Linear Algebra UTM

Linear Algebra UTM

XINLI WANG

Contents

Introduction

This book is an adaptation from the textbook: *Linear Algebra with Applications* by W. Keith Nicholson. The book can be found here: https://lyryx.com/linear-algebra-applications/

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

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

Given a linear equation $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$, a sequence $s_1, s_2, \ldots, s_n$ of n numbers is called a **solution** to the equation if

$$a_1s_1 + a_2s_2 + \cdots + a_ns_n = b$$

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**.

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

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

$$\begin{array}{rrrrl} x_1 - & 2x_2 & +3x_3 & +x_4 = & -3 \\ 2x_1 - & x_2 & +3x_3 & -x_4 = & 0 \end{array}$$

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, $\dots$ are given in terms of new independent variables s, t, etc.).

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

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

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

of three equations in four variables. The array of numbers

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

$$\left[\begin{array}{rrr|r} 1 & 1 & 1 & 1 \\ 0 & 1 & -2 & -2 \\ 0 & -1 & -5 & -6 \end{array}\right]$$

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

$$\left[\begin{array}{rrr|r} 1 & 0 & 3 & 3 \\ 0 & 1 & -2 & -2 \\ 0 & 0 & -7 & -8 \end{array}\right]$$

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

$$\left[\begin{array}{ccc|c} 1 & 0 & 3 & 3 \\ 0 & 1 & -2 & -2 \\ 0 & 0 & 1 & \frac{8}{7} \end{array}\right]$$

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

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -\frac{3}{7} \\ 0 & 1 & 0 & \frac{2}{7} \\ 0 & 0 & 1 & \frac{8}{7} \end{array}\right]$$

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 or media element has been excluded from this version of the text. You can view it online

here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=5

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

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 1 s from left to right, using each of them in turn to create zeros below it. Here is one example.

Example 1.2.2 Solve the following system of equations.

$$\begin{aligned} 3x + y - \ 4z &= -1 \\ x \qquad + 10z &= \ \ \ 5 \\ 4x + y + \ 6z &= \ \ \ 1 \end{aligned}$$

Solution:

The corresponding augmented matrix is

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

Create the first leading one by interchanging rows 1 and 2

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

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

$$\left[\begin{array}{ccr|r} 1 & 0 & 10 & 5 \\ 0 & 1 & -34 & -16 \\ 0 & 1 & -34 & -19 \end{array}\right]$$

Now subtract row 2 from row 3 to obtain

$$\left[\begin{array}{ccr|r} 1 & 0 & 10 & 5 \\ 0 & 1 & -34 & -16 \\ 0 & 0 & 0 & -3 \end{array}\right]$$

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=5

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 $s, t, \dots$, 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 $\left[\begin{array}{cccccc} 0 & 0 & 0 & \cdots & 0 & 1 \end{array}\right]$ 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 or media element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=5

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 $A = \left[\begin{array}{rrr} 1 & -1 & 4 \\ 2 & -1 & 2 \end{array} \right]$ can be carried (by one row operation) to the row-echelon matrix $\left[\begin{array}{rrr} 1 & -1 & 4 \\ 0 & 1 & -6 \end{array} \right]$, and then by another row operation to the (reduced) row-echelon matrix $\left[\begin{array}{rrr} 1 & 0 & -2 \\ 0 & 1 & -6 \end{array} \right]$. 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.

Definition 1.4 Rank of a matrix

The **rank** of matrix A is the number of leading 1s in any row-echelon matrix to which A can be carried by row operations.

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

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

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 or media element has been excluded from this version of the text. You can view it online here: *https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=5*

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.

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.

1.3 Homogeneous equations

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

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$$

Clearly $x_1 = 0, x_2 = 0, \dots, x_n = 0$ 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.

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

Solution:

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

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

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 $n > m$ m" title="Rendered by QuickLaTeX.com" height="11" width="49" style="vertical-align: 0px;">, 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 $m = 2 = n$.)

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

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 $x_1, x_2, \dots, x_n$, 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 $a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$, and suppose that $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ are solutions. Then $a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$ and $a_1y_1 + a_2y_2 + \cdots + a_ny_n = 0$.

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

$$\begin{aligned}
a_1(sx_1 + ty_1) &+ a_2(sx_2 + ty_2) + \cdots + a_n(sx_n + ty_n) \\
&= [a_1(sx_1) + a_2(sx_2) + \cdots + a_n(sx_n)] + [a_1(ty_1) + a_2(ty_2) + \cdots + a_n(ty_n)] \\
&= s(a_1x_1 + a_2x_2 + \cdots + a_nx_n) + t(a_1y_1 + a_2y_2 + \cdots + a_ny_n) \\
&= s(0) + t(0) \\
&= 0
\end{aligned}$$

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

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

$$\left[\begin{array}{rrrr|r} 1 & -2 & 3 & -2 & 0 \\ -3 & 6 & 1 & 0 & 0 \\ -2 & 4 & 4 & -2 & 0 \end{array}\right] \rightarrow \left[\begin{array}{rrrr|r} 1 & -2 & 0 & -\frac{1}{5} & 0 \\ 0 & 0 & 1 & -\frac{3}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

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

$$\left[\begin{array}{rrrrr|r} 1 & -3 & 0 & 2 & 2 & 0 \\ -2 & 6 & 1 & 2 & -5 & 0 \\ 3 & -9 & -1 & 0 & 7 & 0 \\ -3 & 9 & 2 & 6 & -8 & 0 \end{array}\right] \rightarrow \left[\begin{array}{rrrrr|r} 1 & -3 & 0 & 2 & 2 & 0 \\ 0 & 0 & 1 & 6 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

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

$$x = \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array}\right] = \left[\begin{array}{c} 3r - 2s - 2t \\ r \\ -6s + t \\ s \\ t \end{array}\right] = r\left[\begin{array}{r} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{array}\right] + s\left[\begin{array}{r} -2 \\ 0 \\ -6 \\ 1 \\ 0 \end{array}\right] + t\left[\begin{array}{r} -2 \\ 0 \\ 1 \\ 0 \\ 1 \end{array}\right]$$

Hence basic solutions are

$$x_1 = \left[\begin{array}{r} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{array}\right], \quad x_2 = \left[\begin{array}{r} -2 \\ 0 \\ -6 \\ 1 \\ 0 \end{array}\right], \quad x_3 = \left[\begin{array}{r} -2 \\ 0 \\ 1 \\ 0 \\ 1 \end{array}\right]$$

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

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,

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

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

, 2×2, 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 $A = \left[\begin{array}{cc} a & b \\ c & d \end{array} \right]$, $B = \left[\begin{array}{rrr} 1 & 2 & -1 \\ 3 & 0 & 1 \end{array} \right]$ and

$C = \left[\begin{array}{rr} 1 & 0 \\ -1 & 2 \end{array} \right]$

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:

$$\left[\begin{array}{cc} a & b \\ c & d \end{array} \right] = \left[\begin{array}{rr} 1 & 0 \\ -1 & 2 \end{array} \right]$$

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

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

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 $A = [a_{ij}]$ and $B = [b_{ij}]$, this takes the form

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

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 $a=3, b=-1, c=0$.

