1. SYSTEMS OF LINEAR EQUATIONS

The theory of linear equations plays an important and motivating role in the subject of this course. In fact, many problems in linear algebra are equivalent to studying a system of linear equations. Thus the techniques introduced in this chapter will be applicable to the more abstract treatment given later. On the other hand, some of the results of the abstract treatment will give us new insights into the structure of "concrete" systems of linear equations.

This chapter investigates systems of linear equations and describes in detail the Gaussian elimination algorithm which is used to find their solution. Matrices, together with certain operations on them, are also introduced here, since they are closely related to systems of linear equations and their solutions.

All our equations will involve specific numbers called *constants* or *scalars*. For simplicity, we will assume in this chapter that all our scalars are real numbers. The set of real numbers will be denoted, as usual, by $\mathbb{R}$. The solutions of our equations will also involve n-tupes

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

of real numbers called *column vectors*. The set of all such n-tuples is denoted by $\mathbb{R}^n$. We note that the results in this chapter also hold for equations over the complex numbers $\mathbb{C}$.

1.1 Linear equations and their solutions

By *linear equations* in the unknowns or variables $x_1, x_2, \ldots, x_n$ we mean an equation that can be written in the following standard form:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b, \tag{1.1}$$

where $a_1, a_2, \ldots, a_n, b$ are scalars. The scalar a_k is called the ***coefficient*** of x_k and b is called the *independent term* of the equation.

A solution of the linear equation (1.1) is a list of scalars $(s_1, s_2, \ldots, s_n)$ with the property that the following statement (obtained by substituting each s_i for x_i in the equation) is true:

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

This set of values is then said to ***satisfy*** the equation.

The set of all such solutions is called ***solution set*** or ***general solution***, or, simply, the ***solution*** of the equation.

The above notions implicitly assume that there is an ordering of the variables. If the number of variables is small, then, in order to avoid subscripts, we will usually use the variables x,y, as ordered to denote two unknowns, x,y,z, as ordered, to denote three unknowns, x,y,z,t , as ordered, to denote four unknowns, and x,y,z,s,t , as ordered, to denote five unknowns.

Example 1.1

The equations

$$4x_1 - 5x_2 + 2 = x_1 \text{ and } x_2 = 2\left(\sqrt{6} - x_1\right) + x_3$$

are both linear because they can be rearranged algebraically as in (1.1):

$$3x_1 - 5x_2 = -2 \text{ and } 2x_1 + x_2 - x_3 = 2\sqrt{6}$$

The equations

$$4x_1 - 5x_2 = x_1x_2 \text{ and } x_3 = 2\sqrt{x_1} - 6$$

are not linear because of the presence of x_1x_2 in the first equation and $\sqrt{x_1}$ in the second.

Example 1.2

The equation $x + 2y - 4z + t = 3$ is linear in the four unknowns x,y,z,t . The list of scalars $(3,2,1,0)$ is a solution of the equation since

$$3 + 2(2) - 4(1) + 0 = 3$$

is a true statement. However the list $(1,2,4,5)$ is not a solution of the equation since

$$1 + 2(2) - 4(4) + 5 = 3$$

is not a true statement.

At this point, your algebra professor expects you to have carefully and thoroughly read the introduction section of these notes.

Definition 1.1

A linear equation is said to be ***degenerate*** if it has the form

$$0x_1 + 0x_2 + \cdots + 0x_n = b,$$

that is, if every coefficient is equal to zero.

The solutions of degenerate equations are described in the following theorem.

Theorem 1.1

Consider the degenerate linear equation $0x_1 + 0x_2 + \cdots + 0x_n = b$.

1. If the constant $b \neq 0$, then the equation has no solution.
2. If the constant $b = 0$, then every list of scalars $(s_1, s_2, \ldots, s_n)$ is a solution.

Proof. We begin by proving the first statement. Let $(s_1, s_2, \ldots, s_n)$ be any list of scalars.

Suppose that $b \neq 0$. Substituting $x_i = s_i$ in the equation we obtain:

$$0s_1 + 0s_2 + \cdots + 0s_n = b \text{ or } 0 + 0 + \cdots + 0 = b \text{ or } 0 = b.$$

This is not a true statement since $b \neq 0$. Hence no list of scalars is a solution.

Now we prove the second statement. Suppose $b = 0$. Substituting $(s_1, s_2, \ldots, s_n)$ in the equation we obtain:

$$0s_1 + 0s_2 + \cdots + 0s_n = 0 \text{ or } 0 + 0 + \cdots + 0 = 0 \text{ or } 0 = 0$$

which is a true statement. Thus every list of scalars is a solution, as claimed. ■

1.2 Systems of linear equations and elementary operations

We now consider a system of m linear equations, say, $L_1, L_2, \ldots, L_m$, in the n unknowns $x_1, x_2, \ldots, x_n$ which can be put in the *standard form*:

$$\begin{aligned} L_1 &: a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ L_2 &: a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ L_m &: a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m, \end{aligned} \tag{1.2}$$

where the a_{ij}, b_i are constants.

A *solution* (or *particular solution*) of the above system is a list of scalars $(s_1, s_2, \ldots, s_n)$ which is simultaneously a solution of all the equations of the system. The set of all such solutions is called the *solution set* or the *general solution* of the system.

Example 1.3

Consider the system of equations:

$$\begin{aligned} L_1 &: x + 2y - 4z + t = 3, \\ L_2 &: 2x + y + 5z - 4t = 13. \end{aligned}$$

We can verify that $(3,2,1,0)$ is a solution by substituting $x = 3, y = 2, z = 1$, and $t = 0$ in the equations of the system. Similarly, we can verify that $(3,2,1,-3)$ is not a solution.

Systems of linear equations in the same unknowns are said to be *equivalent* if the systems have the same solution set. One way of producing a system which is equivalent to a given system, with linear equations $L_1, L_2, \ldots, L_m$, is by applying a sequence of the following three basic operations called elementary operations:

Definition 1.2: Elementary operations

(E1) Interchange the i th equation and the j th equation: $L_i \leftrightarrow L_j$.

(E2) Multiply the i th equation by a non zero scalar $k : kL_i \to L_i, k \neq 0$.

(E3) Replace the i th equation by itself plus k times the j th equation: $\left(kL_j + L_i\right) \to L_i$.

Remark 1.1

It is important that you understand that if one system of equations is obtained from another by applying elementary operations, then the two systems will be equivalent. This means that they will have the same solution set. The reason for this is that each elementary operation is *reversible*. The basic application of this fact is that we can *replace one system with an equivalent system that is easier to solve*.

Roughly speaking, the simplest equivalent version of a system is obtained using the elementary operations as follows. Use the x_1 term in the first equation of a system to eliminate the x_1 terms in the other equations. Then use the x_2 term in the second equation to eliminate the x_2 terms in the other equations, and so on, until you finally obtain the least amount of variables in common among the equations. The following example illustrates this strategy. Make sure to understand that we replace x_1, x_2, x_3 with x, y, z.

