bmatrix} r \\ s \\ r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \text{ so } r = s = 0. \text{ Hence } \{\vec{u}, \vec{v}\} \text{ is independent,}$$

and so a basis of U. This means $dimU = 2$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=76#h5p-97

While we have found bases in many subspaces of $\mathbb{R}^n$, we have not yet shown that every subspace has a basis.

Let $U \neq \{\vec{0}\}$ be a subspace of $\mathbb{R}^n$. Then:

1. U has a basis and $dimU \leq n$.
2. Any independent set in U can be enlarged (by adding vectors from the standard basis) to a basis of U.
3. Any spanning set for U can be cut down (by deleting vectors) to a basis of U.

Example 5.2.3

Find a basis of $\mathbb{R}^4$ containing $S = \{\vec{u}, \vec{v}\}$ where $\vec{u} = (0, 1, 2, 3)$ and $\vec{v} = (2, -1, 0, 1)$.

Solution:

By Theorem 5.2.6 we can find such a basis by adding vectors from the standard basis of $\mathbb{R}^4$ to S. If we try $\vec{e}_1 = (1, 0, 0, 0)$, we find easily that $\{\vec{e}_1, \vec{u}, \vec{v}\}$ is independent. Now add another vector from the standard basis, say $\vec{e}_2$.

Again we find that $B = \{\vec{e}_1, \vec{e}_2, \vec{u}, \vec{v}\}$ is independent. Since

B has $4 = dim\mathbb{R}^4$ vectors, then B must span $\mathbb{R}^4$ by Theorem 5.2.7 below (or simply verify it directly). Hence B is a basis of $\mathbb{R}^4$.

Theorem 5.2.7

Let U be a subspace of $\mathbb{R}^n$ where $dimU = m$ and let $B = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ be a set of m vectors in U. Then B is independent if and only if B spans U.

Proof:

Suppose B is independent. If B does not span U then, by Theorem 5.2.6, B can be enlarged to a basis of U containing more than m vectors. This contradicts the invariance theorem because $dimU = m$, so B spans U. Conversely, if B spans U but is not independent, then B can be cut down to a basis of U containing fewer than m vectors, again a contradiction. So B is independent, as required.

Theorem 5.2.8

Let $U \subseteq W$ be subspaces of $\mathbb{R}^n$. Then:

1. $dimU \leq dimW$.
2. If $dimU = dimW$, then $U = W$.

Proof:

Write $dimW = k$, and let B be a basis of U.

1. If $dimU > k$, then B is an independent set in W containing more than k vectors, contradicting the fundamental theorem. So $dimU \leq k = dimW$.
2. If $dimU = k$, then B is an independent set in W containing $k = dimW$ vectors, so B spans W by Theorem~??. Hence $W = spanB = U$, proving (2).

It follows from Theorem 5.2.8 that if U is a subspace of $\mathbb{R}^n$, then $dimU$ is one of the integers $0, 1, 2, \ldots, n$, and that:

$$dimU = 0 \quad \text{if and only if} \quad U = \{\vec{0}\},$$
$$dimU = n \quad \text{if and only if} \quad U = \mathbb{R}^n$$

The other subspaces of $\mathbb{R}^n$ are called **proper**. The following example uses Theorem 5.2.8 to show that the proper subspaces of $\mathbb{R}^2$ are the lines through the origin, while the proper subspaces of $\mathbb{R}^3$ are the lines and planes through the origin.

Example 5.2.14

1. If U is a subspace of $\mathbb{R}^2$ or $\mathbb{R}^3$, then $dimU = 1$ if and only if U is a line through the origin.
2. If U is a subspace of $\mathbb{R}^3$, then $dimU = 2$ if and only if U is a plane through the origin.

Solution:

1. Since $dimU = 1$, let $\{\vec{u}\}$ be a basis of U. Then $U = span\{\vec{u}\} = \{t\vec{u} \mid t \text{ in } \mathbb{R}\}$, so U is the line

through the origin with direction vector $\vec{u}$. Conversely each line L with direction vector $\vec{d} \neq \vec{0}$ has the form $L = \{t\vec{d} \mid t \text{ in } \mathbb{R}\}$. Hence $\{\vec{d}\}$ is a basis of U, so U has dimension 1.

2. If $U \subseteq \mathbb{R}^3$ has dimension 2, let $\{\vec{v}, \vec{w}\}$ be a basis of U. Then $\vec{v}$ and $\vec{w}$ are not parallel (by Example 5.2.7) so $\vec{n} = \vec{v} \times \vec{w} \neq \vec{0}$. Let $P = \{\vec{x} \text{ in } \mathbb{R}^3 \mid \vec{n} \cdot \vec{x} = 0\}$ denote the plane through the origin with normal $\vec{n}$. Then P is a subspace of $\mathbb{R}^3$ (Example 5.1.1) and both $\vec{v}$ and $\vec{w}$ lie in P (they are orthogonal to $\vec{n}$), so $U = span\{\vec{v}, \vec{w}\} \subseteq P$ by Theorem 5.1.1. Hence

$$U \subseteq P \subseteq \mathbb{R}^3$$

Since $dimU = 2$ and $dim(\mathbb{R}^3) = 3$, it follows from Theorem 5.2.8 that $dimP = 2$ or 3, whence $P = U$ or $\mathbb{R}^3$. But $P \neq \mathbb{R}^3$ (for example, $\vec{n}$ is not in P) and so $U = P$ is a plane through the origin.

Conversely, if U is a plane through the origin, then $dimU = 0$, 1, 2, or 3 by Theorem 5.2.8. But $dimU \neq 0$ or 3 because $U \neq \{\vec{0}\}$ and $U \neq \mathbb{R}^3$, and $dimU \neq 1$ by (1). So $dimU = 2$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-94

5.3 Orthogonality

Length and orthogonality are basic concepts in geometry and, in $\mathbb{R}^2$ and $\mathbb{R}^3$, they both can be defined using the dot product. In this section we extend the dot product to vectors in $\mathbb{R}^n$, and so endow $\mathbb{R}^n$ with euclidean geometry. We then introduce the idea of an orthogonal basis—one of the most useful concepts in linear algebra, and begin exploring some of its applications.

Dot Product, Length, and Distance

If $\vec{x} = (x_1, x_2, \ldots, x_n)$ and $\vec{y} = (y_1, y_2, \ldots, y_n)$ are two n-tuples in $\mathbb{R}^n$, recall that their **dot product** was defined in Section 2.2 as follows:

$$\vec{x} \cdot \vec{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

Observe that if $\vec{x}$ and $\vec{y}$ are written as columns then $\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y}$ is a matrix product (and $\vec{x}\vec{y} = x\vec{y}^T$ if they are written as rows). Here $\vec{x} \cdot \vec{y}$ is a 1×1 matrix, which we take to be a number.

Definition Length in $\mathbb{R}^n$

As in $\mathbb{R}^3$, the **length** of the vector is defined by

$$||\vec{x}|| = \sqrt{\vec{x} \cdot \vec{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

Where $\sqrt{(\quad)}$ indicates the positive square root.

A vector $\vec{x}$ of length 1 is called a **unit vector**. If $\vec{x} \neq \vec{0}$, then $||\vec{x}|| \neq 0$ and it follows easily that $\frac{1}{||\vec{x}||}\vec{x}$ is a unit vector (see Theorem 5.3.6 below), a fact that we shall use later.