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

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

$$A+B=\left[\begin{array}{c} a_{ij}+b_{ij} \end{array}\right]=\left[\begin{array}{c} b_{ij}+a_{ij} \end{array}\right]=B+A$$

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

$B = \left[\begin{array}{rrr} 1 & -1 & 1 \\ -2 & 0 & 6 \end{array} \right], C = \left[\begin{array}{rrr} 1 & 0 & -2 \\ 3 & 1 & 1 \end{array} \right]$.

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

Solution:

$$
\begin{aligned}
-A &= \left[\begin{array}{rrr} -3 & 1 & 0 \\ -1 & -2 & 4 \end{array}\right] \\
A-B &= \left[\begin{array}{ccc} 3-1 & -1-(-1) & 0-1 \\ 1-(-2) & 2-0 & -4-6 \end{array}\right] = \left[\begin{array}{rrr} 2 & 0 & -1 \\ 3 & 2 & -10 \end{array}\right] \\
A+B-C &= \left[\begin{array}{ccc} 3+1-1 & -1-1-0 & 0+1-(-2) \\ 1-2-3 & 2+0-1 & -4+6-1 \end{array}\right] = \left[\begin{array}{rrr} 3 & -2 & 3 \\ -4 & 1 & 1 \end{array}\right]
\end{aligned}
$$

Example 2.1.5

Solve

$$\left[\begin{array}{rr} 3 & 2 \\ -1 & 1 \end{array}\right] + X = \left[\begin{array}{rr} 1 & 0 \\ -1 & 2 \end{array}\right]$$

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

$$\left[\begin{array}{rr} 3 & 2 \\ -1 & 1 \end{array}\right] + X = \left[\begin{array}{rr} 1 & 0 \\ -1 & 2 \end{array}\right]$$

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 $A = \left[\begin{array}{rrr} 3 & -1 & 4 \\ 2 & 0 & 1 \end{array} \right]$
and $B = \left[\begin{array}{rrr} 1 & 2 & -1 \\ 0 & 3 & 2 \end{array} \right]$,
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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 $(A + B) + (C + D)$ and $A + [B + (C + D)]$. 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.

$$\begin{aligned}
2(A+3C)-3(2C-B)-3\left[2(2A+B-4C)-4(A-2C)\right] \\
&= 2A+6C-6C+3B-3\left[4A+2B-8C-4A+8C\right] \\
&= 2A+3B-3\left[2B\right] \\
&= 2A-3B
\end{aligned}$$

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

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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

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:

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

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

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

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

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

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

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 $e = e'$.

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 $A = \frac{1}{3}(0) = 0$.

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

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

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

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 $(a_1, a_2, \dots, a_n)$ 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,

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

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 $(r_1, r_2, \dots, r_n)$ 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 **x** or **v**. For example, an $m \times n$ matrix A will be written as a row of columns:

$$A = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{array} \right] \text{ where } \mathbf{a}_j \text{ denotes column } j \text{ of } A \text{ for each } j.$$

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:

$$\begin{aligned} ax_1 + \ bx_2 + \ cx_3 &= b_1 \\ a'x_1 + b'x_2 + c'x_3 &= b_1 \end{aligned} \tag{2.2}$$

and let $A = \left[\begin{array}{ccc} a & b & c \\ a' & b' & c' \end{array} \right]$, $\mathbf{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} \right]$, $\mathbf{b} = \left[\begin{array}{c} b_1 \\ b_2 \end{array} \right]$

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

$$\left[\begin{array}{rrr} ax_1 + & bx_2 + & cx_3 \\ a'x_1 + & b'x_2 + & c'x_3 \end{array} \right] = \left[\begin{array}{c} b_1 \\ b_2 \end{array} \right]$$

which in turn can be written as follows:

$$x_1 \left[\begin{array}{c} a \\ a' \end{array} \right] + x_2 \left[\begin{array}{c} b \\ b' \end{array} \right] + x_3 \left[\begin{array}{c} c \\ c' \end{array} \right] = \left[\begin{array}{c} b_1 \\ b_2 \end{array} \right]$$

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

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

Example 2.2.1

Write the system

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

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 $x_1, x_2, \dots, x_n$ 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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

Example 2.2.2

If $A = \left[\begin{array}{rrrr} 2 & -1 & 3 & 5 \\ 0 & 2 & -3 & 1 \\ -3 & 4 & 1 & 2 \end{array} \right]$ and $\mathbf{x} = \left[\begin{array}{r} 2 \\ 1 \\ 0 \\ -2 \end{array} \right]$, compute $A\mathbf{x}$.

Solution:

By Definition 2.5:
$A\mathbf{x} = 2 \left[\begin{array}{r} 2 \\ 0 \\ -3 \end{array} \right] + 1 \left[\begin{array}{r} -1 \\ 2 \\ 4 \end{array} \right] + 0 \left[\begin{array}{r} 3 \\ -3 \\ 1 \end{array} \right] - 2 \left[\begin{array}{r} 5 \\ 1 \\ 2 \end{array} \right] = \left[\begin{array}{r} -7 \\ 0 \\ -6 \end{array} \right]$

.

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

Example 2.2.3

Given columns $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, and $\mathbf{a}_4$ in $\mathbb{R}^3$, 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

$$\mathbf{x} = \left[\begin{array}{r} 2 \\ -3 \\ 5 \\ 1 \end{array} \right].$$

Hence Definition 2.5 gives

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

where $A = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{array} \right]$ is the matrix with $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, and $\mathbf{a}_4$ as its columns.

Example 2.2.4

Let $A = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \end{array} \right]$ be the 3×4 matrix given in terms of its columns

$$\mathbf{a}_1 = \left[\begin{array}{r} 2 \\ 0 \\ -1 \end{array} \right], \mathbf{a}_2 = \left[\begin{array}{r} 1 \\ 1 \\ 1 \end{array} \right], \mathbf{a}_3 = \left[\begin{array}{r} 3 \\ -1 \\ -3 \end{array} \right], \text{ and}$$

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

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

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

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

Example 2.2.6

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 $A = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{array} \right]$ and $B = \left[\begin{array}{cccc} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{array} \right]$ be given in terms of their columns. Since adding two matrices is the same as adding their columns, we have

$$A+B = \left[\begin{array}{cccc} \mathbf{a}_1+\mathbf{b}_1 & \mathbf{a}_2+\mathbf{b}_2 & \cdots & \mathbf{a}_n+\mathbf{b}_n \end{array} \right]$$

If we write $\mathbf{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array} \right]$.