Example 1.4

The solution of the system

$$\begin{aligned} L_1 &: x + 2y - 4z = -4 \\ L_2 &: 5x + 11y - 21y = -22 \\ L_3 &: 3x - 2y + 3z = 11 \end{aligned}$$

is obtained as follows:

Use x in the first equation to eliminate it from the other equations. First we replace L_2 with $-5L_1 + L_2$:

$$\begin{array}{lrcrr} L_1: & x + 2y - 4z = -4 & \text{Operations:} & L_1: & x + 2y - 4z = -4 \\ L_2: & 5x + 11y - 21z = -22 & \left(-5L_1 + L_2\right) \to L_2 & \text{new } L_2: & y - z = -2 \\ L_3: & 3x - 2y + 3z = 11 & & L_3: & (3x - 2y + 3z = 11 \end{array}$$

Next we keep the new L_2 and replace L_3 with $-3L_1 + L_3$:

$$\begin{array}{lrcrr} L_1: & x + 2y - 4z = -4 & \text{Operations:} & L_1: & x + 2y - 4z = -4 \\ L_2: & y - z = -2 & \left(-3L_1 + L_3\right) \to L_3 & L_2: & y - z = -2 \\ L_3: & 3x - 2y + 3z = 11 & & \text{new } L_3: & -8y + 15z = 23 \end{array}$$

Use y in the second equation to eliminate it from the other equations. We replace L_1 with $-2L_2 + L_1$ and replace L_3 with $8L_2 + L_3$. We show both replacements at once:

$$\begin{array}{lrcrr} L_1: & x + 2y - 4z = -4 & \text{Operations:} & \text{new } L_1: & x - 2z = 0 \\ L_2: & y - z = -2 & \left(-2L_2 + L_1\right) \to L_1 & L_2: & y - z = -2 \\ L_3: & -8y + 15z = 23 & \left(8L_2 + L_3\right) \to L_3 & \text{new } L_3: & 7z = 7 \end{array}$$

At this point, it is evident that $z = 1$ and that you can use this to solve for the other two variables. Nevertheless, our philosophy of using elementary operations to reach a system of equations with the least amount of variables in common among them will be the keystone to developing all the theory in the rest of the course. So, even if it seems somewhat unnecessary right row, we will continue to solve the system with elementary operations for illustrative purposes.

To simplify calculations, we can replace L_3 with $\frac{1}{7}L_3$:

$$\begin{array}{lrcrr} L_1: & x - 2z = 0 & \text{Operations:} & L_1: & x - 2z = 0 \\ L_2: & y - z = -2 & \frac{1}{7}L_3 \to L_3 & L_2: & y - z = -2 \\ L_3: & 7z = 7 & & \text{new } L_3: & z = 1 \end{array}$$

Use z in the third equation and use it eliminate from the other equations. We replace L_1 with $2L_3 + L_1$ and replace L_2 with $L_3 + L_2$:

$$\begin{array}{lrcr}
L_1: & x - 2z = 0 & \text{Operations:} & \text{new } L_1: & x = 2 \\
L_2: & y - z = -2 & (2L_3 + L_1) \to L_1 & \text{new } L_2: & y = -1 \\
L_3: & z = 1 & (L_3 + L_2) \to L_2 & L_3: & z = 1
\end{array}$$

We have clearly arrived at the simplest equivalent system of equations. Notice that the three equations have no variables in common. This is the ideal scenario since we can readily deduce that the solution of the system is $x = 2, y = -1, z = 1$.

Suppose that you start with the system of equation (1.2) and use elementary operations to obtain a simpler equivalent version such that the equations have the least amount of variables in common. Then the first variable with non zero coefficient in each equation is called ***leading variable*** and the variables that are not leading variables are called ***free variables***.

Recall from the beginning of this chapter that we have implicitly assumed that there is an ordering of the variables. Moreover, there may be more than one simpler system equivalent to (1.2). For this reason, we impose some additional requirements for writing a simpler equivalent version of (1.2). First, we will require that the equations should be rearranged so that the leading variables are order not only horizontally from left to right but also vertically from top to bottom. This can be accomplished by interchanging the necessary equations (recall that interchanging equations is an elementary operation). Second, we will require that the coefficient of each leading variable is 1*. This is possible by rescaling the necessary equations. These two additional conditions ensure that the resulting simpler equivalent system is ***the simplest*** one, independently of the sequence of elementary operations that you use to arrive at it.

This notion of simplest equivalent system will be, in our opinion, the most useful tool at your disposal to develop and understand the material in this course. We will illustrate how to obtain such simplest equivalent system in the following examples.

* This format of writing the simplest version is called the *row reduced echelon form*, but we are reserving this name for matrices.

Example 1.5

Consider the following system of equations:

$$\begin{aligned} L_1 &: \quad y-4z=8, \\ L_2 &: \quad 2x-3y+2z=1 \\ L_3 &: \quad 5x-8y+7z=1 \end{aligned}$$

1. Find the simplest system of equations that is equivalent to the given one.
2. Indicate the leading variables and the free variables.
3. Find the solution of the given system of equation.

1. First, you need to remember that the variables need to be in the correct order horizontally and vertically, and that the coefficients of the leading variables should be equal to 1. This means that the variable x should appear as the leading variable of the first equation. This is not the case for the given system. One way to fix this is to interchange L_1 with either L_2 or L_3. Either option is good.

$$\begin{array}{llcll} L_1: & y-4z=8 & \text{Operations:} & \text{new } L_1: & 2x-3y+2z=1 \\ L_2: & 2x-3y+2z=1 & L_1 \leftrightarrow L_2 & \text{new } L_2: & y-4z=8 \\ L_3: & 5x-8y+7z=1 & & L_3: & 5x-8y+7z=1 \end{array}$$

The coefficient of x in L_1 should be 1 , so we rescale as follows:

$$\begin{array}{llcll} L_1: & 2x-3y+2z=1 & \text{Operations:} & \text{new } L_1: & x-\frac{3}{2}y+z=\frac{1}{2} \\ L_2: & y-4z=8 & \frac{1}{2}L_1 \leftrightarrow L_1 & L_2: & y-4z=8 \\ L_3: & 5x-8y+7z=1 & & L_3: & 5x-8y+7z=1 \end{array}$$

Use x in the first equation to eliminate it from the other equations.

$$\begin{array}{llcll} L_1: & x-\frac{3}{2}y+z=\frac{1}{2} & \text{Operations:} & L_1: & x-\frac{3}{2}y+z=\frac{1}{2} \\ L_2: & y-4z=8 & \left(-5L_1+L_3\right) \to L_3 & L_2: & y-4z=8 \\ L_3: & 5x-8y+7z=1 & & \text{new } L_3: & -\frac{1}{2}y+2z=-\frac{3}{2} \end{array}$$

Use y in the second equation and use it eliminate from the other equations.

$$\begin{array}{llcll} L_1: & x-\frac{3}{2}y+z=\frac{1}{2} & \text{Operations:} & \text{new } L_1: & x-5z=\frac{25}{2} \\ L_2: & y-4z=8 & \left(\frac{3}{2}L_2+L_1\right) \to L_1 & L_2: & y-4z=8 \\ L_3: & -\frac{1}{2}y+2z=-\frac{3}{2} & & \text{new } L_3: & 0=-\frac{5}{2} \end{array}$$

We have ran out of non-degenerate equations, there is the least amount of variables in common among the equation, the variables are in the correct order, and the coefficients of the leading variables are 1. Therefore we have found the simplest equivalent system.

2. The leading variables are x and y because they are the first variables in each nondegenerate equation with non zero coefficient. The remainder variable, z, is a free variable.

3. It turns out that $L_3 : 0 = -\frac{5}{2}$ is a degenerate equation with no solution which means that there is no values for x, y, z that will *simultaneously* satisfy all equations. Therefore, the given system of equations has no solution.

Before continuing on to the next example, we need to discuss free variables further. Suppose that after reducing (1.2) to its simplest equivalent version, you find that all of its degenerate equations (if any) are of the form $0x_1 + 0x_2 + \cdots + 0x_n = 0$, and that at least one of its variables is a free variable. Then the system has infinite number of solutions since each of the free variables may be assigned any real number. The general solution of the system is obtained as follows. Once the system has been reduced to its simplest form, arbitrary values, called *parameters*, are assigned to the free variables, and the non free variables can be solved in terms of these parameters.