Example 5.3.1

If $\vec{x} = (1, -1, -3, 1)$ and $\vec{y} = (2, 1, 1, 0)$ in $\mathbb{R}^4$, then $\vec{x} \cdot \vec{y} = 2 - 1 - 3 + 0 = -2$ and $||\vec{x}|| = \sqrt{1 + 1 + 9 + 1} = \sqrt{12} = 2\sqrt{3}$. Hence $\frac{1}{2\sqrt{3}} \vec{x}$ is a unit vector; similarly $\frac{1}{\sqrt{6}} \vec{y}$ is a unit vector.

Theorem 5.3.1

Let $\vec{x}, \vec{y}$, and $\vec{z}$ denote vectors in $\mathbb{R}^n$. Then:

1. $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$.
2. $\vec{x} \cdot (\vec{y} + \vec{z}) = \vec{x} \cdot \vec{y} + \vec{x} \cdot \vec{z}$.
3. $(a\vec{x}) \cdot \vec{y} = a(\vec{x} \cdot \vec{y}) = \vec{x} \cdot (a\vec{y})$ for all scalars a.
4. $||\vec{x}||^2 = \vec{x} \cdot \vec{x}$.
5. $||\vec{x}|| \geq 0$, and $||a\vec{x}|| = 0$ if and only if $\vec{x} = \vec{0}$.
6. $||a\vec{x}|| = |a|||\vec{x}||$ for all scalars a.

Proof:

(1), (2), and (3) follow from matrix arithmetic because $\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y}$; (4) is clear from the definition; and (6) is a routine verification since $|a| = \sqrt{a^2}$. If $\vec{x} = (x_1, x_2, \ldots, x_n)$, then $||\vec{x}|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ so $||\vec{x}|| = 0$ if and only if $x_1^2 + x_2^2 + \cdots + x_n^2 = 0$. Since each x_i is a real number this happens if and only if $x_i = 0$ for each i; that is, if and only if $\vec{x} = \vec{0}$. This proves (5).

Because of Theorem 5.3.1, computations with dot products in $\mathbb{R}^n$ are similar to those in $\mathbb{R}^3$. In particular, the dot product

$$(\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_m) \cdot (\vec{y}_1 + \vec{y}_2 + \cdots + \vec{y}_k)$$

equals the sum of mk terms, $\vec{x}_i \cdot \vec{y}_j$, one for each choice of i and j. For example:

$$(3\vec{x} - 4\vec{y}) \cdot (7\vec{x} + 2\vec{y}) = 21(\vec{x} \cdot \vec{x}) + 6(\vec{x} \cdot \vec{y}) - 28(\vec{y} \cdot \vec{x}) - 8(\vec{y} \cdot \vec{y})$$
$$= 21||\vec{x}||^2 - 22(\vec{x} \cdot \vec{y}) - 8||\vec{y}||^2$$

holds for all vectors $\vec{x}$ and $\vec{y}$.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-103

Example 5.3.2

Show that $||\vec{x} + \vec{y}||^2 = ||\vec{x}||^2 + 2(\vec{x} \cdot \vec{y}) + ||\vec{y}||^2$ for any $\vec{x}$ and $\vec{y}$ in $\mathbb{R}^n$.

Solution:

Using Theorem 5.3.1 several times:

$$||\vec{x} + \vec{y}||^2 = (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) = \vec{x} \cdot \vec{x} + \vec{x} \cdot \vec{y} + \vec{y} \cdot \vec{x} + \vec{y} \cdot \vec{y}$$
$$= ||\vec{x}||^2 + 2(\vec{x} \cdot \vec{y}) + ||\vec{y}||^2$$

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-108

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-109

Example 5.3.3

Suppose that $\mathbb{R}^n = span\{\vec{f_1}, \vec{f_2}, \ldots, \vec{f_k}\}$ for

some vectors $\vec{f_i}$. If $\vec{x} \cdot \vec{f_i} = 0$ for each i where $\vec{x}$ is in $\mathbb{R}^n$, show that $\vec{x} = \vec{0}$.

Solution:

We show $\vec{x} = \vec{0}$ by showing that $||\vec{x}|| = 0$ and using (5) of Theorem 5.3.1. Since the $\vec{f_i}$ span $\mathbb{R}^n$, write $\vec{x} = t_1 \vec{f_1} + t_2 \vec{f_2} + \cdots + t_k \vec{f_k}$ where the t_i are in $\mathbb{R}$. Then

$$||\vec{x}||^2 = \vec{x} \cdot \vec{x} = \vec{x} \cdot (t_1 \vec{f_1} + t_2 \vec{f_2} + \cdots + t_k \vec{f_k})$$

$$= t_1 (\vec{x} \cdot \vec{f_1}) + t_2 (\vec{x} \cdot \vec{f_2}) + \cdots + t_k (\vec{x} \cdot \vec{f_k})$$

$$= t_1 (0) + t_2 (0) + \cdots + t_k (0)$$

$$= 0$$

We saw in Section 4.2 that if $\vec{u}$ and v are nonzero vectors in $\mathbb{R}^3$, then $\frac{\vec{u} \cdot \vec{v}}{||\vec{u}|| ||\vec{v}||} = \cos \theta$ where θ is the angle between $\vec{u}$ and $\vec{v}$. Since $|\cos \theta| \leq 1$ for any angle θ, this shows that $|\vec{u} \cdot \vec{v}| \leq ||\vec{u}|| ||\vec{v}||$. In this form the result holds in $\mathbb{R}^n$.

Theorem 5.3.2 Cauchy Inequality

If $\vec{x}$ and $\vec{y}$ are vectors in $\mathbb{R}^n$, then

$$|\vec{x} \cdot \vec{y}| \leq ||\vec{x}|| ||\vec{y}||$$

Proof:

The inequality holds if $\vec{x} = \vec{0}$ or $\vec{y} = \vec{0}$ (in fact it is equality). Otherwise, write $||\vec{x}|| = a > 0$ and $||\vec{y}|| = b > 0$ for convenience. A computation like that preceding Example 5.3.2 gives

(5.1)
$$||b\vec{x} - a\vec{y}||^2 = 2ab(ab - \vec{x} \cdot \vec{y}) \text{ and } ||b\vec{x} + a\vec{y}||^2 = 2ab(ab + \vec{x} \cdot \vec{y})$$

It follows that $ab - \vec{x} \cdot \vec{y} \geq 0$ and $ab + \vec{x} \cdot \vec{y} \geq 0$, and hence that $-ab \leq \vec{x} \cdot \vec{y} \leq ab$. Hence $|\vec{x} \cdot \vec{y}| \leq ab = ||\vec{x}|| ||\vec{y}||$, proving the Cauchy inequality.

If equality holds, then $|\vec{x} \cdot \vec{y}| = ab$, so $\vec{x} \cdot \vec{y} = ab$ or $\vec{x} \cdot \vec{y} = -ab$. Hence Equation (5.1) shows that $b\vec{x} - a\vec{y} = 0$ or $b\vec{x} + a\vec{y} = 0$, so one of $\vec{x}$ and $\vec{y}$ is a multiple of the other (even if $a = 0$ or $b = 0$).