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.

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.

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

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.)

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

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 $\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$. Write $\text{rank}A = r$.

1. $\text{rank}\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right]$ is either r or $r+1$.
2. The system is consistent if and only if $\text{rank}\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] = r$.
3. The system is inconsistent if and only if $\text{rank}\left[\begin{array}{c|c} A & \mathbf{b} \end{array}\right] = r+1$.

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 $(a_1, a_2, \dots, a_n)$ and $(b_1, b_2, \dots, b_n)$ 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

$$\mathbf{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} \right] \quad \text{and} \quad A = \left[\begin{array}{cccc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{array} \right]$$

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.

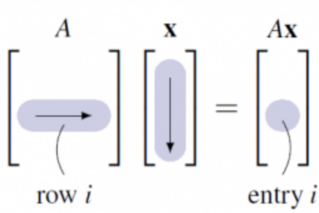
An interactive or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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

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

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

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

An interactive or media element has been excluded from this version of the text. You can view it online here: *https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66*

Example 2.2.9

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

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

Solution:

Write $A = \left[\begin{array}{rrrrr} 5 & -1 & 2 & 1 & -3 \\ 1 & 1 & 3 & -5 & 2 \\ -1 & 1 & -2 & 0 & -3 \end{array}\right]$, $\mathbf{b} = \left[\begin{array}{r} 8 \\ -2 \\ 0 \end{array}\right]$, and $\mathbf{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array}\right]$. Then the dot product rule gives $A\mathbf{x} = \left[\begin{array}{rcrcrcrcr} 5x_1 & - & x_2 & + & 2x_3 & + & x_4 & - & 3x_5 \\ x_1 & + & x_2 & + & 3x_3 & - & 5x_4 & + & 2x_5 \\ -x_1 & + & x_2 & - & 2x_3 & & & - & 3x_5 \end{array}\right]$, 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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 $n > 2$ 2" title="Rendered by QuickLaTeX.com" height="12" width="42" style="vertical-align: 0px;">, the $\textbf{identity matrix}$ 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

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

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 $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$. 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 $j = 1, 2, \dots, n$.

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 2.2.12will 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 $A = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{array} \right]$ and $B = \left[\begin{array}{cccc} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{array} \right]$ 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 $A = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{array} \right]$ 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

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

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.

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:

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

Thus the product matrix AB is given in terms of its columns $A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_n$: 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

$$A(B\vec{x}) = \left[\begin{array}{cccc} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_n \end{array} \right] \vec{x} = (AB)\vec{x}$$

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 $A = \left[\begin{array}{ccc} 2 & 3 & 5 \\ 1 & 4 & 7 \\ 0 & 1 & 8 \end{array} \right]$

and

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

Solution:

The columns of B are

$\vec{b}_1 = \begin{bmatrix} 8 \\ 7 \\ 6 \end{bmatrix}$ and $\vec{b}_2 = \begin{bmatrix} 9 \\ 2 \\ 1 \end{bmatrix}$, 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.

Theorem 2.3.2 Dot Product Rule

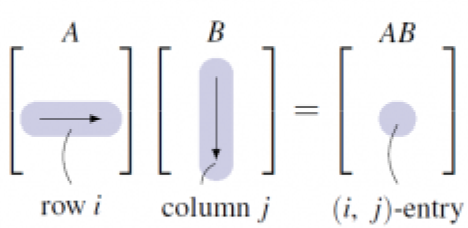
Let A and B be matrices of sizes $m \times n$ and $n \times k$, respectively. Then the (i, j)-entry of AB is the dot product of row i of A with column j of B.

Proof:

Write $B = \left[\begin{array}{cccc} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \end{array} \right]$ 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):

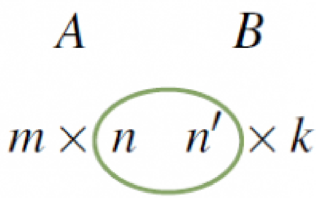
Go *across* row i of A, and *down* column j of B, multiply corresponding entries, and add the results.



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.



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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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.

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

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

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

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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:
$A^2 = \begin{bmatrix} 6 & 9 \\ -4 & -6 \end{bmatrix} \begin{bmatrix} 6 & 9 \\ -4 & -6 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, 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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 $A = \left[\begin{array}{cccc} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{array} \right]$ where $\vec{a}_j$ is column j of A. Then Definition 2.9 and Example 2.2.1 give

$$IA = \left[\begin{array}{cccc} I\vec{a}_1 & I\vec{a}_2 & \cdots & I\vec{a}_n \end{array} \right] = \left[\begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array} \right] = A$$

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:

$$AI = A\left[\begin{array}{cccc} \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \end{array} \right] = \left[\begin{array}{cccc} A\vec{e}_1 & A\vec{e}_2 & \cdots & A\vec{e}_n \end{array} \right] = \left[\begin{array}{cccc} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{array} \right] = A$$

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

Theorem 2.3.3

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 $C = \left[\begin{array}{cccc} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_k \end{array} \right]$ in terms of its columns, then $BC = \left[\begin{array}{cccc} B\vec{c}_1 & B\vec{c}_2 & \cdots & B\vec{c}_k \end{array} \right]$ by Definition 2.9, so

$$
\begin{aligned}
A(BC) &= \left[\begin{array}{cccc} A(B\vec{c}_1) & A(B\vec{c}_2) & \cdots & A(B\vec{c}_k) \end{array}\right] && \text{Definition 2.9} \\
&= \left[\begin{array}{cccc} (AB)\vec{c}_1 & (AB)\vec{c}_2 & \cdots & (AB)\vec{c}_k) \end{array}\right] && \text{Theorem 2.3.1} \\
&= (AB)C && \text{Definition 2.9}
\end{aligned}
$$

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 $A = \left[\begin{array}{cccc} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{array}\right]$ in terms of its columns, we get

$$
\begin{aligned}
(B+C)A &= \left[\begin{array}{cccc} (B+C)\vec{a}_1 & (B+C)\vec{a}_2 & \cdots & (B+C)\vec{a}_n \end{array}\right] && \text{Definition 2.9} \\
&= \left[\begin{array}{cccc} B\vec{a}_1 + C\vec{a}_1 & B\vec{a}_2 + C\vec{a}_2 & \cdots & B\vec{a}_n + C\vec{a}_n \end{array}\right] && \text{Theorem 2.2.2} \\
&= \left[\begin{array}{cccc} B\vec{a}_1 & B\vec{a}_2 & \cdots & B\vec{a}_n \end{array}\right] + \left[\begin{array}{cccc} C\vec{a}_1 & C\vec{a}_2 & \cdots & C\vec{a}_n \end{array}\right] && \text{Adding Columns} \\
&= BA + CA && \text{Definition 2.9}
\end{aligned}
$$

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

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

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

Example 2.3.9 and Example 2.3.10 below show how we can use the properties in Theorem 2.3.2to deduce other facts about matrix multiplication. Matrices A and B are said to **commute** if $AB = BA$.