Example 1.6

Find the general solution of the system

$$\begin{aligned} L_1 &: \quad x + 4y - 3z + 2t = 5 \\ L_2 &: \quad -2x - 8y + 7z - 8t = -8 \end{aligned}$$

If the solution of the system is not obvious, then our best option to find it is to reduce the system to its simplest form. This is done as follows: First, we eliminate x from the second equation.

$$\begin{array}{lr|c|lr} L_1 : & x + 4y - 3z + 2t = 5 & \text{Operations:} & L_1 : & x + 4y - 3z + 2t = 5 \\ L_2 : & -2x - 8y + 7z - 8t = -8 & (2L_1 + L_2) \to L_2 & L_2 : & z - 4t = 2 \end{array}$$

It is not possible to eliminate y from L_1 since the leading variable of L_2 is z, which means that our next task is to eliminate z from L_1.

$$\begin{array}{lr|c|lr} L_1 : & x + 4y - 3z + 2t = 5 & \text{Operations:} & L_1 : & x + 4y - 10t = 11 \\ L_2 : & z - 4t = 2 & (3L_2 + L_1) \to L_1 & L_2 : & z - 4t = 2 \end{array}$$

The last system is the simplest equivalent version of the original one. Make sure to understand why this is so. The first variables with nonzero coefficients in each equation are x and z which means that these are the leading variables and that y and t are free variables. From the discussion above, we deduce that the system has infinite number solutions and the general solution is obtained as follows. We assign arbitrary values to the free variables y and t, say,

$$y = \lambda \text{ and } t = \mu, \text{ where } \lambda, \mu \in \mathbb{R}$$

Then we solve for x and z in terms of λ and μ. We write the general solution:

$$\begin{aligned} x &= 11 - 4\lambda + 10\mu, \\ y &= \lambda, \\ z &= 2 + 4\mu, \\ t &= \mu. \end{aligned}$$

It is always a good idea to check that your final result is indeed the correct answer. In this case, you should verify that these expressions for x, y, z, t satisfy **both** equations of the original system. If you obtain two true statements then your answer is correct.

We finish this section by answering two fundamental questions about systems of linear equations:

1. Does at least one solution exist?

2. If a solution exists, is it the only one; that is, is the solution unique?

The answer to these questions are stated as a theorem.

Theorem 1.2

Any system of linear equations has either: (i) a unique solution, (ii) no solution, or (iii) an infinite number of solutions.

Proof. Given any system of linear equations, we apply elementary operations to reduce it to its simplest equivalent form. After this is done, we may get one of the following cases:

I. There is at least one degenerate equation of the form $0x_1 + 0x_2 + \cdots + 0x_n = b$ with $b \neq 0$. This degenerate equation has no solution and, consequently, the system has no solution.

II. There is no degenerate equation of the form $0x_1 + 0x_2 + \cdots + 0x_n = b$ with $b \neq 0$. We have two subcases.

   a. There is at least one free variable. This means that the system has an infinite number of solutions.

b. There are no free variables. This means that all the variables are leading variables. This is only possible if the simplest equivalent version is of the form

$$\begin{aligned} x_1 &= s_1 \\ x_2 &= s_2 \\ &\vdots \\ x_n &= s_n, \end{aligned}$$

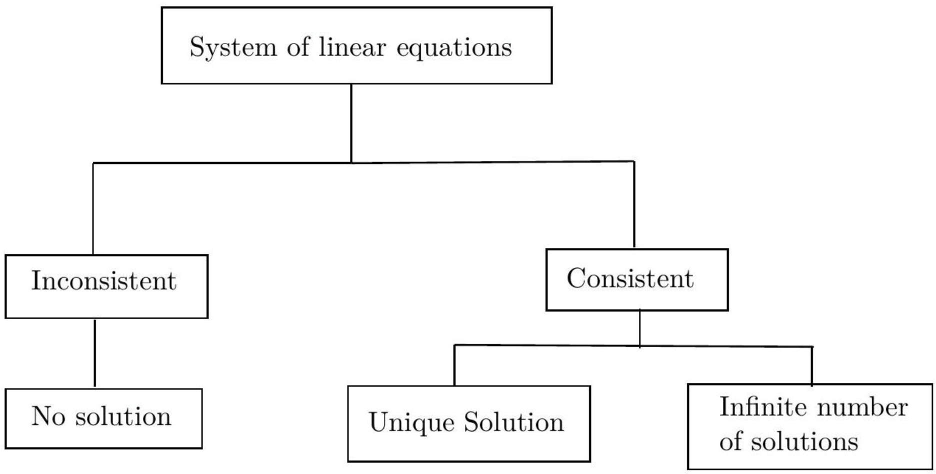
where $s_1, s_2, \ldots, s_n$ are real numbers. Then each variable has a unique value which means that the solution of the system is unique.

No other cases are possible, which proves the theorem. ■

Remark 1.2

Case II in the previous proof does not say that all the equations are non-degenerate. It only says that there are no degenerate equations that have no solution. In fact, this case may include degenerate equations of the form $0x_1 + 0x_2 + \cdots + 0x_n = 0$. In the future, make sure you understand what a definition, theorem or proof says and *does not* say.

In light of Theorem 1.2, systems of linear equations can be classified according to the existence of solutions. A system is said to be *consistent* if it has at least one solution, and is said to be *inconsistent* if it has no solution.



1.3 Matrices and matrix operations

A *matrix* is a rectangular array of numbers:

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

Such matrix may be denoted by $A = (a_{ij})$. Note that the element a_{ij}, called the (i,j)-entry, appears in the i-th row and the j-th column. A matrix with m rows and n columns is called and $m \times n$ matrix.

The first non zero entry of a row is called its *leading entry*. A row whose all of its entries consists of zeros has no leading entries.

Definition 1.3

A matrix is in *echelon form* (or *row echelon form*) if it has the following properties:

1. All zero rows, if any, are at the bottom of the matrix.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zeros.

If a matrix in echelon form satisfies the following additional conditions, then it is in *reduced echelon form* (or *row reduced echelon form*):

4. The leading entry in each non zero row is 1 .
5. Each leading 1 is the only non zero entry in its column.

Example 1.7

The following matrices are in echelon form. The leading entries (represented with ■) may have any non zero value; the non leading entries (represented with *) may have any values (including zero).

$$\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & \blacksquare & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * \end{bmatrix}$$

The following matrices are in reduced echelon form because the leading entries are 1's, and there are 0 's below and above each leading 1, that is, each leading 1 is the only non zero entry in its column.

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

Just like with systems of equations, we can operate with the rows of a matrix by applying the following three basic operations.

Definition 1.4: Elementary row operations

(E1) Interchange the i th row and the j th row: $R_i \leftrightarrow R_j$.

(E2) Multiply the i th row by a non zero scalar $k : kR_i \to R_i, k \neq 0$.

(E3) Replace the i th row by itself plus k times the j th row: $\left(kR_j + R_i\right) \to R_i$.

A matrix A is said to be *row equivalent* to a matrix B, written $A \sim B$, if B can be obtained from A by applying a finite sequence of elementary row operations.

Any non zero matrix is row equivalent to more than one matrix in echelon form that can be obtained by using different sequences of row operations. However, the reduced echelon form one obtains from a matrix is unique, independently of the row operations used to obtain it. We state this as a theorem without proof.

Theorem 1.3

Each matrix is row equivalent to one and only one reduced echelon matrix.

When row operations on a matrix produce an echelon form, further row operations to obtain the reduced echelon form do not change the positions of the leading entries. Since the reduced echelon form is unique, *the leading entries are always in the same positions in any echelon form obtained from a given matrix*. These leading entries correspond to leading 1's in the reduced echelon form.

Definition 1.5

A *pivot position* in a matrix A is a location in A that corresponds to the leading 1 in the row reduced echelon form of A. A *pivot column* is a column of A that contains a pivot position.

Heuristically, pivot positions are non zero entries of a matrix that are used to create zeros as needed via elementary row operations to obtain a row reduced echelon form.

Example 1.8

Consider the matrix

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

1. Find *an* echelon form of A .
2. Find *the* row reduced echelon form of A .
3. Indicate the pivot positions of A .

1. There are more than one possible echelon form; each one is obtained depending on the sequence of elementary row operations used. We compute one of them. First, we start by using $a_{11} = 1$ as pivot to create 0 's below a_{11}, that is, apply $-2R_1 + R_2 \to R_2$ and $-3R_1 + R_3 \to R_3$ to obtain

$$\begin{bmatrix} 1 & 2 & -3 & 0 \\ 0 & 0 & 4 & 2 \\ 0 & 0 & 5 & 3 \end{bmatrix}.$$

Now use the new (2,3)-entry as a pivot to create 0 's below it, that is, apply $-\frac{5}{4}R_2 + R_3 \to R_3$ to obtain

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

The matrix is now in echelon form.