There is an important consequence of the Cauchy inequality. Given $\vec{x}$ and $\vec{y}$ in $\mathbb{R}^n$, use Example 5.3.2 and the fact that $\vec{x} \cdot \vec{y} \leq ||\vec{x}|| ||\vec{y}||$ to compute

$$||\vec{x} + \vec{y}||^2 = ||\vec{x}||^2 + 2(\vec{x} \cdot \vec{y}) + ||\vec{y}||^2 \leq ||\vec{x}||^2 + 2||\vec{x}|| ||\vec{y}|| + ||\vec{y}||^2 = (||x + y||)^2$$

Taking positive square roots gives:

Corollary 5.3.1 Triangle Inequality

If $\vec{x}$ and $\vec{y}$ are vectors in $\mathbb{R}^n$, then
$$||\vec{x} + \vec{y}|| \leq ||x|| + ||\vec{y}||.$$

The reason for the name comes from the observation that in $\mathbb{R}^3$ the inequality asserts that the sum of the lengths of two sides of a triangle is not less than the length of the third side.

Definition 5.7 Distance in $\mathbb{R}^n$

If $\vec{x}$ and $\vec{y}$ are two vectors in $\mathbb{R}^n$, we define the **distance** $d(\vec{x}, \vec{y})$ between $\vec{x}$ and $\vec{y}$ by

$$d(\vec{x}, \vec{y}) = ||x - y||$$

Theorem 5.3.3

If $\vec{x}, \vec{y}$, and $\vec{z}$ are three vectors in $\mathbb{R}^n$ we have:

1. $d(\vec{x}, \vec{y}) \geq 0$ for all $\vec{x}$ and $\vec{y}$.
2. $d(\vec{x}, \vec{y}) = 0$ if and only if $\vec{x} = \vec{y}$.
3. $d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x})$ for all $\vec{x}$ and $\vec{y}$.
4. $d(\vec{x}, \vec{z}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$ for all $\vec{x}, \vec{y}$, and $\vec{z}$. \quad Triangle inequality.

Proof:

(1) and (2) restate part (5) of Theorem 5.3.1 because $d(\vec{x}, \vec{y}) = ||\vec{x} - \vec{y}||$, and (3) follows because $||\vec{u}|| = || - \vec{u}||$

for every vector $\vec{u}$ in $\mathbb{R}^n$. To prove (4) use the Corollary to Theorem 5.3.2:

$$d(\vec{x}, \vec{z}) = ||\vec{x} - \vec{z}|| = ||(\vec{x} - \vec{y}) + (\vec{y} - \vec{z})||$$
$$\leq ||(\vec{x} - \vec{y})|| + ||(\vec{y} - \vec{z})|| = d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$$

Orthogonal Sets and the Expansion Theorem

Definition 5.8 Orthogonal and Orthonormal Sets

We say that two vectors $\vec{x}$ and $\vec{y}$ in $\mathbb{R}^n$ are **orthogonal** if $\vec{x} \cdot \vec{y} = 0$, extending the terminology in $\mathbb{R}^3$ (See Theorem 4.2.3). More generally, a set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ of vectors in R^n is called an **orthogonal set** if

$$\vec{x}_i \cdot \vec{x}_j = 0 \text{ for all } i \neq j \quad \text{and} \quad x_i \neq 0 \text{ for all } i$$

Note that $\{\vec{x}\}$ is an orthogonal set if $\vec{x} \neq \vec{0}$. A set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ of vectors in $\mathbb{R}^n$ is called **orthonormal** if it is orthogonal and, in addition, each $\vec{x}_i$ is a unit vector:

$$||\vec{x}_i|| = 1 \text{ for each } i.$$

Example 5.3.4

The standard basis $\left\{ \vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n \right\}$ is an orthonormal set in $\mathbb{R}^n$.

Example 5.3.5

If $\left\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\right\}$ is orthogonal, so also is $\left\{a_1\vec{x}_1, a_2\vec{x}_2, \ldots, a_k\vec{x}_k\right\}$ for any nonzero scalars a_i.

🖳 *An interactive H5P element has been excluded from this version of the text. You can view it online here:*

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-106

🖳 *An interactive H5P element has been excluded from this version of the text. You can view it online here:*

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-105

If $\vec{x} \neq \vec{0}$, it follows from item (6) of Theorem 5.3.1 that $\frac{1}{||\vec{x}||}\vec{x}$ is a unit vector, that is it has length 1.

Definition 5.9 Normalizing an Orthogonal Set

If $\left\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\right\}$ is an orthogonal set, then $\left\{\frac{1}{||\vec{x_1}||}\vec{x}_1, \frac{1}{||\vec{x_2}||}\vec{x}_2, \cdots, \frac{1}{||\vec{x_k}||}\vec{x}_k\right\}$ is an orthonormal

set, and we say that it is the result of **normalizing** the orthogonal set $\{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_k\}$.

If $\vec{f}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix}, \vec{f}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}, \vec{f}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix},$

and $\vec{f}_4 = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 1 \end{bmatrix}$

then $\{\vec{f}_1, \vec{f}_2, \vec{f}_3, \vec{f}_4\}$ is an orthogonal set in $\mathbb{R}^4$ as is easily verified. After normalizing, the corresponding orthonormal set is $\{\frac{1}{2}\vec{f}_1, \frac{1}{\sqrt{6}}\vec{f}_2, \frac{1}{\sqrt{2}}\vec{f}_3, \frac{1}{2\sqrt{3}}\vec{f}_4\}$

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-101

The most important result about orthogonality is Pythagoras' theorem. Given orthogonal vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^3$, it asserts that

$$||v + w||^2 = ||\vec{v}||^2 + ||\vec{w}||^2$$

. In this form the result holds for any orthogonal set in $\mathbb{R}^n$.

Theorem 5.3.4 Pythagoras' Theorem

If $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ is an orthogonal set in $\mathbb{R}^n$, then

$$||\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_k||^2 = \vec{x_1}||^2 + \vec{x_2}||^2 + \cdots + \vec{x_k}||^2.$$

Proof:

The fact that $\vec{x}_i \cdot \vec{x}_j = 0$ whenever $i \neq j$ gives

$$
\begin{aligned}
||\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_k||^2 &= (\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_k) \cdot (\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_k) \\
&= (\vec{x}_1 \cdot \vec{x}_1 + \vec{x}_2 \cdot \vec{x}_2 + \cdots + \vec{x}_k \cdot \vec{x}_k) + \sum_{i \neq j} \vec{x}_i \cdot \vec{x}_j \\
&= ||\vec{x}_1||^2 + ||\vec{x}_2||^2 + \cdots + ||\vec{x}_k||^2 + 0
\end{aligned}
$$

This is what we wanted.

If $\vec{v}$ and $\vec{w}$ are orthogonal, nonzero vectors in $\mathbb{R}^3$, then they are certainly not parallel, and so are linearly independent Example 5.2.7. The next theorem gives a far-reaching extension of this observation.

Every orthogonal set in $\mathbb{R}^n$ is linearly independent.