Example 2.3.9

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 $BC = CB$.

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

Example 2.3.10

Show that $AB = BA$ if and only if $(A - B)(A + B) = A^2 - B^2$.

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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 $\textbf{invertible matrix}$.

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 $B = \left[\begin{array}{cc} -1 & 1 \\ 1 & 0 \end{array} \right]$ is an inverse of $A = \left[\begin{array}{cc} 0 & 1 \\ 1 & 1 \end{array} \right]$.

Solution:

Compute AB and BA.

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

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

Example 2.4.2

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 $A = \left[\begin{array}{cc} 0 & -1 \\ 1 & -1 \end{array} \right]$, 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

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

Hence $A^3 = I$, as asserted. This can be written as $A^2A = I = AA^2$, 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

$$A^{-1} = \frac{1}{detA} adjA$$

Solution:

For convenience, write $e = detA = ad - bc$ and $B = adjA = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. 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 $detA = 2 \cdot 8 - 4 \cdot (-3) = 28 \neq 0$. Hence A is invertible and $A^{-1} = \frac{1}{detA} adjA = \frac{1}{28} \begin{bmatrix} 8 & -4 \\ 3 & 2 \end{bmatrix}$, 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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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

$$\begin{aligned} A^{-1}A\vec{x} &= A^{-1}\vec{b} \\ I\vec{x} &= A^{-1}\vec{b} \\ \vec{x} &= A^{-1}\vec{b} \end{aligned}$$

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 anysolution, 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

$$\begin{cases} 5x_1 - 3x_2 = -4 \\ 7x_1 + 4x_2 = \quad 8 \end{cases}.$$

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

$$A^{-1} = \frac{1}{41}\left[\begin{array}{rr} 4 & 3 \\ -7 & 5 \end{array}\right]$$

by Example 2.4.4. Thus Theorem 2.4.2 gives

$$\vec{x} = A^{-1}\vec{b} = \frac{1}{41}\left[\begin{array}{rr} 4 & 3 \\ -7 & 5 \end{array}\right]\left[\begin{array}{r} -4 \\ 8 \end{array}\right] = \frac{1}{41}\left[\begin{array}{r} 8 \\ 68 \end{array}\right]$$

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

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

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

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:

$$\left[\begin{array}{cc} A & I \end{array}\right] \rightarrow \left[\begin{array}{cc} I & A^{-1} \end{array}\right]$$

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

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

Solution:

Apply elementary row operations to the double matrix

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

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 $A^{-1} = \frac{1}{2}\begin{bmatrix} -3 & -3 & 11 \\ 1 & 1 & -3 \\ 1 & -1 & -1 \end{bmatrix}$, 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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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

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:

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

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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

Theorem 2.4.4

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 $A_1, A_2, \dots, A_k$ are all invertible, so is their product $A_1A_2\cdots A_k$, and

$$(A_1A_2\cdots A_k)^{-1} = A_k^{-1}\cdots A_2^{-1}A_1^{-1}.$$

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 $k > 2$ 2" title="Rendered by QuickLaTeX.com" height="13" width="41" style="vertical-align: 0px;">, assume inductively that $(A_1A_2\cdots A_{k-1})^{-1} = A_{k-1}^{-1}\cdots A_2^{-1}A_1^{-1}$. We apply this fact together with property 3 as follows:

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

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 $left\text{-}multiplied$ by a matrix A to yield $AB = AC$. Similarly, $right\text{-}multiplication$ gives $BA = CA$. However, we cannot mix the two: If $B = C$, it need notbe the case that $AB = CA$ even if A is invertible, for example, $A = \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right], B = \left[\begin{array}{cc} 0 & 0 \\ 1 & 0 \end{array}\right] = C$.

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

Example 2.4.10

Find A if $(A^T - 2I)^{-1} = \left[\begin{array}{rr} 2 & 1 \\ -1 & 0 \end{array} \right]$.

Solution:

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

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

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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 $\left[\begin{array}{c|c} A & \vec{0} \end{array}\right]$ of the system $A\vec{x} = \vec{0}$. Then $\left[\begin{array}{c|c} A & \vec{0} \end{array}\right] \to \left[\begin{array}{c|c} R & \vec{0} \end{array}\right]$ is the reduced form, and $\left[\begin{array}{c|c} R & \vec{0} \end{array}\right]$ 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 $\left[\begin{array}{c|c} A & \vec{b} \end{array}\right]$ 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 $\left[\begin{array}{c|c} A & \vec{b} \end{array}\right] \to \left[\begin{array}{c|c} I_n & \vec{c} \end{array}\right]$ 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 $I_n = \left[\begin{array}{cccc} \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \end{array}\right]$ where

$\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$ are the columns of I_n. For each \newline $j = 1, 2, \dots, n$, 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 $C = \left[\begin{array}{cccc} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{array} \right]$ be the $n \times n$ matrix with these matrices $\vec{c}_j$ as its columns. Then Definition 2.9 gives (5):

$$AC = A \left[\begin{array}{cccc} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{array} \right] = \left[\begin{array}{cccc} A\vec{c}_1 & A\vec{c}_2 & \cdots & A\vec{c}_n \end{array} \right] = \left[\begin{array}{cccc} \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \end{array} \right] = I_n$$

(5) $\Longrightarrow$ (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) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (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) $\Longrightarrow$ (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

$$\left[\begin{array}{rrr} 1 & 2 & 1 \\ 1 & 1 & 1 \end{array}\right]\left[\begin{array}{rr} -1 & 1 \\ 1 & -1 \\ 0 & 1 \end{array}\right] = I_2 \quad \text{but} \quad \left[\begin{array}{rr} -1 & 1 \\ 1 & -1 \\ 0 & 1 \end{array}\right]\left[\begin{array}{rrr} 1 & 2 & 1 \\ 1 & 1 & 1 \end{array}\right] \neq I_3$$

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=66

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:

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

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 $detA = det\,[a] = a$ 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 $detA$ 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

$$detA = aei + bfg + cdh - ceg - afh - bdi \tag{3.1}$$

and observe that $detA \neq 0$ because $detA = w \neq 0$ (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:

$$\begin{aligned} detA = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= aei + bfg + cdh - ceg - afh - bdi \\ &= a(ei - fh) - b(di - fg) + c(dh - eg) \\ &= a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix} \end{aligned}$$

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.