2. We apply further elementary row operations to the echelon form above to obtain the unique row reduced echelon form. Each pivot position should be 1,

so we apply $\frac{1}{4}R_2 \to R_2$ and $2R_3 \to R_3$ to obtain

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

At this point, the fastest way to obtain the row reduced echelon form is to apply $\frac{1}{2}R_3 + \ R_2 \to R_2$ to obtain

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

followed by applying $3R_2 + R_1 \to R_1$ to obtain

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

The matrix is now in row reduced echelon form.

3. The locations in A that correspond to the leading 1's in the row reduced echelon form are $a_{11} = 1, a_{23} = -2$, and $a_{34} = 3$. These are the pivot positions of A.

Matrices will usually be denoted by capital letters $A, B, \ldots$ Two matrices A and B are *equal*, written $A = B$, if they have the same size, that is, the same number of rows and columns, and if the corresponding elements are equal. Thus the equality of two $m \times n$ matrices is equivalent to a system of mn equalities, one for each pair of elements. For instance, the statement

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

is equivalent to the following system of equations:

$$\begin{aligned} x + y &= 3 \\ x - y &= 1 \\ 2z + w &= 5 \\ z - 2 &= 4. \end{aligned}$$

The solution of the system is $x = 2, y = 1, z = 3, w = -1$. (Do not just accept that these are the answers. Try to solve the system by yourself.)

A matrix with one row is also referred to as a *row matrix*, and with one column as a *column vector*. In particular, a scalar can be viewed as a 1×1 matrix.

Let $A = (a_{ij})$ and $B = (b_{ij})$ be two matrices with the same size, say, $m \times n$ matrices. The *sum* of A and B, written $A + B$, is obtained by adding the corresponding entries:

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} + b_{m1} & a_{2m} + b_{2m} & \cdots & a_{mn} + b_{mn} \end{bmatrix},$$

or, more compactly, $A + B = (a_{ij} + b_{ij})$.

The *product* of a matrix $A = (a_{ij})$ by a scalar k, written kA, is the matrix obtained by multiplying each entry of A by k : $kA = (ka_{ij})$. We also define $-A = (-1)A$ and $A - B = A + (-B)$.

The $m \times n$ matrix whose entries are all zeros is called the zero matrix. We usually write the zero matrix as 0 and its size is usually understood from context. The zero matrix is similar to the scalar 0. For any matrix $A, A + 0 = 0 + A = A$.

Basic properties of matrices under the operations of matrix addition and scalar multiplication follow.

Theorem 1.4

For any $m \times n$ matrices A, B, C and scalars, α, β

a) $(A + B) + C = A + (B + C)$ b) $A + 0 = A$

c) $A + (-A) = 0$ d) $A + B = B + A$

e) $\alpha(A + B) = \alpha A + \alpha B$ f) $(\alpha + \beta)A = \alpha A + \beta A$

g) $(\alpha\beta)A = \alpha(\beta A)$ h) $1A = A$

Proof. We only prove the first property here. The rest are left as an exercise for the reader. Let $A = (a_{ij}), B = (b_{ij})$, and $C = (c_{ij})$ be matrices of the same size. Then

$$(A + B) + C = ((a_{ij} + b_{ij}) + c_{ij}) = (a_{ij} + (b_{ij} + c_{ij})) = A + (B + C).$$

■

There are several ways to define matrix-vector multiplication. You have probably seen one way in high school. Nevertheless, we choose the following definition because it is the most convenient for developing the theoretical ideas throughout the course. It should not be difficult for you verify that the definition you learned in high school is equivalent to the one we present here.

Definition 1.6

If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and if $\mathbf{x}$ is a column vector in $\mathbb{R}^n$, then the *product of A and $\mathbf{x}$*, denoted by $A\mathbf{x}$, is defined as follows,

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

In words, the product of a matrix A and a column vector $\mathbf{x}$ *is the column vector that is equal to the weighted sum of the columns of A using the entries $\mathbf{x}$ as weights*. Note that $A\mathbf{x}$ is defined only if the number of columns of A equals the number of entries in $\mathbf{x}$.

We take this opportunity to give this type of weighted sum a special name. Given vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ in $\mathbb{R}^n$ and given scalars $c_1, c_2, \ldots, c_p$, the vector $y \in \mathbb{R}^n$ defined by

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$$

is called a *linear combination* of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ with *weights* $c_1, c_2, \ldots, c_p$.

Example 1.9

We compute

$$\begin{bmatrix} 3 & 1 & 0 \\ 2 & 2 & 5 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}.$$

By the definition of matrix-vector multiplication, we must compute the linear combination of the columns of the matrix using the entries of the vector as weights. Therefore,

$$\begin{bmatrix} 3 & 1 & 0 \\ 2 & 2 & 5 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = 2\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 5 \\ 4 \end{bmatrix} = \begin{bmatrix} 7 \\ 21 \\ 14 \end{bmatrix}.$$

Remark 1.3

A common error that we see repeatedly is students treating vectors and matrices like fractions. For instance, some students will notice that the entries of the vector in the final result of the previous example share a common factor of 7 ; then they will proceed to “cancel” this common factor out like they would do with the numerator and denominator of a fraction and write the following as a final answer:

$$\begin{bmatrix} 3 & 1 & 0 \\ 2 & 2 & 5 \\ 1 & 0 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}.$$

This is, of course, absolutely wrong. If you write this, your algebra professor will realize that life is too short, immediately resign from teaching, and join a traveling circus. But not before changing your grade in the course to a zero.

Example 1.10

The product of a row matrix and a column vector is

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

Make sure you understand how the definition was used to find this product.

The properties of the matrix-vector multiplication in the next theorem are important and will be used throughout the course.

Theorem 1.5

For an $m \times n$ matrix A, vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, and scalar c,

a. $A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v}$

b. $A(c\mathbf{u}) = c(\mathbf{Au})$

Proof. Let u_i and v_i be the i-th entries of **u** and **v**, respectively. To prove the first statement, use the definition of matrix-vector multiplication to compute $A(\mathbf{u}+\mathbf{v})$:

$$A(\mathbf{u}+\mathbf{v}) = [\mathbf{a}_1\mathbf{a}_2 \dots \mathbf{a}_n]\begin{bmatrix} u_1+v_1 \\ u_2+v_2 \\ \vdots \\ u_n+v_n \end{bmatrix}$$
$$= (u_1+v_1)\mathbf{a}_1 + (u_2+v_2)\mathbf{a}_2 + \cdots + (u_n+v_n)\mathbf{a}_n$$
$$= (u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + \cdots + u_n\mathbf{a}_n) + (v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + \cdots + v_n\mathbf{a}_n)$$
$$= A\mathbf{u} + A\mathbf{v}.$$

The proof of the second statement is left as an exercise for the reader. ■

You have certainly seen how to multiply matrices in high school. Nevertheless, such a definition is not very useful for theoretical purposes. We now give an equivalent abstract definition of matrix multiplication.

Definition 1.7

Let A be an $m \times p$ matrix, and let B be a $p \times n$ matrix. The product of A and B, denoted by AB, is the unique $m \times n$ matrix such that, for every vector $\mathbf{x} \in \mathbb{R}^n, (AB)\mathbf{x} = A(B\mathbf{x})$.

It is possible to find a (less abstract) representation of AB in terms of matrix-vector multiplication. We provide this representation in the following theorem.