Proof:

Let $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ be an orthogonal set in $\mathbb{R}^n$ and suppose a linear combination vanishes, say: $t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k = \vec{0}$. Then

$$
\begin{aligned}
0 = \vec{x}_1 \cdot \vec{0} &= \vec{x}_1 \cdot (t_1\vec{x}_1 + t_2\vec{x}_2 + \cdots + t_k\vec{x}_k) \\
&= t_1(\vec{x}_1 \cdot \vec{x}_1) + t_2(\vec{x}_1 \cdot \vec{x}_2) + \cdots + t_k(\vec{x}_1 \cdot \vec{x}_k) \\
&= t_1||\vec{x}_1||^2 + t_2(0) + \cdots + t_k(0) \\
&= t_1||\vec{x}_1||^2
\end{aligned}
$$

Since $||\vec{x}_1||^2 \neq 0$, this implies that $t_1 = 0$. Similarly $t_i = 0$ for each i.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

Theorem 5.3.6

Let $\{\vec{f_1}, \vec{f_2}, \ldots, \vec{f_m}\}$ be an orthogonal basis of a subspace U of $\mathbb{R}^n$. If $\vec{x}$ is any vector in U, we have

$$\vec{x} = \left(\frac{\vec{x} \cdot \vec{f_1}}{||\vec{f_1}||^2}\right) \vec{f_1} + \left(\frac{\vec{x} \cdot \vec{f_2}}{||\vec{f_2}||^2}\right) \vec{f_1} + \cdots + \left(\frac{\vec{x} \cdot \vec{f_m}}{||\vec{f_m}||^2}\right) \vec{f_m}$$

Proof:

Since $\{\vec{f_1}, \vec{f_2}, \ldots, \vec{f_m}\}$ spans U, we have $\vec{x} = t_1 \vec{f_1} + t_2 \vec{f_2} + \cdots + t_m \vec{f_m}$ where the t_i are scalars. To find t_1 we take the dot product of both sides with $\vec{f_1}$:

$$\vec{x} \cdot \vec{f_1} = (t_1\vec{f_1} + t_2\vec{f_2} + \cdots + t_m\vec{f_m}) \cdot \vec{f_1}$$

$$= t_1(\vec{f_1} \cdot \vec{f_1}) + t_2(\vec{f_2} \cdot \vec{f_1}) + \cdots + t_m(\vec{f_m} \cdot \vec{f_1})$$

$$= t_1||\vec{f_1}||^2 + t_2(0) + \cdots + t_m(0)$$

$$= t_1||\vec{f_1}||^2$$

Since $\vec{f_1} \neq \vec{0}$, this gives $t_1 = \frac{\vec{x} \cdot \vec{f_1}}{||\vec{f_1}||^2}$. Similarly, $t_i = \frac{\vec{x} \cdot \vec{f_i}}{||\vec{f_i}||^2}$ for each i.

The expansion in Theorem 5.3.6 of $\vec{x}$ as a linear combination of the orthogonal basis $\{\vec{f_1}, \vec{f_2}, \ldots, \vec{f_m}\}$ is called the **Fourier expansion** of $\vec{x}$, and the coefficients $t_1 = \frac{\vec{x} \cdot \vec{f_i}}{||f_i||^2}$ are called the **Fourier coefficients.** Note that if $\{\vec{f_1}, \vec{f_2}, \ldots, \vec{f_m}\}$ is actually orthonormal, then $t_i = \vec{x} \cdot \vec{f_i}$ for each i.

Example 5.3.7

Expand $\vec{x} = (a, b, c, d)$ as a linear combination of the orthogonal basis $\{\vec{f_1}, \vec{f_2}, \vec{f_3}, \vec{f_4}\}$ of $\mathbb{R}^4$ given in Example 5.3.6.

Solution:

We have $\vec{f_1} = (1, 1, 1, -1)$, $\quad \vec{f_2} = (1, 0, 1, 2)$, $\vec{f_3} = (-1, 0, 1, 0)$, and $\vec{f_4} = (-1, 3, -1, 1)$ so the Fourier coefficients are

$$t_1 = \frac{\vec{x} \cdot \vec{f_1}}{||\vec{f_1}||^2} = \tfrac{1}{4}(a + b + c + d) \qquad\qquad t_3 = \frac{\vec{x} \cdot \vec{f_3}}{||\vec{f_3}||^2} = \tfrac{1}{2}(-a + c)$$

$$t_2 = \frac{\vec{x} \cdot \vec{f_2}}{||\vec{f_2}||^2} = \tfrac{1}{6}(a + c + 2d) \qquad\qquad t_4 = \frac{\vec{x} \cdot \vec{f_4}}{||\vec{f_4}||^2} = \tfrac{1}{12}(-a + 3b - c + d)$$

The reader can verify that indeed $\vec{x} = t_1 \vec{f_1} + t_2 \vec{f_2} + t_3 \vec{f_3} + t_4 \vec{f_4}.$

5.4 Rank of a Matrix

In this section we use the concept of dimension to clarify the definition of the rank of a matrix given in Section 1.2, and to study its properties. This requires that we deal with rows and columns in the same way. While it has been the custom to write the n-tuples as columns, in this section we will frequently write them as rows. Subspaces, independence, spanning, and dimension are defined for rows using matrix operations, just as for columns. If A is an $m \times n$ matrix, we define:

Definition 5.10 Column and Row Space of a Matrix

The **column space**, $colA$, of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A.

The **row space**, $rowA$, of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A.

Lemma 5.4.1

Let A and B denote $m \times n$ matrices.

1. If $A \to B$ by elementary row operations, then $rowA = rowB$.
2. If $A \to B$ by elementary column operations, then $colA = colB$.

Proof:

We prove (1); the proof of (2) is analogous. It is enough to do it in the case when $A \to B$ by a single row operation. Let $R_1, R_2, \ldots, R_m$ denote the rows of A. The row operation $A \to B$ either interchanges two rows, multiplies a row by a nonzero constant, or adds a multiple of a row to a different row. We leave the first two cases to the reader. In the last case, suppose that a times row p is added to row q where $p < q$. Then the rows of B are $R_1, \ldots, R_p, \ldots, R_q + aR_p, \ldots, R_m$, and Theorem 5.1.1 shows that

$$span\{R_1, \ldots, R_p, \ldots, R_q, \ldots, R_m\} = span\{R_1, \ldots, R_p, \ldots, R_q + aR_p, \ldots, R_m\}$$

That is, $rowA = rowB$.

If A is any matrix, we can carry $A \to R$ by elementary row operations where R is a row-echelon matrix. Hence $rowA = rowR$ by Lemma 5.4.1; so the first part of the following result is of interest.

Lemma 5.4.2

If R is a row-echelon matrix, then

1. The nonzero rows of R are a basis of $rowR$.
2. The columns of R containing leading ones are a basis of $colR$.

Proof:

The rows of R are independent, and they span $rowR$ by definition. This proves (1).

Let $\vec{c}_{j_1}, \vec{c}_{j_2}, \ldots, \vec{c}_{j_r}$ denote the columns of R containing leading 1s. Then $\{\vec{c}_{j_1}, \vec{c}_{j_2}, \ldots, \vec{c}_{j_r}\}$ is independent because the leading 1s are in different rows (and have zeros below and to the left of them). Let U denote the subspace of all columns in $\mathbb{R}^m$ in which the last $m - r$ entries are zero. Then $dimU = r$ (it is just $\mathbb{R}^r$ with extra zeros). Hence the independent set $\{\vec{c}_{j_1}, \vec{c}_{j_2}, \ldots, \vec{c}_{j_r}\}$ is a basis of U by Theorem 5.2.7. Since each $\vec{c}_{j_i}$ is in $colR$, it follows that $colR = U$, proving (2).