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

Example 3.1.1

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

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70

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

$$A = \left[\begin{array}{rrr} 3 & 4 & 5 \\ 1 & 7 & 2 \\ 9 & 8 & -6 \end{array} \right].$$

Solution:
The cofactor expansion along the first row is as follows:

$$\begin{aligned} detA &= 3c_{11}(A) + 4c_{12}(A) + 5c_{13}(A) \\ &= 3\begin{vmatrix} 7 & 2 \\ 8 & -6 \end{vmatrix} - 4\begin{vmatrix} 1 & 2 \\ 9 & -6 \end{vmatrix} + 3\begin{vmatrix} 1 & 7 \\ 9 & 8 \end{vmatrix} \\ &= 3(-58) - 4(-24) + 5(-55) \\ &= -353 \end{aligned}$$

Note that the signs alternate along the row (indeed along any row or column). Now we compute $detA$ by expanding along the first column.

$$\begin{aligned} detA &= 3c_{11}(A) + 1c_{21}(A) + 9c_{31}(A) \\ &= 3\begin{vmatrix} 7 & 2 \\ 8 & -6 \end{vmatrix} - \begin{vmatrix} 4 & 5 \\ 8 & -6 \end{vmatrix} + 9\begin{vmatrix} 4 & 5 \\ 7 & 2 \end{vmatrix} \\ &= 3(-58) - (-64) + 9(-27) \\ &= -353 \end{aligned}$$

The reader is invited to verify that $detA$ 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.

Example 3.1.4

Compute $detA$ where

$$A = \left[\begin{array}{rrrr} 3 & 0 & 0 & 0 \\ 5 & 1 & 2 & 0 \\ 2 & 6 & 0 & -1 \\ -6 & 3 & 1 & 0 \end{array} \right].$$

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

$$\begin{aligned} det A &= 3c_{11}(A) + 0c_{12}(A) + 0c_{13}(A) + 0c_{14}(A) \\ &= 3 \left| \begin{array}{rrr} 1 & 2 & 0 \\ 6 & 0 & -1 \\ 3 & 1 & 0 \end{array} \right| \end{aligned}$$

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

$$
\begin{aligned}
det A &= 3\left(0\begin{vmatrix} 6 & 0 \\ 3 & 1 \end{vmatrix} - (-1)\begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} + 0\begin{vmatrix} 1 & 2 \\ 6 & 0 \end{vmatrix}\right) \\
&= 3[0 + 1(-5) + 0] \\
&= -15
\end{aligned}
$$

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

$$A = \left[\begin{array}{rrr} 1 & -1 & 3 \\ 1 & 0 & -1 \\ 2 & 1 & 6 \end{array} \right].$$

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

$$det A = \left| \begin{array}{rrr} 1 & -1 & 3 \\ 1 & 0 & -1 \\ 2 & 1 & 6 \end{array} \right| = \left| \begin{array}{rrr} 1 & -1 & 4 \\ 1 & 0 & 0 \\ 2 & 1 & 8 \end{array} \right| = - \left| \begin{array}{rr} -1 & 4 \\ 1 & 8 \end{array} \right| = 12$$

where we expanded the second 3×3 matrix along row 2.

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

Example 3.1.6

If $det \left[\begin{array}{ccc} a & b & c \\ p & q & r \\ x & y & z \end{array} \right] = 6,$
evaluate $det A$ where
$A = \left[\begin{array}{ccc} a+x & b+y & c+z \\ 3x & 3y & 3z \\ -p & -q & -r \end{array} \right].$

Solution:

First take common factors out of rows 2 and 3.

$$detA = 3(-1)det \begin{bmatrix} a+x & b+y & c+z \\ x & y & z \\ p & q & r \end{bmatrix}$$

Now subtract the second row from the first and interchange the last two rows.

$$detA = -3det \begin{bmatrix} a & b & c \\ x & y & z \\ p & q & r \end{bmatrix} = 3det \begin{bmatrix} a & b & c \\ p & q & r \\ x & y & z \end{bmatrix} = 3 \cdot 6 = 18$$

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

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

$$A = \left[\begin{array}{ccc} 1 & x & x \\ x & 1 & x \\ x & x & 1 \end{array} \right].$$

Solution:

To evaluate $det A$, first subtract x times row 1 from rows 2 and 3.

$$det A = \left| \begin{array}{ccc} 1 & x & x \\ x & 1 & x \\ x & x & 1 \end{array} \right| = \left| \begin{array}{ccc} 1 & x & x \\ 0 & 1-x^2 & x-x^2 \\ 0 & x-x^2 & 1-x^2 \end{array} \right| = \left| \begin{array}{cc} 1-x^2 & x-x^2 \\ x-x^2 & 1-x^2 \end{array} \right|$$

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 $x = -\frac{1}{2}$.

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

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 $(a_2 - a_1)$ and $(a_3 - a_1)$ are common factors in rows 1 and 2, respectively, so

$$\begin{aligned} det \begin{bmatrix} 1 & a_1 & a_1^2 \\ 1 & a_2 & a_2^2 \\ 1 & a_3 & a_3^2 \end{bmatrix} &= (a_2 - a_1)(a_3 - a_1) det \begin{bmatrix} 1 & a_2 + a_1 \\ 1 & a_3 + a_1 \end{bmatrix} \\ &= (a_2 - a_1)(a_3 - a_1)(a_3 - a_2) \end{aligned}$$

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

$$A = \left[\begin{array}{cccc} a & 0 & 0 & 0 \\ u & b & 0 & 0 \\ v & w & c & 0 \\ x & y & z & d \end{array} \right].$$

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 or media element has been excluded from this version of the text. You can view it online here:

https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70

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 $A_1, \dots, A_k$ 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,

$$det(A^{-1}) = \frac{1}{detA}$$

Proof:

If A is invertible, then $AA^{-1} = I$; so the product theorem gives

$$1 = detI = det(AA^{-1}) = detAdetA^{-1}$$

Hence, $detA \neq 0$ and also $detA^{-1} = \frac{1}{detA}$.

Conversely, if $detA \neq 0$, 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

$$detR = detE_k \cdots detE_2 detE_1 detA$$

Since $detE \neq 0$ for all elementary matrices E, this shows $detR \neq 0$. 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 or media element has been excluded from this version of the text. You can view it online here: [*https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70*](https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70)

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70

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

$$
\begin{aligned}
det A^T = det E_1^T det E_2^T \cdots det E_k^T &= det E_1 det E_2 \cdots det E_k \\
&= det E_k \cdots det E_2 det E_1 \\
&= det A
\end{aligned}
$$

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.

$$
\begin{aligned}
det(A^3B^{-1}A^TB^2) &= det(A^3)det(B^{-1})det(A^T)det(B^2) \\
&= (detA)^3\frac{1}{detB}detA(detB)^2 \\
&= 2^3 \cdot \frac{1}{5} \cdot 2 \cdot 5^2 \\
&= 80
\end{aligned}
$$

An interactive or media element has been excluded from this version of the text. You can view it online here: *https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70*

Example 3.2.5

A square matrix is called **orthogonal** if $A^{-1} = A^T$. What are the possible values of $detA$ if A is orthogonal?