Theorem 1.6

Let A be an $m \times p$ matrix, and let B be a $p \times n$ matrix with columns $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$, then the product AB is the unique $m \times n$ matrix whose columns are $A\mathbf{b}_1, A\mathbf{b}_2, \dots, A\mathbf{b}_n$. That is,

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \dots & A\mathbf{b}_n \end{bmatrix}.$$

Proof. Following the definition of AB, let $\mathbf{x}$ be an arbitrary vector in $\mathbb{R}^n$. Let

$$\mathbf{y} = B\mathbf{x} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_n \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n.$$

Then,

$$AB\mathbf{x} = A(B\mathbf{x}) = A\mathbf{y}$$
$$= A(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n)$$
$$= x_1A\mathbf{b}_1 + x_2A\mathbf{b}_2 + \cdots + x_nA\mathbf{b}_n$$
$$= \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \dots & A\mathbf{b}_n \end{bmatrix}\mathbf{x}.$$

This is true for every choice of $\mathbf{x}$ and each product $A\mathbf{b}_i$ is uniquely determined. This proves the theorem. ■

Example 1.11

Compute AB, where

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

Write $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{bmatrix}$, and compute

$$A\mathbf{b}_1 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 11 \\ -1 \end{bmatrix}, \quad A\mathbf{b}_2 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 13 \end{bmatrix}, \quad A\mathbf{b}_3 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 21 \\ -9 \end{bmatrix}.$$

Then,

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix}.$$

The following theorem lists the standard properties of matrix multiplication. We note that I_m denotes the unique $m \times m$, called the *identity matrix*, such that $I_m\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$. The identity matrix has the following representation:

$$I_m = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Theorem 1.7

Let A be an $m \times n$ matrix, and let B and C have sizes for which the indicated sums and products are defined.

a. $A(BC) = (AB)C$ (associative law of multiplication)

b. $A(B+C) = AB + AC$ (left distributive law)

c. $(B+C)A = BA + CA$ (right distributive law)

d. $r(AB) = (rA)B = A(rB)$ for any scalar r

e. $I_mA = A = AI_m$ (identity for matrix multiplication)

Proof. Let $C = \begin{bmatrix} \mathbf{c}_1 & \dots & \mathbf{c}_p \end{bmatrix}$. By Theorem 1.6,

$$BC = \begin{bmatrix} B\mathbf{c}_1 & \dots & B\mathbf{c}_p \end{bmatrix}$$
$$A(BC) = \begin{bmatrix} A(B\mathbf{c}_1) & \dots & A(B\mathbf{c}_p) \end{bmatrix},$$

By the definition of matrix multiplication, $A(B\mathbf{x}) = (AB)\mathbf{x}$ for all $\mathbf{x}$, so

$$A(BC) = \begin{bmatrix} (AB)\mathbf{c}_1 & \dots & (AB)\mathbf{c}_p \end{bmatrix} = (AB)C.$$

The rest of the properties are left for the reader to prove. ■

The left-to-right order in matrix products is critical because AB and BA are usually not the same. On one hand, the sizes of A and B are such that AB is well defined but BA may not be even possible to compute. For instance, if A is a 3×5 matrix and B is a 5×1 matrix, then we can compute AB because the number of columns of A is equal to the number of rows of B, but it is not possible to compute BA because the number of columns of B is not equal to the number of rows of A. on the other hand, the columns of AB are linear combinations of the columns of A, whereas the columns of BA are constructed from the columns of B. If $AB = BA$, we say that A and B *commute* with one another.

Remark 1.4

We have already discussed that, in general, matrix products do not commute, that is, $AB \neq BA$. There two more important differences between matrix algebra and ordinary algebra that you need to keep in mind.

1. The cancellation laws *do not* hold for matrix multiplication. That is, if $AB = AC$, then it is not true, in general, that $B = C$. For instance, let

$$A = \begin{bmatrix} 2 & -3 \\ -4 & 6 \end{bmatrix}, \; B = \begin{bmatrix} 8 & 4 \\ 5 & 5 \end{bmatrix}, \; C = \begin{bmatrix} 5 & -2 \\ 3 & 1 \end{bmatrix}.$$

   Verify that $AB = AC$ and yet $B \neq C$.

2. If a product AB is the zero matrix, you *cannot* conclude, in general, that either $A = 0$ or $B = 0$. For instance, let

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

   Verify that $AB = 0$ and yet $A \neq 0$ and $B \neq 0$.

Every now and then we need to compute a particular entry of the product AB without computing the whole matrix product. In this case, we do so by recalling the matrix multiplication rule you learned in high school (which can be deduced from Theorem 1.6).

Theorem 1.8

Let $A=\left(a_{ij}\right)$ be an $m\times p$ matrix and let $B=\left(b_{ij}\right)$ be a $p\times n$ matrix. If $(AB)_{ij}$ denotes the $\left(i,j\right)$-entry in AB, then

$$(AB)_{ij}=\begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{ip}\end{bmatrix}\begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{pj}\end{bmatrix}=a_{i1}b_{1j}+a_{i2}b_{2j}+\dots+a_{ip}b_{pj}.$$

Given an $m\times n$ matrix $A=\left(a_{ij}\right)$, the transpose of A, denoted by $A^\top$, is the $n\times m$ matrix $A^\top=\left(a_{ij}^\top\right)$, where $a_{ij}^\top=a_{ji}$. That is, $A^\top$ is obtained by changing the columns of A into columns, or, equivalently, by changing the rows of A into columns. So, if we write $A=\left[\mathbf{a}_1\dots\mathbf{a}_n\right]$, then

$$A^\top=\begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_n^\top\end{bmatrix}.$$

Theorem 1.9

Let A and B denote matrices whose sizes are appropriate for the following sums and products.

a. $\left(A^\top\right)^\top=A$

b. $(A+B)^\top=A^\top+B^\top$

c. For any scalar $r,(rA)^\top=rA^\top$

d. $(AB)^\top=B^\top A^\top$

Proof. We only prove the last statement. The rest are left for the reader as exercises. First, observe that

$$\left[(AB)^\top\right]_{ij}=(AB)_{ji}=a_{j1}b_{1i}+a_{j2}b_{2i}+\dots+a_{jp}b_{pi}=\begin{bmatrix} b_{1i} & b_{2i} & \dots & b_{pi}\end{bmatrix}\begin{bmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jp}\end{bmatrix}.$$

On the other hand,

$$\left[B^\top A^\top\right]_{ij} = \begin{bmatrix} b_{i1}^\top & b_{i2}^\top & \cdots & b_{ip}^\top \end{bmatrix} \begin{bmatrix} a_{1j}^\top \\ a_{2j}^\top \\ \vdots \\ a_{pj}^\top \end{bmatrix} = \begin{bmatrix} b_{1i} & b_{2i} & \cdots & b_{pi} \end{bmatrix} \begin{bmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jp} \end{bmatrix},$$

which shows that the (i,j)-entry of $(AB)^\top$ is equal to the corresponding entry of $B^\top A^\top$. Therefore, $(AB)^\top = B^\top A^\top$. ■

1.4 Systems of linear equations and matrices

Consider again a system of m linear equations and n unknowns:

$$\begin{aligned} L_1 \quad &: a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ L_2 \quad &: a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ L_m \quad &: a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m. \end{aligned} \tag{1.3}$$

We can interpret (1.3) the equations resulting from considering two equal column vectors:

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Then, L_i is obtained by setting the i-th entry of the vector on the left side equal to the i-th entry of the vector on the right side. We can also operate on the left-hand side vector as follows:

$$\begin{aligned} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} &= \begin{bmatrix} a_{11}x_1 \\ a_{21}x_1 \\ \vdots \\ a_{m1}x_1 \end{bmatrix} + \begin{bmatrix} a_{12}x_2 \\ a_{22}x_2 \\ \vdots \\ a_{m2}x_2 \end{bmatrix} + \cdots + \begin{bmatrix} a_{1n}x_n \\ a_{2n}x_n \\ \vdots \\ a_{mn}x_n \end{bmatrix} \\ &= x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

So, if we let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

we can write (1.3) more compactly as the vector equation

$$A\mathbf{x} = \mathbf{b}. \tag{1.4}$$

The matrix A is called the *matrix coefficient*, $\mathbf{x}$ is the *unknown vector*, and $\mathbf{b}$ is the *independent term*. We make the following key observation:

Observation 1.1

Every solution of the system (1.3) is solution of the vector equation (1.4), and vice versa.

The *augmented matrix* of (1.3) is the following matrix:

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_n \end{bmatrix}.$$

That is, the augmented matrix of the system (1.3) is the matrix which consists of the matrix A concatenated by the vector $\mathbf{b}$. Observe that the system (1.3) is completely determined by the vector equation (1.4) which, in turn, is completely determined by its augmented matrix. In light of this, we have the following theorem.