Let A be any matrix and suppose A is carried to some row-echelon matrix R by row operations. Note that R is not unique. In Section 1.2 we defined the **rank** of A, denoted $rankA$, to be the number of leading 1s in R, that is the number of nonzero rows of R. The fact that this number does not depend on the choice of R was not proved. However part 1 of Lemma 5.4.2 shows that

$$rankA = dim(rowA)$$

and hence that $rankA$ is independent of R.

Lemma 5.4.2 can be used to find bases of subspaces of $\mathbb{R}^n$ (written as rows). Here is an example.

Find a basis of
$$U = span\{(1, 1, 2, 3), (2, 4, 1, 0), (1, 5, -4, -9)\}$$
.

Solution:

U is the row space of $\begin{bmatrix} 1 & 1 & 2 & 3 \\ 2 & 4 & 1 & 0 \\ 1 & 5 & -4 & -9 \end{bmatrix}$. This matrix has

row-echelon form $\begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 1 & -\frac{3}{2} & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, so

$\{(1, 1, 2, 3), (0, 1, -\frac{3}{2}, -3)\}$ is basis of U by Lemma 5.4.1.

Note that $\{(1, 1, 2, 3), (0, 2, -3, -6)\}$ is another basis that avoids fractions.

Theorem 5.4.1 Rank Theorem

Let A denote any $m \times n$ matrix of rank r. Then

$$dim(colA) = dim(rowA) = r$$

Moreover, if A is carried to a row-echelon matrix R by row operations, then

1. The r nonzero rows of R are a basis of $rowA$.
2. If the leading 1s lie in columns $j_1, j_2, \ldots, j_r$ of R, then columns $j_1, j_2, \ldots, j_r$ of A are a basis of $colA$.

Proof:

We have $rowA = rowR$ by Lemma 5.4.1, so (1) follows from 5.4.2. Moreover, $R = UA$ for some invertible matrix U. Now write $A = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \ldots & \vec{c}_n \end{bmatrix}$ where $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n$ are the columns of A. Then

$$R = UA = U \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \ldots & \vec{c}_n \end{bmatrix} = \begin{bmatrix} U\vec{c}_1 & U\vec{c}_2 & \cdots & U\vec{c}_n \end{bmatrix}$$

Thus, in the notation of (2), the set $B = \{U\vec{c}_{j_1}, U\vec{c}_{j_2}, \ldots, U\vec{c}_{j_r}\}$ is a basis of $colR$ by Lemma 5.4.2. So, to prove (2) and the fact that $dim(colA) = r$, it is enough to show that $D = \{\vec{c}_{j_1}, \vec{c}_{j_2}, \ldots, \vec{c}_{j_r}\}$ is a basis of $colA$. First, D is linearly independent because U is invertible (verify), so we show that, for each j, column $\vec{c}_j$ is a linear combination of the $\vec{c}_{j_i}$. But $U\vec{c}_j$ is column j of R, and so is a linear combination of the $U\vec{c}_{j_i}$, say $U\vec{c}_j = a_1 U\vec{c}_{j_1} + a_2 U\vec{c}_{j_2} + \cdots + a_r U\vec{c}_{j_r}$ where each a_i is a real number.

Since U is invertible, it follows that $\vec{c}_j = a_1 \vec{c}_{j_1} + a_2 c_{j_2} + \cdots + a_r \vec{c}_{j_r}$ and the proof is complete.

Example 5.4.2

Compute the rank of $A = \begin{bmatrix} 1 & 2 & 2 & -1 \\ 3 & 6 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{bmatrix}$ and find bases for $row\,A$ and $col\,A$.

Solution:

The reduction of A to row-echelon form is as follows:

$$\begin{bmatrix} 1 & 2 & 2 & -1 \\ 3 & 6 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 2 & -1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & -1 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 2 & -1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Hence $rank\,A$ = 2, and
$$\{ \begin{bmatrix} 1 & 2 & 2 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 & -3 \end{bmatrix} \}$$
is a basis of $row\,A$ by 5.4.2. Since the leading 1s are in columns 1 and 3 of the row-echelon matrix, Theorem 5.4.1 shows that columns 1 and 3 of A are a basis
$$\left\{ \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 1 \end{bmatrix} \right\}$$
of $col\,A$.

If A is any matrix, then $rank\,A = rank(A^T)$.

If A is an $m \times n$ matrix, we have $col\,A \subseteq \mathbb{R}^m$ and $row\,A \subseteq \mathbb{R}^n$. Hence Theorem 5.2.8 shows that $dim(col\,A) \leq dim(\mathbb{R}^m) = m$ and $dim(row\,A) \leq dim(\mathbb{R}^n) = n$. Thus Theorem 5.4.1 gives:

If A is an $m \times n$ matrix, then $rank\,A \leq m$ and $rank\,A \leq n$.

$$rankA = rank(UA) = rank(AV) \text{ whenever}$$

U and V are invertible.

Proof:

Lemma 5.4.1 gives $rankA = rank(UA)$. Using this and Corollary 5.4.1 we get

$$rank(AV) = rank(AV)^T = rank(V^T A^T) = rank(A^T) = rankA$$

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=76#h5p-117

Let A, U, and V be matrices of sizes $m \times n, p \times m,$ and $n \times q$ respectively.

1. $col(AV) \subseteq colA$, with equality if $VV' = I_n$ for some V'.
2. $row(UA) \subseteq rowA$, with equality if $U'U = I_m$ for some U'.

Proof:

For (1), write $V = [\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_q]$ where $\vec{v}_j$ is column j of V. Then we have $AV = [A\vec{v}_1, A\vec{v}_2, \ldots, A\vec{v}_q]$, and each $A\vec{v}_j$ is in $colA$ by Definition 2.4. It follows that $col(AV) \subseteq colA$. If $VV' = I_n,$ we obtain $colA = col[(AV)V'] \subseteq col(AV)$ in the same way. This proves (1).

As to (2), we have $col[(UA)^T] = col(A^TU^T) \subseteq col(A^T)$ by (1), from which $row(UA) \subseteq rowA$. If $U'U = I_m,$ this is equality as in the proof of (1).

Corollary 5.4.4

If A is $m \times n$ and B is $n \times m$, then $rankAB \leq rankA$ and $rankAB \leq rankB$.

Proof:

By Lemma 5.4.3 $col(AB) \subseteq colA$ and $row(BA) \subseteq rowA$, so Theorem 5.4.1 applies.

In Section 5.1 we discussed two other subspaces associated with an $m \times n$ matrix A: the null space $null(A)$ and the image space $im(A)$

$$null(A) = \{\vec{x} \text{ in } \mathbb{R}^n \mid A\vec{x} = \vec{0}\} \text{ and } im(A) = \{A\vec{x} \mid \vec{x} \text{ in } \mathbb{R}^n\}$$

Using rank, there are simple ways to find bases of these spaces. If A has rank r, we have $im(A) = col(A)$ by Example 5.1.8, so $dim[im(A)] = dim[col(A)] = r$. Hence Theorem 5.4.1 provides a method of finding a basis of $im(A)$. This is recorded as part (2) of the following theorem.