Solution:

If A is orthogonal, we have $I = AA^T$. Take determinants to obtain

$$1 = detI = det(AA^T) = detAdetA^T = (detA)^2$$

Since $detA$ is a number, this means $detA = \pm 1$.

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 $A(adjA) = (detA)I = (adjA)A$ and hence that, if $detA \neq 0$, $A^{-1} = \frac{1}{detA} adjA$. 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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70

The relationship $A(adj\,A) = (det A)I$ holds for any square matrix A.

Theorem 3.2.4 Adjugate formula

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 notan 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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70

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

$$detA = \begin{vmatrix} 2 & 1 & 3 \\ 5 & -7 & 1 \\ 3 & 0 & -6 \end{vmatrix} = \begin{vmatrix} 2 & 1 & 7 \\ 5 & -7 & 11 \\ 3 & 0 & 0 \end{vmatrix} = 3 \begin{vmatrix} 1 & 7 \\ -7 & 11 \end{vmatrix} = 180$$

Since $A^{-1} = \frac{1}{detA} adj A = \frac{1}{180}\left[c_{ij}(A)\right]^T$,
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

$$det(adjA) = (detA)^{n-1}.$$

Solution:

Write $d = detA$; we must show that $det(adjA) = d^{n-1}$. We have $A(adjA) = dI$ by Theorem 3.2.4, so taking determinants gives $ddet(adjA) = d^n$. Hence we are done if $d \neq 0$. Assume $d = 0$; we must show that $det(adjA) = 0$, that is, $adjA$ is not invertible. If $A \neq 0$, this follows from $A(adjA) = dI = 0$; if $A = 0$, it follows because then $adjA = 0$.

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 $x_1, x_2, \dots, x_n$. 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 $detA \neq 0$, we left multiply by A^{-1} to obtain the solution $\vec{x} = A^{-1}\vec{b}$. When we use the adjugate formula, this becomes

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{detA}(adjA)\vec{b}$$

$$= \frac{1}{detA}\begin{bmatrix} c_{11}(A) & c_{21}(A) & \cdots & c_{n1}(A) \\ c_{12}(A) & c_{22}(A) & \cdots & c_{n2}(A) \\ \vdots & \vdots & & \vdots \\ c_{1n}(A) & c_{2n}(A) & \cdots & c_{nn}(A) \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Hence, the variables $x_1, x_2, \ldots, x_n$ are given by

$$x_1 = \frac{1}{detA}\left[b_1c_{11}(A) + b_2c_{21}(A) + \cdots + b_nc_{n1}(A)\right]$$
$$x_2 = \frac{1}{detA}\left[b_1c_{12}(A) + b_2c_{22}(A) + \cdots + b_nc_{n2}(A)\right]$$
$$\vdots \qquad \vdots$$
$$x_n = \frac{1}{detA}\left[b_1c_{1n}(A) + b_2c_{2n}(A) + \cdots + b_nc_{nn}(A)\right]$$

Now the quantity $b_1c_{11}(A) + b_2c_{21}(A) + \cdots + b_nc_{n1}(A)$ occurring in the formula for x_1 looks like the cofactor expansion of the determinant of a matrix. The cofactors involved are $c_{11}(A), c_{21}(A), \dots, c_{n1}(A)$, 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

$$
\begin{aligned}
det A_1 &= b_1 c_{11}(A_1) + b_2 c_{21}(A_1) + \cdots + b_n c_{n1}(A_1) \\
&= b_1 c_{11}(A) + b_2 c_{21}(A) + \cdots + b_n c_{n1}(A) \\
&= (det A) x_1
\end{aligned}
$$

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

$$x_1 = \frac{det A_1}{det A}, \; x_2 = \frac{det A_2}{det A}, \; \cdots, \; x_n = \frac{det A_n}{det A}$$

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.

$$\begin{aligned} 5x_1 + x_2 - x_3 &= 4 \\ 9x_1 + x_2 - x_3 &= 1 \\ x_1 - x_2 + 5x_3 &= 2 \end{aligned}$$

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.

$$detA = det\begin{bmatrix} 5 & 1 & -1 \\ 9 & 1 & -1 \\ 1 & -1 & 5 \end{bmatrix} = -16$$

$$detA_1 = det\begin{bmatrix} 4 & 1 & -1 \\ 1 & 1 & -1 \\ 2 & -1 & 5 \end{bmatrix} = 12$$

Hence, $x_1 = \frac{detA_1}{detA} = -\frac{3}{4}$ 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 or media element has been excluded from this version of the text. You can view it online here:
https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70

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 $\frac{1}{2}$).
3. One-quarter of the juvenile females in any year survive into adulthood (the **juvenile survival rate** is $\frac{1}{4}$).

If there were 100 adult females and 40 juvenile females alive initially, compute the population of females k years later.

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

$$P^{-1}AP = D \text{ is a diagonal matrix} \tag{3.8}$$

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).

Theorem 3.3.1

If $A = PDP^{-1}$ then $A^k = PD^kP^{-1}$ for each $k = 1, 2, \dots$.

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.

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.

Example 3.3.4

Find the characteristic polynomial, eigenvalues, and basic eigenvectors for

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & -1 \\ 1 & 3 & -2 \end{bmatrix} Solution : Herethecharacteristicpolynomialisgivenby c_A(x) = det \begin{bmatrix} x-2 & 0 & 0 \\ -1 & x-2 & 1 \\ -1 & -3 & x+2 \end{bmatrix} = (x-2)(x-1)(x+1)$$

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

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

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 or media element has been excluded from this version of the text. You can view it online here: https://ecampusontario.pressbooks.pub/linearalgebrautm/?p=70

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

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

$$A = \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array}\right]$$

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.

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.

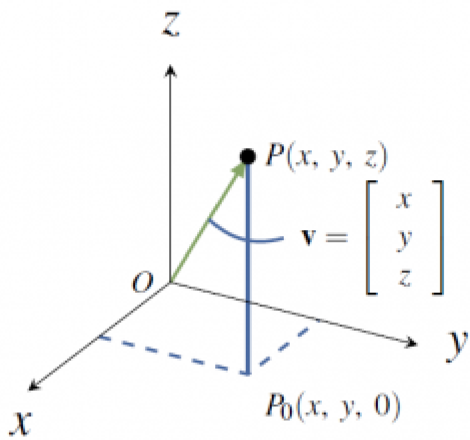


Figure 4.1.1

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 $\vec{v} = \left[\begin{array}{c} x \\ y \\ z \end{array} \right]$ 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 $\vec{0} = \left[\begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right]$.