Theorem 1.10

If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$, and if $\mathbf{b}$ is in $\mathbb{R}^m$, the matrix equation

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

has the same solution set as the vector equation

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

which, in turn, has the same solution set as the system of linear equations whose augmented matrix is

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n & \mathbf{b} \end{bmatrix}.$$

From now on, we will use the language and theory of matrices to study systems of linear equations. We reframe the question of existence of solutions in terms of matrices. Given a particular vector **b**, it is easy to deduce from Theorem 1.10 that the equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if **b** is a linear combination of the columns of A.

Example 1.12

Consider the following system of equations

$$\begin{aligned} x_1 + 2x_2 + 7x_3 &= b_1 \\ -2x_1 + 5x_2 + 4x_3 &= b_2, \end{aligned}$$

where b_1 and b_2 are arbitrary real numbers. Let

$$A = \begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_2 \end{bmatrix}, \ \text{and } \ \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Then the system can be written more compactly as the vector equation $A\mathbf{x} = \mathbf{b}$. Notice that **b** is a generic vector. Keeping this vector arbitrary will allow us to deduce general results about the system of equations since these results will not depend on a particular choice of **b**. Here, we will study the existence of solutions by solving the system with a generic vector **b**. To do so, we row reduce the augmented $\begin{bmatrix} A\mathbf{b} \end{bmatrix}$ as follows:

$$\begin{bmatrix} 1 & 2 & 7 & b_1 \\ -2 & 5 & 4 & b_2 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 & b_1 \\ 0 & 9 & 18 & 2b_1 + b_2 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 & b_1 \\ 0 & 1 & 2 & \frac{2b_1+b_2}{9} \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 & \frac{5b_1-2b_2}{9} \\ 0 & 1 & 2 & \frac{2b_1+b_2}{9} \end{bmatrix}$$

The last matrix is the row reduced echelon form which, in turn, corresponds to an augmented matrix of a simpler equivalent system of equations. In fact, it corresponds to the simplest equivalent system. Indeed, if we write down the system that is determined by the row reduced echelon form:

$$\begin{aligned} x_1 + 3x_3 &= \frac{5b_1 - 2b_2}{9} \\ x_2 + 2x_3 &= \frac{2b_1 + b_2}{9}, \end{aligned}$$

it becomes clear that the equations have the least amount of variables in common, the variables are in the correct order, and the coefficients of the leading variables are 1's. The presence of the free variable x_3 indicates that this system has an infinite number of solutions, which are as follows:

$$\begin{aligned} x_1 &= \frac{5b_1 - 2b_2}{9} - 3\lambda, \\ x_2 &= \frac{2b_1 + b_2}{9} - 2\lambda, \\ x_3 &= \lambda,\ \lambda \in \mathbb{R}. \end{aligned}$$

Observe that it is possible to find a solution without putting any restriction on the vector **b**, which allows us to conclude that for each **b** in $\mathbb{R}^2$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution. Moreover, by the definition of the product $A\mathbf{x}$, we can write

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \left(\frac{5b_1 - 2b_2}{9} - 3\lambda\right)\begin{bmatrix} 1 \\ -2 \end{bmatrix} + \left(\frac{2b_1 + b_2}{9} - 2\lambda\right)\begin{bmatrix} 2 \\ 5 \end{bmatrix} + \lambda\begin{bmatrix} 7 \\ 4 \end{bmatrix},\ \lambda \in \mathbb{R}.$$

The most useful interpretation of the equation above is that each **b** in $\mathbb{R}^2$ can be written as a linear combination of the columns of A. In this case, there is an infinite number of ways to write each **b** as a linear combination of the columns of A since there is an infinite number of options for the choice of values for the parameter λ.

You should always be on the look out for important observations when studying the material. In the previous example, observe that the pivot positions in the row reduced echelon form correspond to leading variables and that non pivot positions correspond to free variables. A moment's thought should convince you that this is true in general. This is such an important observation that we must highlight it.

Observation 1.2

The pivot columns of *coefficient matrix* of a system correspond to the leading variables of the system, and the non pivot columns correspond to free variables. Therefore, a system has an infinite number of solutions if its coefficient matrix has at least one non pivot column. Conversely, if every column of the coefficient matrix is a pivot column, then every variable is a leading variable and, thus, the system has a unique solution.

Warning: The last sentence of the observation above is about the ***columns of the coefficient matrix*** not about the augmented matrix nor about the rows of the coefficient matrix. On one hand, if every column of the augmented matrix is a pivot column, then it has a row of the form

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

which corresponds to a degenerate equation of the form $0=1$, which clearly has no solution. On the other hand, the uniqueness of the solution cannot be deduced if the coefficient matrix has a pivot in every row. The system in Example 1.12 has an augmented matrix with a pivot in every row and yet it has an infinite number of solutions.

Example 1.13

Let

$$A=\begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix} \text{ and } \mathbf{b}=\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Is the equation $A\mathbf{x}=\mathbf{b}$ consistent for all possible $\mathbf{b}$?

Row reduce the augment matrix of $A\mathbf{x}=\mathbf{b}$:

$$\begin{bmatrix} 1 & 3 & 4 & b_1 \\ -4 & 2 & -6 & b_2 \\ -3 & -2 & -7 & b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & 4b_1+b_2 \\ 0 & 7 & 5 & 3b_1+b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & 4b_1+b_2 \\ 0 & 0 & 0 & b_1-\frac{1}{2}b_2+b_3 \end{bmatrix}$$

The third row in the last matrix corresponds to a degenerate equation of the form

$$0=b_1-\frac{1}{2}b_2+b_3.$$

If we proceed further applying elementary operations, this row will not change. Therefore, the system is consistent only if the entries of $\mathbf{b}$ satisfy the equation $0=b_1-\frac{1}{2}b_2+b_3$. Clearly, not every $\mathbf{b}$ satisfies this condition (an easy example is $\mathbf{b}=\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^\top$). If A had a pivot in all three rows, we would not care about the calculations in the augmented column because in this case an echelon form of the augmented matrix could not have a row such as $\begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$ corresponding to a degenerate equation with no solutions.

1.5 Solution sets of linear systems

The system of linear equations (1.3) is said to be *homogeneous* if all the independent terms are equal to zero, that is, if it can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is its $m \times n$ coefficient matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. The homogeneous equation $A\mathbf{x} = \mathbf{0}$ always has a solution, namely $\mathbf{x} = \mathbf{0}$, the zero vector in $\mathbb{R}^n$, called the trivial solution. Any other solution, if it exists, is called a non trivial solution.

For a given equation $A\mathbf{x} = \mathbf{0}$, the important question is whether there is a non trivial solution. A non trivial solution exists if and only if the trivial solution is not a unique solution. The proof of Theorem 1.2 leads immediately to the following fact:

Remark 1.5

The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a non trivial solution if and only if the equation has at least one free variable.