Theorem 5.4.2

Let A denote an $m \times n$ matrix of rank r. Then

1. The $n - r$ basic solutions to the system $A\vec{x} = \vec{0}$ provided by the gaussian algorithm are a basis of $null(A)$, so $dim[null(A)] = n - r$.
2. Theorem 5.4.1 provides a basis of $im(A) = col(A)$, and $dim[im(A)] = r$.

Proof:

It remains to prove (1). We already know (Theorem 2.2.1) that $null(A)$ is spanned by the $n - r$ basic solutions of $A\vec{x} = \vec{0}$. Hence using Theorem 5.2.7, it suffices to show that $dim[null(A)] = n - r$. So let $\{\vec{x}_1, \dots, \vec{x}_k\}$ be a basis of $null(A)$, and extend it to a basis $\{\vec{x}_1, \dots, \vec{x}_k, \vec{x}_{k+1}, \dots, \vec{x}_n\}$ of $\mathbb{R}^n$ (by Theorem 5.2.6). It is

enough to show that $\{A\vec{x}_{k+1}, \ldots, A\vec{x}_n\}$ is a basis of $im(A)$; then $n - k = r$ by the above and so $k = n - r$ as required.

Spanning. Choose $A\vec{x}$ in $im(A)$, $\vec{x}$ in $\mathbb{R}^n$, and write
$$\vec{x} = a_1\vec{x}_1 + \cdots + a_k\vec{x}_k + a_{k+1}\vec{x}_{k+1} + \cdots + a_n\vec{x}_n$$
where the a_i are in $\mathbb{R}$.

Then $\quad A\vec{x} = a_{k+1}A\vec{x}_{k+1} + \cdots + a_nA\vec{x}_n \quad$ because $\{\vec{x}_1, \ldots, \vec{x}_k\} \subseteq null(A)$.

Independence. Let $t_{k+1}A\vec{x}_{k+1} + \cdots + t_nA\vec{x}_n = \vec{0}$, t_i in $\mathbb{R}$. Then $t_{k+1}\vec{x}_{k+1} + \cdots + t_n\vec{x}_n$ is in $nullA$, so $t_{k+1}\vec{x}_{k+1} + \cdots + t_n\vec{x}_n = t_1\vec{x}_1 + \cdots + t_k\vec{x}_k$ for some $t_1, \ldots, t_k$ in $\mathbb{R}$. But then the independence of the $\vec{x}_i$ shows that $t_i = 0$ for every i.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-111

Example 5.4.3

If $A = \begin{bmatrix} 1 & -2 & 1 & 1 \\ -1 & 2 & 0 & 1 \\ 2 & -4 & 1 & 0 \end{bmatrix}$, find bases of $null(A)$ and $im(A)$, and so find their dimensions.

Solution:

If $\vec{x}$ is in $null(A)$, then $A\vec{x} = \vec{0}$, so $\vec{x}$ is given by solving the system $A\vec{x} = \vec{0}$. The reduction of the augmented matrix to reduced form is

$$
\left[\begin{array}{rrrr|r}
1 & -2 & 1 & 1 & 0 \\
-1 & 2 & 0 & 1 & 0 \\
2 & -4 & 1 & 0 & 0
\end{array}\right]
\rightarrow
\left[\begin{array}{rrrr|r}
1 & -2 & 0 & -1 & 0 \\
0 & 0 & 1 & 2 & 0 \\
0 & 0 & 0 & 0 & 0
\end{array}\right]
$$

Hence $r = rank(A) = 2$. Here, $im(A) = col(A)$ has basis

$$
\left\{
\left[\begin{array}{r} 1 \\ -1 \\ 2 \end{array}\right] ,
\left[\begin{array}{r} 1 \\ 0 \\ 1 \end{array}\right]
\right\}
$$

by Theorem 5.4.1 because the leading 1s are in columns 1 and 3. In particular, $dim[im(A)] = 2 = r$ as in Theorem 5.4.2.

Turning to $null(A)$, we use gaussian elimination. The leading variables are x_1 and x_3, so the nonleading variables become parameters: $x_2 = s$ and $x_4 = t$. It follows from the reduced matrix that $x_1 = 2s + t$ and $x_3 = -2t$, so the general solution is

$$
\vec{x} =
\left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array}\right]
=
\left[\begin{array}{c} 2s + t \\ s \\ -2t \\ t \end{array}\right]
= s\vec{x}_1 + t\vec{x}_2 \text{ where } \vec{x}_1 =
\left[\begin{array}{r} 2 \\ 1 \\ 0 \\ 0 \end{array}\right] , \text{ and } x_2 =
\left[\begin{array}{r} 1 \\ 0 \\ -2 \\ 1 \end{array}\right] .
$$

Hence $null(A)$. But $\vec{x}_1$ and $\vec{x}_2$ are solutions (basic), so

$$null(A) = span\{\vec{x}_1, \vec{x}_2\}$$

However Theorem 5.4.2 asserts that $\{\vec{x}_1, \vec{x}_2\}$ is a basis of $null(A)$. (In fact it is easy to verify directly that $\{\vec{x}_1, \vec{x}_2\}$ is independent in this case.) In particular, $dim[null(A)] = 2 = n - r$.

Let A be an $m \times n$ matrix. Corollary 5.4.2 asserts that $rankA \leq m$ and $rankA \leq n$, and it is natural to ask when these extreme cases arise. If $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n$ are the columns of A, Theorem 5.2.2 shows that $\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$ spans $\mathbb{R}^m$ if and only if the system $A\vec{x} = \vec{b}$ is consistent for every $\vec{b}$ in $\mathbb{R}^m$, and that $\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}$ is independent if and only if $A\vec{x} = \vec{0}, \vec{x}$ in $\mathbb{R}^n$, implies $\vec{x} = \vec{0}$. The next two useful theorems improve on both these results, and relate them to when the rank of A is n or m.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-113

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-114

Theorem 5.4.3

The following are equivalent for an $m \times n$ matrix A:

1. $rank A = n$.
2. The rows of A span $\mathbb{R}^n$.
3. The columns of A are linearly independent in $\mathbb{R}^m$.
4. The $n \times n$ matrix $A^T A$ is invertible.
5. $CA = I_n$ for some $n \times m$ matrix C.
6. If $A\vec{x} = \vec{0}, \vec{x}$ in $\mathbb{R}^n$, then $\vec{x} = 0$.

Proof:

(1) $\Longrightarrow$ (2). We have $row A \subseteq \mathbb{R}^n$, and $dim(row A) = n$ by (1), so $row A = \mathbb{R}^n$ by Theorem 5.2.8. This is (2).

(2) $\Longrightarrow$ (3). By (2), $row A = \mathbb{R}^n$, so $rank A = n$. This means $dim(col A) = n$. Since the n columns of A span $col A$, they are independent by Theorem 5.2.7.