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 $\vec{v} = \left[\begin{array}{c} x \\ y \\ z \end{array} \right]$ 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:

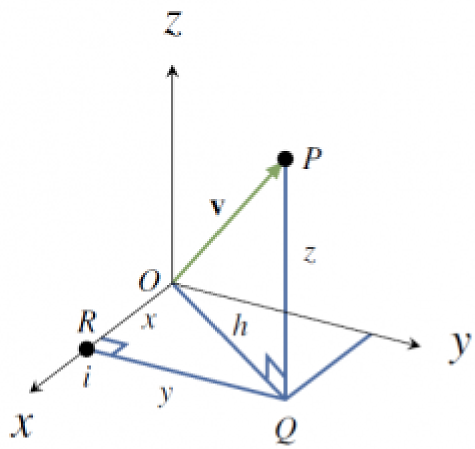


Figure 4.1.2

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 $a\vec{v} = \left[\begin{array}{ccc} ax & ay & az \end{array} \right]^T$ 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 $\vec{v} = \left[\begin{array}{r} 2 \\ -3 \\ 3 \end{array} \right]$
then $||\vec{v}|| = \sqrt{4+9+9} = \sqrt{22}$. Similarly if
$\vec{v} = \left[\begin{array}{r} 2 \\ -3 \end{array} \right]$
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.

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.

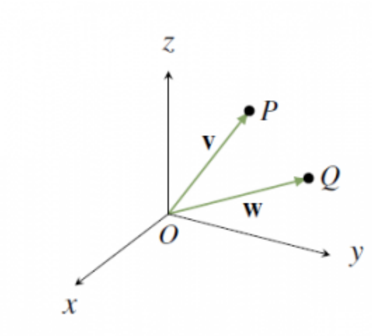


Figure 4.1.3

Hence $x = x_1$, $y = y_1$, and $z = z_1$, that is

$$\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \vec{w}.$$

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.

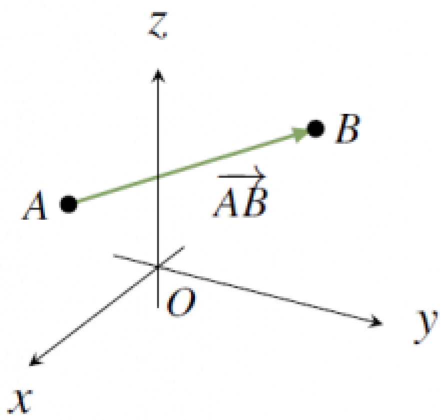


Figure 4.1.4

Hence,

Definition 4.1 Geometric vectors

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.

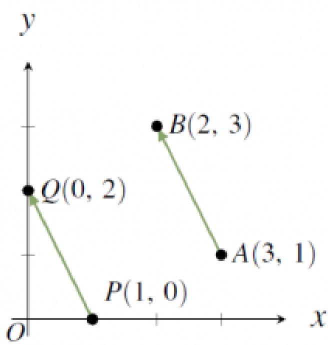


Figure 4.1.5

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

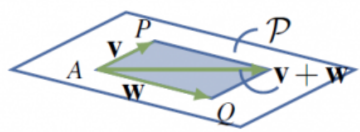


Figure 4.1.6

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 $\vec{w}$.

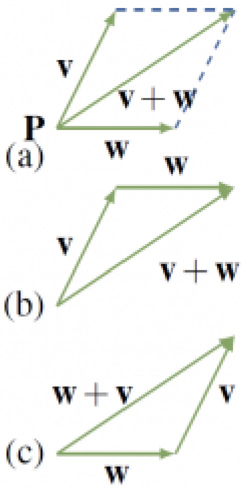


Figure 4.1.7

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.

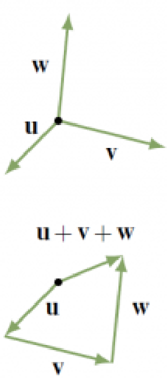


Figure 4.1.8

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 $\vec{w}$.

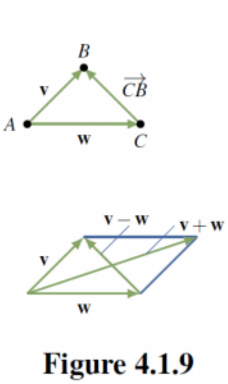


Figure 4.1.9.

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

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?

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$0" title="Rendered by QuickLaTeX.com" height="12" width="42" style="vertical-align: 0px;">; 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 $\vec{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\vec{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, 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 $a > 0$ 0" title="Rendered by QuickLaTeX.com" height="12" width="42" style="vertical-align: 0px;">. But $||a\vec{v}|| = |a|||\vec{v}|| = a||\vec{v}||$ when $a > 0$ 0" title="Rendered by QuickLaTeX.com" height="12" width="42" style="vertical-align: 0px;">, so $a\vec{v}$ is a unit vector if and only if $a = \frac{1}{||v||}$.

Definition 4.2 Parallel vectors in $\mathbb{R}^3$

Two nonzero vectors are called **parallel** if they have the same or opposite direction.

Theorem 4.1.5

Two nonzero vectors $\vec{v}$ and $\vec{w}$ are parallel if and only if one is a scalar multiple of the other.

Example 4.1.5

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.

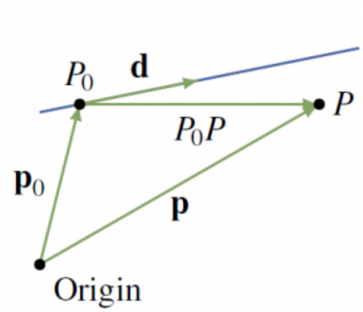


Figure 4.1.10

Then

$$\vec{p} = \vec{p_0} + \overrightarrow{P_0P}$$

Hence P lies on the line if and only if $\overrightarrow{P_0P}$ is parallel to $\vec{d}$ –that is, if and only if $\overrightarrow{P_0P} = t\vec{d}$ 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 $\vec{d} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \neq \vec{0}$ is given by

$$\begin{aligned} x &= x_0 + ta \\ y &= y_0 + tb \quad\quad t \text{ any scalar} \\ z &= z_0 + tc \end{aligned}$$

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 $z = z_0 + tc$.

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

$$\vec{d} = \vec{P_0P_1} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

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

$$\begin{aligned} x &= 2 + 2t \\ y &= -t \quad\quad\quad t \text{ a parameter} \\ z &= 1 \end{aligned}$$

Note that if P_1 is used (rather than P_0), the equations are

$$\begin{aligned} x &= 4 + 2s \\ y &= -1 - s \quad s \text{ a parameter} \\ z &= 1 \end{aligned}$$

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.