Example 1.14

Determine if the following homogeneous system has a non trivial solution. Then describe the solution set.

$$\begin{aligned} x + 2y - 3z + w &= 0, \\ x - 3y + z - 2w &= 0, \\ 2x + y - 3z + 5w &= 0. \end{aligned}$$

Row reduce the augmented matrix to row reduced echelon form:

$$\begin{bmatrix} 1 & 2 & -3 & 1 & 0 \\ 1 & -3 & 1 & -2 & 0 \\ 2 & 1 & -3 & 5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & -3 & 1 & 0 \\ 0 & -5 & 4 & -3 & 0 \\ 0 & -3 & 3 & 3 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & -3 & 1 & 0 \\ 0 & -5 & 4 & -3 & 0 \\ 0 & 1 & -1 & -1 & 0 \end{bmatrix}$$

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

$$\sim \begin{bmatrix} 1 & 0 & 0 & 11 & 0 \\ 0 & 1 & 0 & 7 & 0 \\ 0 & 0 & 1 & 8 & 0 \end{bmatrix}$$

The last matrix is the row reduced echelon form which, in turn, is the augmented matrix of the simplest equivalent system:

$$\begin{aligned} x + 11w &= 0, \\ y + 7w &= 0, \\ z + 8w &= 0. \end{aligned}$$

Notice that w is a free variable. It is possible to deduce this in advanced from the row reduced echelon form since the fourth column is **not** a pivot column (which means that w is not a leading variable; thus, w is free). By the previous remark, the system has a non trivial solution. After solving the simplified system, we deduce that the solution set consists of

$$\begin{aligned} x &= -11\lambda, \\ y &= -7\lambda, \\ z &= -8\lambda, \\ w &= \lambda \in \mathbb{R}, \end{aligned}$$

or, equivalently, the solution of the equation $A\mathbf{x} = 0$ where A is the coefficient matrix of the system, consists of all vector in $\mathbb{R}^4$ of the form

$$\mathbf{x} = \begin{bmatrix} -11\lambda \\ -7\lambda \\ -8\lambda \\ \lambda \end{bmatrix} = \lambda \begin{bmatrix} -11 \\ -7 \\ -8 \\ 1 \end{bmatrix}, \ \lambda \in \mathbb{R}.$$

That is, the solution of $A\mathbf{x} = 0$ consists of all the scalar multiples of the vector

$$\begin{bmatrix} -11 \\ -7 \\ -8 \\ 1 \end{bmatrix}.$$

Example 1.15

A single linear equation can be treated as a system with one equation. Describe the solution set of the homogeneous system

$$10x_1 - 3x_2 - 2x_3 = 0.$$

Obviously, there is no need to write its augmented matrix or to apply elementary row operations. We can directly solve for x_1 (respecting the implicit ordering of the variables). In other words, x_1 is the leading variable and the rest of variables are free:

$$\begin{aligned} x_1 &= \frac{3}{10}\lambda + \frac{1}{5}\mu,\ \lambda, \mu \in \mathbb{R}, \\ x_2 &= \lambda, \\ x_3 &= \mu. \end{aligned}$$

As a vector equation $A\mathbf{x} = 0$. (here $0 = 0 \in \mathbb{R}$), where

$$A = \begin{bmatrix} 10 & -3 & -2 \end{bmatrix}.$$

the general solution is

$$\mathbf{x} = \begin{bmatrix} \frac{3}{10}\lambda + \frac{1}{5}\mu \\ \lambda \\ \mu \end{bmatrix} = \lambda \begin{bmatrix} \frac{3}{10} \\ 1 \\ 0 \end{bmatrix} + \mu \begin{bmatrix} \frac{1}{5} \\ 0 \\ 1 \end{bmatrix}.$$

In words, the general solution consists of all possible linear combinations of two vectors:

$$\begin{bmatrix} \frac{3}{10} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{5} \\ 0 \\ 1 \end{bmatrix}.$$

When a non homogeneous system has more than one solution, two distinct solutions satisfy a simple but important relation between them, and the set of solutions of the homogeneous system characterizes this relation. We make this precise in the following theorem.

Theorem 1.11

Suppose that the equation $A\mathbf{x} = \mathbf{b}$ is consistent for some given non zero $\mathbf{b}$, and let $\mathbf{u}$ and $\mathbf{v}$ be two distinct solutions. Then there is some $\mathbf{h}$ satisfying $A\mathbf{h} = 0$, such that $\mathbf{u} = \mathbf{v} + \mathbf{h}$.

Proof. Let $\mathbf{u}$ and $\mathbf{v}$ be two distinct solutions of the equation $A\mathbf{x} = \mathbf{b}$. Then

$$A(\mathbf{u} - \mathbf{v}) = \mathbf{b} - \mathbf{b} = 0.$$

This means that $\mathbf{h} = \mathbf{u} - \mathbf{v}$ is a solution of the homeneous equation $A\mathbf{x} = 0$. This proves the theorem. ■

Another way to sate the previous theorem is as follows.

Theorem 1.12

Suppose that the equation $A\mathbf{x}=\mathbf{b}$ is consistent for some given non zero $\mathbf{b}$, and let $\mathbf{v}$ be a particular solution. Then the solution set of $A\mathbf{x}=\mathbf{b}$ consists of the set of all vectors of the form $\mathbf{u}=\mathbf{v}+\mathbf{h}$, where $\mathbf{h}$ is any solution of the homogeneous equation $A\mathbf{x}=0$.

Warning: The previous theorem applies only when $A\mathbf{x}=\mathbf{b}$. If the equation has no solution, then the solution set is empty.

Example 1.16

Recall the system in Example 1.12:

$$\begin{aligned} x_1+2x_2+7x_3&=b_1\\ -2x_1+5x_2+4x_3&=b_2, \end{aligned}$$

which is consistent for each $b_1,b_2\in\mathbb{R}$. Its matrix coefficient is

$$A=\begin{bmatrix} 1 & 2 & 7\\ -2 & 5 & 4 \end{bmatrix},$$

and, in vector form, its general solution is

$$\mathbf{x}=\begin{bmatrix} \frac{5b_1-2b_2}{9}\\ \frac{2b_1+b_2}{9}\\ 0 \end{bmatrix}+\lambda\begin{bmatrix} -3\\ -2\\ 1 \end{bmatrix},\ \lambda\in\mathbb{R}$$

We now verify that this general solution is of the form $\mathbf{u}+\mathbf{h}$, where $\mathbf{u}$ satisfies $A\mathbf{u}=\mathbf{b}$ and $\mathbf{h}$ satisfies $A\mathbf{h}=0$. Let

$$\mathbf{u}=\begin{bmatrix} \frac{5b_1-2b_2}{9}\\ \frac{2b_1+b_2}{9}\\ 0 \end{bmatrix}\ \text{and}\ \mathbf{h}=\lambda\begin{bmatrix} -3\\ -2\\ 1 \end{bmatrix}.$$

Then

$$A\mathbf{u}=\begin{bmatrix} 1 & 2 & 7\\ -2 & 5 & 4 \end{bmatrix}\begin{bmatrix} \frac{5b_1-2b_2}{9}\\ \frac{2b_1+b_2}{9}\\ 0 \end{bmatrix}=\left(\frac{5b_1-2b_2}{9}\right)\begin{bmatrix} 1\\ -2 \end{bmatrix}+\left(\frac{2b_1+b_2}{9}\right)\begin{bmatrix} 2\\ 5 \end{bmatrix}+0\begin{bmatrix} 7\\ 4 \end{bmatrix}=\begin{bmatrix} b_1\\ b_2 \end{bmatrix},$$

and

$$A\mathbf{h} = \begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \end{bmatrix} \begin{bmatrix} -3\lambda \\ -2\lambda \\ \lambda \end{bmatrix} = -3\lambda \begin{bmatrix} 1 \\ -2 \end{bmatrix} - 2\lambda \begin{bmatrix} 2 \\ 5 \end{bmatrix} + \lambda \begin{bmatrix} 7 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

1.6 Invertible matrices

An $n \times n$ matrix A is said to be *invertible* if there is an $n \times n$ matrix C such that

$$CA = I \text{ and } AC = I$$

where $I = I_n$ is the identity matrix. In this case, C is an *inverse* of A.

Theorem 1.13

The inverse of an $n \times n$ matrix A is unique.

Proof. Suppose that C and B are both inverses of A. Then,

$$B = BI = B(AC) = (BA)C = IC = C.$$

Therefore, $B = C$ which means that the inverse is uniquely determined by A.
■

Since the inverse of a matrix A is uniquely determined by A, we can assign it a special notation. This unique inverse is denoted by A^{-1}, so that

$$A^{-1}A = I \text{ and } AA^{-1} = I$$

A matrix that is not invertible is sometimes called a *singular* matrix, and an invertible matrix is called a *non singular* matrix.

Invertible matrices are indispensable in linear algebra – mainly for algebraic calculations and formula derivation. In particular, they play an important role in solving systems of equations.