$(3) \Rightarrow (4)$. If $(A^T A)\vec{x} = \vec{0}$, $\vec{x}$ in $\mathbb{R}^n$, we show that $\vec{x} = \vec{0}$ (Theorem 2.4.5). We have

$$||A\vec{x}||^2 = (A\vec{x})^T A\vec{x} = \vec{x}^T A^T A\vec{x} = \vec{x}^T \vec{0} = \vec{0}$$

Hence $A\vec{x} = \vec{0}$, so $\vec{x} = \vec{0}$ by (3) and Theorem 5.2.2.

$(4) \Rightarrow (5)$. Given (4), take $C = (A^T A)^{-1} A^T$.

$(5) \Rightarrow (6)$. If $A\vec{x} = \vec{0}$, then left multiplication by C (from (5)) gives $\vec{x} = \vec{0}$.

$(6) \Rightarrow (1)$. Given (6), the columns of A are independent by Theorem 5.2.2. Hence $dim(col\,A) = n$, and (1) follows.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-118

Theorem 5.4.4

The following are equivalent for an $m \times n$ matrix A:

1. $rank\,A = m$.
2. The columns of A span $\mathbb{R}^m$.
3. The rows of A are linearly independent in $\mathbb{R}^n$.
4. The $m \times m$ matrix AA^T is invertible.

5. $AC = I_m$ for some $n \times m$ matrix C.

6. The system $A\vec{x} = \vec{b}$ is consistent for every $\vec{b}$ in $\mathbb{R}^m$.

Proof:

(1) $\implies$ (2). By (1), $dim(colA = m$, so $colA = \mathbb{R}^m$ by Theorem 5.2.8.

(2) $\implies$ (3). By (2), $colA = \mathbb{R}^m$, so $rankA = m$. This means $dim(rowA) = m$. Since the m rows of A span $rowA$, they are independent by Theorem 5.2.7.

(3) $\implies$ (4). We have $rankA = m$ by (3), so the $n \times m$ matrix A^T has rank m. Hence applying Theorem 5.4.3 to A^T in place of A shows that $\left(A^T\right)^T A^T$ is invertible, proving (4).

(4) $\implies$ (5). Given (4), take $C = A^T \left(AA_T\right)^{-1}$ in (5).

(5) $\implies$ (6). Comparing columns in $AC = I_m$ gives $A\vec{c}_j = \vec{e}_j$ for each j, where $\vec{c}_j$ and $\vec{e}_j$ denote column j of C and I_m respectively. Given $\vec{b}$ in $\mathbb{R}^m$, write $\vec{b} = \sum_{j=1}^{m} r_j\vec{e}_j$, r_j in $\mathbb{R}$. Then $A\vec{x} = \vec{b}$ holds with $\vec{x} = \sum_{j=1}^{m} r_j\vec{c}_j$ as the reader can verify.

(6) $\implies$ (1). Given (6), the columns of A span $\mathbb{R}^m$ by Theorem 5.2.2. Thus $colA = \mathbb{R}^m$ and (1) follows.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-110

5.5 Similarity and Diagonalization

Similar Matrices

Definition 5.11 Similar Matrices

If A and B are $n \times n$ matrices, we say that A and B are **similar**, and write $A \sim B$, if $B = P^{-1}AP$ for some invertible matrix P.

Note that $A \sim B$ if and only if $B = QAQ^{-1}$ where Q is invertible (write $P^{-1} = Q$). The language of similarity is used throughout linear algebra. For example, a matrix A is diagonalizable if and only if it is similar to a diagonal matrix.

If $A \sim B$, then necessarily $B \sim A$. To see why, suppose that $B = P^{-1}AP$. Then $A = PBP^{-1} = Q^{-1}BQ$ where $Q = P^{-1}$ is invertible. This proves the second of the following properties of similarity:

1. $A \sim A$ *for all square matrices* A.

2. *If* $A \sim B$, *then* $B \sim A$.

(5.2) 3. *If* $A \sim B$ *and* $B \sim A$, *then* $A \sim C$.

These properties are often expressed by saying that the similarity relation $\sim$ is an **equivalence relation** on the set of $n \times n$ matrices. Here is an example showing how these properties are used.

Example 5.5.1

If A is similar to B and either A or B is diagonalizable, show that the other is also diagonalizable.

Solution:

We have $A \sim B$. Suppose that A is diagonalizable, say $A \sim D$ where D is diagonal. Since $B \sim A$ by (2) of (5.2), we have $B \sim A$ and $A \sim D$. Hence $B \sim D$ by (3) of (5.2), so B is diagonalizable too. An analogous argument works if we assume instead that B is diagonalizable.

Similarity is compatible with inverses, transposes, and powers:

If $A \sim B$ then $\quad A^{-1} \sim B^{-1}, \quad A^T \sim B^T, \quad$ and $\quad A^k \sim B^k$ for all integers $k \geq 1$.

Definition 5.12 Trace of a matrix

The **trace** $tr\,A$ of an $n \times n$ matrix A is defined to be the sum of the main diagonal elements of A.

In other words:

$$\text{If } A = [a_{ij}], \text{ then } tr\,A = a_{11} + a_{22} + \cdots + a_{nn}.$$

It is evident that $tr(A + B) = tr\,A + tr\,B$ and that $tr(cA) = c\,tr\,A$ holds for all $n \times n$ matrices A and B and all scalars c. The following fact is more surprising.

Lemma 5.5.1

Let A and B be $n \times n$ matrices. Then
$tr(AB) = tr(BA)$.

Proof:

Write $A = [a_{ij}]$ and $B = b_{ij}]$. For each i, the (i,i)-entry d_i of the matrix AB is given as follows: $d_i = a_{i1}b_{1i} + a_{i2}b_{2i} + \cdots + a_{in}b_{ni} = \sum_j a_{ij}b_{ji}$.
Hence

$$tr(AB) = d_1 + d_2 + \cdots + d_n = \sum_i d_i = \sum_i \left(\sum_j a_{ij}b_{ji} \right)$$

Similarly we have $tr(BA) = \sum_i (\sum_j b_{ij}a_{ji})$. Since these two double sums are the same, Lemma 5.5.1 is proved.

Theorem 5.5.1

If A and B are similar $n \times n$ matrices, then A and B have the same determinant, rank, trace, characteristic polynomial, and eigenvalues.

Proof:

Let $B = P^{-1}AP$ for some invertible matrix P. Then we have

$det B = det(P^{-1}) det A det P = det A$ because $det(P^{-1}) = 1/det P$

Similarly, $rank B = rank(P^{-1}AP) = rank A$ by Corollary 5.4.2. Next Lemma 5.5.1 gives

$$tr(P^{-1}AP) = tr\left[P^{-1}(AP)\right] = tr\left[(AP)P^{-1}\right] = tr A$$

As to the characteristic polynomial,

$$
\begin{aligned}
c_B(x) = det(xI - B) &= det\{x(P^{-1}IP) - P^{-1}AP\} \\
&= det\{P^{-1}(xI - A)P\} \\
&= det(xI - A) \\
&= c_A(x)
\end{aligned}
$$

Finally, this shows that A and B have the same eigenvalues because the eigenvalues of a matrix are the roots of its characteristic polynomial.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-119

Sharing the five properties in Theorem 5.5.1 does not guarantee that two matrices are similar. The matrices $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ have the same determinant, rank, trace, characteristic polynomial, and eigenvalues, but they are not similar because $P^{-1}IP = I$ for any invertible matrix P.