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

Solution:
Suppose $P(x, y, z)$ with vector $\vec{p}$ lies on both lines. Then

$$\begin{bmatrix} 1 - 3t \\ 2 + 5t \\ 1 + t \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -1 + s \\ 3 - 4s \\ 1 - s \end{bmatrix} \text{ for some } t \text{ and } s,$$

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

$$\begin{aligned} 1-3t &= -1+s \\ 2+5t &= 3-4s \\ 1+t &= 1-s \end{aligned}$$

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

$$\vec{p} = \left[\begin{array}{c} 1-3t \\ 2+5t \\ 1+t \end{array} \right] = \left[\begin{array}{r} -2 \\ 7 \\ 2 \end{array} \right]$$

using $t=1$. Of course, this point can also be found from

$$\vec{p} = \left[\begin{array}{c} -1+s \\ 3-4s \\ 1-s \end{array} \right] \text{ using } s=-1.$$

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.

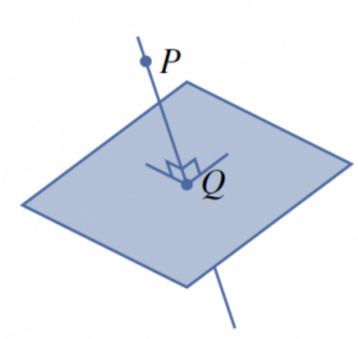


Figure 4.2.1

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
$\vec{v} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$ and
$\vec{w} = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$, their **dot product** $\vec{v} \cdot \vec{w}$ is a number defined

$$\vec{v}\vec{w} = x_1x_2 + y_1y_2 + z_1z_2 = \vec{v}^T\vec{w}$$

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:

$$\begin{aligned}
||\vec{v} - 3\vec{w}||^2 &= (\vec{v} - 3\vec{w}) \cdot (\vec{v} - 3\vec{w}) \\
&= \vec{v} \cdot (\vec{v} - 3\vec{w}) - 3\vec{w} \cdot (\vec{v} - 3\vec{w}) \\
&= \vec{v} \cdot \vec{v} - 3(\vec{v} \cdot \vec{w}) - 3(\vec{w} \cdot \vec{v}) + 9(\vec{w} \cdot \vec{w}) \\
&= ||\vec{v}||^2 - 6(\vec{v} \cdot \vec{w}) + 9||\vec{w}||^2 \\
&= 4 - 12 + 9 = 1
\end{aligned}$$

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

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

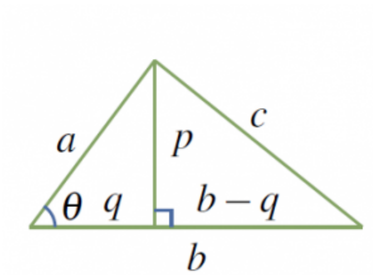


Figure 4.2.2

Proof:

We prove it when is θ acute, that is $0 \leq \theta < \frac{\pi}{2}$; 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

$$c^2 = p^2 + (b-q)^2 = a^2\sin^2\theta + (b - a\cos\theta)^2$$
$$= a^2(\sin^2\theta + \cos^2\theta) + b^2 - 2ab\cos\theta$$

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

$$0 \leq \theta \leq \pi$$

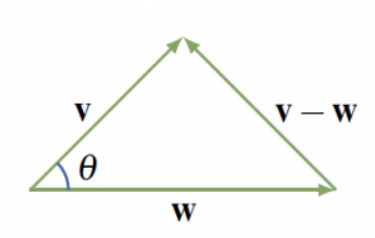
This angle θ will be called the **angle between** $\vec{v}$ and $\vec{w}$. Clearly v and w are parallel if θ is either 0 or π. Note that we do not define the angle between $\vec{v}$ and $\vec{w}$ if one of these vectors is $\vec{0}$.

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

$$\vec{v} \cdot \vec{w} = ||\vec{v}||||\vec{w}|| \cos \theta$$



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:

Figure 4.2.4

$$||\vec{v} - \vec{w}||^2 = ||\vec{v}||^2 + ||\vec{w}||^2 - 2||\vec{v}||||\vec{w}|| \cos \theta$$

On the other hand, we use Theorem 4.2.1:

$$\begin{aligned} ||\vec{v} - \vec{w}||^2 &= (\vec{v} - \vec{w}) \cdot (\vec{v} - \vec{w}) \\ &= \vec{v} \cdot \vec{v} - \vec{v} \cdot w - \vec{w} \cdot \vec{v} + \vec{w} \cdot \vec{w} \\ &= ||\vec{v}||^2 - 2(\vec{v} \cdot \vec{w}) + ||\vec{w}||^2 \end{aligned}$$

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.

Example 4.2.3

Compute the angle between

$$\vec{u} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} \text{ and}$$

$$\vec{v} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}.$$

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

$$\begin{array}{lll} \vec{v} \cdot \vec{w} > 0 & \mbox{if and only if } & \theta \mbox{ is acute } (0 \leq \theta < \frac{\pi}{2}) \\ \vec{v} \cdot \vec{w} < 0 & \mbox{if and only if } & \theta \mbox{ is obtuse } (\frac{\pi}{2} < \theta \leq 0) \\ \vec{v} \cdot \vec{w} = 0 & \mbox{if and only if } & \theta = \frac{\pi}{2} \end{array}$$

0 & \mbox{if and only if } & \theta \mbox{ is acute } (0 \leq \theta < \frac{\pi}{2}) \\ \vec{v} \cdot \vec{w} < 0 & \mbox{if and only if } & \theta \mbox{ is obtuse } (\frac{\pi}{2} < \theta \leq 0) \\ \vec{v} \cdot \vec{w} = 0 & \mbox{if and only if } & \theta = \frac{\pi}{2} \end{array} \end{equation*}" title="Rendered by QuickLaTeX.com">

In this last case, the (nonzero) vectors are perpendicular. The following terminology is used in linear algebra:

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 $\frac{\pi}{2}$.

Since $\vec{v} \cdot \vec{w} = 0$ if either $\vec{v} = \vec{0}$ or $\vec{w} = \vec{0}$, we have the following theorem:

Theorem 4.2.3

Two vectors $\vec{v}$ and $\vec{w}$ are orthogonal if and only if $\vec{v} \cdot \vec{w} = 0$.

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.

Projections

Planes

The Cross Product

4.3 More on the Cross Product

5. Vector Space [latex size ="40"]\mathbb{R}^n[/latex]

5.1 Subspaces and Spanning

5.2 Independence and Dimension

5.3 Orthogonality

5.4 Rank of a Matrix

5.5 Similarity and Diagonalization

This is where you can add appendices or other back matter.