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-124

Diagonalization Revisited

Recall that a square matrix A is **diagonalizable** if there exists an invertible matrix P such that $P^{-1}AP = D$ is a diagonal matrix, that is if A is similar to a diagonal matrix D\index{diagonal matrices}. Unfortunately, not all matrices are diagonalizable, for

example $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Determining whether A is diagonalizable is closely related to the eigenvalues and eigenvectors of A. Recall that a number λ is called an **eigenvalue** of A if $A\vec{x} = \lambda\vec{x}$ for some nonzero column $\vec{x}$ in $\mathbb{R}^n$, and any such nonzero vector $\vec{x}$ is called an **eigenvector** of A corresponding to λ (or simply a λ-eigenvector of A). The eigenvalues and eigenvectors of A are closely related to the **characteristic polynomial** $c_A(x)$ of A, defined by

$$c_A(x) = det(xI - A)$$

If A is $n \times n$ this is a polynomial of degree n, and its relationship to the eigenvalues is given in the following theorem.

Theorem 5.5.2

Let A be an $n \times n$ matrix.

1. The eigenvalues λ of A are the roots of the characteristic polynomial $c_A(x)$ of A.
2. The λ-eigenvectors $\vec{x}$ are the nonzero solutions to the homogeneous system

$$(\lambda I - A)\vec{x} = \vec{0}$$

of linear equations with $\lambda I - A$ as coefficient matrix.

The next theorem will show us the condition when a square matrix is diagonalizable.

Theorem 5.5.3

Let A be an $n \times n$ matrix.

1. A is diagonalizable if and only if $\mathbb{R}^n$ has a basis $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\}$ consisting of eigenvectors of A.
2. When this is the case, the matrix $P = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \end{bmatrix}$ is invertible and $P^{-1}AP = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ where, for each i, λ_i is the eigenvalue of A corresponding to x_i

.

The next result is a basic tool for determining when a matrix is diagonalizable. It reveals an important connection between eigenvalues and linear independence: Eigenvectors corresponding to distinct eigenvalues are necessarily linearly independent.

Theorem 5.5.4

Let $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k$ be eigenvectors corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ of an $n \times n$ matrix A. Then $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ is a linearly independent set.

Theorem 5.5.5

If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable

Show that $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 3 \\ -1 & 1 & 0 \end{bmatrix}$ is diagonalizable.

Solution:

A routine computation shows that $c_A(x) = (x - 1)(x - 3)(x + 1)$ and so has distinct eigenvalues $1, 3$, and -1. Hence Theorem 5.5.5 applies.

Definition 5.1.3 Eigenspace of a Matrix

If λ is an eigenvalue of an $n \times n$ matrix A, define the **eigenspace** of A corresponding to λ by

$$E_\lambda(A) = \{\vec{x} \text{ in } \mathbb{R}^n \mid A\vec{x} = \lambda\vec{x}\}$$

This is a subspace of $\mathbb{R}^n$ and the eigenvectors corresponding to λ are just the nonzero vectors in $E_\lambda(A)$. In fact $E_\lambda(A)$ is the null space of the matrix $(\lambda I - A)$:

$$E_\lambda(A) = \{\vec{x} \mid (\lambda I - A)\vec{x} = \vec{0}\} = null(\lambda I - A)$$

The basic solutions of the homogeneous system $(\lambda I - A)\vec{x} = \vec{0}$ given by the gaussian algorithm form a basis for $E_\lambda(A)$. In particular

(5.5)

$dim E_\lambda(A)$ is the number of basic solutions $\vec{x}$ of $(\lambda I - A)\vec{x} = \vec{0}$

Now recall that the **multiplicity** of an eigenvalue λ of A is the number of times λ occurs as a root of the characteristic polynomial $c_A(x)$ of A. In other words, the multiplicity of λ is the largest integer $m \geq 1$ such that

$$c_A(x) = (x - \lambda)^m g(x)$$

for some polynomial $g(x)$. Because of (5.5), a square matrix is diagonalizable if and only if the multiplicity of each eigenvalue λ equals $dim[E_\lambda(A)]$. We are going to prove this, and the proof requires the following result which is valid for any square matrix, diagonalizable or not.

Lemma 5.5.3

Let λ be an eigenvalue of multiplicity m of a square matrix A. Then $dim\left[E_\lambda(A)\right] \le m$.

It turns out that this characterizes the diagonalizable $n \times n$ matrices A for which $c_A(x)$ **factors completely** over $\mathbb{R}$. By this we mean that $c_A(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$, where the λ_i are real numbers (not necessarily distinct); in other words, every eigenvalue of A is real. This need not happen (consider $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$), which leads us to the general conclusion regarding when a square matrix is diagonalizable.

Theorem 5.5.6

The following are equivalent for a square matrix A for which $c_A(x)$ factors completely.

1. A is diagonalizable.
2. $dim[E_\lambda(A)]$ equals the multiplicity of λ for every eigenvalue λ of the matrix A

An interactive H5P element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=76#h5p-123

Example 5.5.5

If $A = \begin{bmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{bmatrix}$ and

$$B = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 1 & -2 \\ -1 & 0 & -2 \end{bmatrix} \text{ show that } A \text{ is diagonalizable}$$

but B is not.

Solution:

We have $c_A(x) = (x+3)^2(x-1)$ so the eigenvalues are $\lambda_1 = -3$ and $\lambda_2 = 1$. The corresponding eigenspaces are $E_{\lambda_1}(A) = span\{\vec{x}_1, \vec{x}_2\}$ and $E_{\lambda_2}(A) = span\{\vec{x}_3\}$ where

$$\vec{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \vec{x}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \vec{x}_3 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

as the reader can verify. Since $\{\vec{x}_1, \vec{x}_2\}$ is independent, we have $dim(E_{\lambda_1}(A)) = 2$ which is the multiplicity of λ_1. Similarly, $dim(E_{\lambda_2}(A)) = 1$ equals the multiplicity of λ_2. Hence A is diagonalizable
by 5.5.6, and a diagonalizing matrix is $P = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \vec{x}_3 \end{bmatrix}$.

Turning to B, $c_B(x) = (x+1)^2(x-3)$ so the eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = 3$. The corresponding eigenspaces are $E_{\lambda_1}(B) = span\{\vec{y}_1\}$ and $E_{\lambda_2}(B) = span\{\vec{y}_2\}$ where

$$\vec{y}_1 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \vec{y}_2 = \begin{bmatrix} 5 \\ 6 \\ -1 \end{bmatrix}$$

Here $dim(E_{\lambda_1}(B)) = 1$ is smaller than the multiplicity of λ_1, so the matrix B is not diagonalizable, again by Theorem 5.5.6.

The fact that $dim(E_{\lambda_1}(B)) = 1$ means that there is no possibility of finding three linearly independent eigenvectors.

> 📇 *An interactive H5P element has been excluded from this version of the text. You can view it online here:*
> *https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=76#h5p-121*

> 📇 *An interactive H5P element has been excluded from this version of the text. You can view it online here:*
> *https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=76#h5p-125*

> 📇 *An interactive H5P element has been excluded from this version of the text. You can view it online here:*
> *https://ecampusontario.pressbooks.pub/ linearalgebrautm/?p=76#h5p-126*