Theorem 1.14

If A is an invertible $n \times n$ matrix, then for each $\mathbf{b} \in \mathbb{R}^r$, the equation $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof. For any $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^n, \mathbf{x} = A^{-1}\mathbf{b}$ is solution since:

$$A\mathbf{x} = A\left(A^{-1}\mathbf{b}\right) = \left(AA^{-1}\right)\mathbf{b} = I\mathbf{b} = \mathbf{b}.$$

If $\mathbf{u}$ is another solution, then

$$A\left(\mathbf{u} - A^{-1}\mathbf{b}\right) = \mathbf{b} - \mathbf{b} = 0.$$

Therefore,

$$\mathbf{u} - A^{-1}\mathbf{b} = I\left(\mathbf{u} - A^{-1}\mathbf{b}\right) = A^{-1}A\left(\mathbf{u} - A^{-1}\mathbf{b}\right) = A^{-1}0 = 0.$$

This implies $\mathbf{u} = A^{-1}\mathbf{b}$, which means that the solution $\mathbf{x} = A^{-1}\mathbf{b}$ is unique. ■

The following theorem provides three useful properties of invertible matrices.

Theorem 1.15

a. If A is an invertible matrix, then A^{-1} is invertible and

$$\left(A^{-1}\right)^{-1} = A.$$

b. If A and B are $n \times n$ invertible matrices, then AB is invertible and

$$(AB)^{-1} = B^{-1}A^{-1}.$$

c. If A is invertible, then $A^{\top}$ is invertible and

$$\left(A^{\top}\right)^{-1} = \left(A^{-1}\right)^{\top}.$$

Proof. a. We already know that A satisfies

$$AA^{-1} = I \text{ and } A^{-1}A = I.$$

Therefore, the inverse of A^{-1} is A, that is, $\left(A^{-1}\right)^{-1} = A$.

b. On one hand,

$$\left(B^{-1}A^{-1}\right)AB = B^{-1}\left(A^{-1}A\right)B = B^{-1}IB = B^{-1}B = I.$$

On the other hand,

$$AB\left(B^{-1}A^{-1}\right) = A\left(B^{-1}B\right)A^{-1} = AIA^{-1} = AA^{-1} = I.$$

Therefore, $(AB)^{-1} = B^{-1}A^{-1}$.

c. We compute

$$\left(A^{-1}\right)^{\top} A^{\top} = \left(AA^{-1}\right)^{\top} = I^{\top} = I.$$

Similarly,

$$A^{\top}\left(A^{-1}\right)^{\top} = \left(A^{-1}A\right)^{\top} = I^{\top} = I.$$

Therefore, $\left(A^{\top}\right)^{-1} = \left(A^{-1}\right)^{\top}$. ■

We have treated inverses abstractly. Now we describe a procedure for computing them.

Theorem 1.16

If A is an invertible matrix, then

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

Consequently, A is invertible if and only if it is row equivalent to the identity matrix I.

Proof. Write $I = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{bmatrix}$. Recall that A is invertible if and only if there is a unique matrix C such that

$$AC = I \text{ and } CA = I.$$

Write $C = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{bmatrix}$. Since

$$AC = \begin{bmatrix} A\mathbf{c}_1 & A\mathbf{c}_2 & \dots & A\mathbf{c}_n \end{bmatrix},$$

then $AC = I$ if and only if all the equations

$$A\mathbf{c}_i = \mathbf{e}_i \text{ for } i = 1,2,\dots,n.$$

have unique solutions. Each equation has an augmented matrix $\begin{bmatrix} A\mathbf{e}_i \end{bmatrix}$. We can apply the same sequence of row elementary operations to reduce all the augmented matrix to a row reduced echelon form. But instead of applying these same sequences of row operations n times, we could just apply them once to the *augmented* matrix

$$\begin{bmatrix} A & \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} A & I \end{bmatrix}.$$

Recall that each $A\mathbf{c}_i = \mathbf{e}_i$ has a unique solution if and only if all the columns of A are pivot columns implying that $A \sim I$ and

$$\begin{bmatrix} A & I \end{bmatrix} \sim \begin{bmatrix} I & \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{bmatrix} = \begin{bmatrix} A & C \end{bmatrix}.$$

This proves the theorem because $C = A^{-1}$. ■

We use this procedure to recover the well-known formula for the inverse of a 2×2 matrix.

Example 1.17

Let

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

Suppose that $ad - bc \neq 0$. We compute

$$\begin{bmatrix} A & I_2 \end{bmatrix} = \begin{bmatrix} a & b & 1 & 0 \\ c & d & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} ac & bc & c & 0 \\ ac & ad & 0 & a \end{bmatrix} \sim \begin{bmatrix} ac & bc & c & 0 \\ 0 & ad-bc & -c & a \end{bmatrix}$$

$$\sim \begin{bmatrix} ac & bc & c & 0 \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix} \sim \begin{bmatrix} ac & 0 & c + \frac{bc^2}{ad-bc} & \frac{-abc}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}$$

$$\sim \begin{bmatrix} ac & 0 & \frac{acd}{ad-bc} & \frac{-abc}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & c + \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}.$$

Hence,

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

We must admit that we cheated. In the last reduction step, we divided by ac without knowing if $ac \neq 0$. Nevertheless, it is possible to deduce the same formula for A^{-1} when $ac = 0$. We leave it to the reader to treat this case. (Hint: consider two cases: $a \neq 0, c = 0$ and $a = 0, c \neq 0$. The case $a = c = 0$ is excluded because this implies $ad - bc = 0$.)

1.7 Determinants

Each $n \times n$ matrix A is assigned a special scalar called the determinant of A, **denoted** by $\det A$ or $|A|$. We give a recursive definition of a determinant.

Definition 1.8

The determinant of a 1×1 matrix $A=\left[a\right]$ is defined as $\left|A\right|=a$. For $n\geq2$, the determinant of an $n\times n$ matrix $A=\left(a_{ij}\right)$ is defined as

$$\det A=\sum_{j=1}^{n}(-1)^{1+j}a_{1j}\det A_{1j},$$

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

Example 1.18

Let us recover the well-known expression for the determinant of a 2×2 matrix. Let

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

Then $A_{11}=d$ and $A_{12}=c$. Using the definition of $\det A$, we get

$$\det A=a\det A_{11}-b\det A_{12}=ad-bc.$$

Example 1.19

Let us show another well-known result. Let

$$A=\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix},\ \mathbf{x}=\begin{bmatrix}x_1\\ x_2\end{bmatrix},\ \text{and}\ \mathbf{b}=\begin{bmatrix}b_1\\ b_2\end{bmatrix}.$$

If $\det A=a_{11}a_{22}-a_{12}a_{21}\neq0$, then it is possible to express the solution of $A\mathbf{x}=\mathbf{b}$ in terms of determinants. This result is known as Cramer's rule. Here we show how to obtain x_1. The remaining variable is left for the reader to compute.

Write $I_2=\left[\mathbf{e}_1\mathbf{e}_2\right]$. Notice that if $\mathbf{x}$ satisfies $A\mathbf{x}=\mathbf{b}$, we have

$$A\left[\mathbf{x}\quad\mathbf{e}_2\right]=\left[A\mathbf{x}\quad A\mathbf{e}_2\right]=\begin{bmatrix}b_1 & a_{12}\\ b_2 & a_{22}\end{bmatrix},$$

were we have used the definition of matrix multiplication. On the other hand, we also have

$$A\left[\mathbf{x}\quad\mathbf{e}_2\right]=\left[A\mathbf{x}\quad A\mathbf{e}_2\right]=\begin{bmatrix}a_{11}x_1+a_{12}x_2 & a_{12}\\ a_{21}x_1+a_{22}x_2 & a_{22}\end{bmatrix}$$

Then,

$$\det\begin{bmatrix} a_{11}x_1 + a_{12}x_2 & a_{12} \\ a_{21}x_1 + a_{22}x_2 & a_{22} \end{bmatrix} = \det\begin{bmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{bmatrix}.$$

We use the definition of determinant to compute the left hand side of the equation:

$$\begin{aligned} \det\begin{bmatrix} a_{11}x_1 + a_{12}x_2 & a_{12} \\ a_{21}x_1 + a_{22}x_2 & a_{22} \end{bmatrix} &= a_{22}\left(a_{11}x_1 + a_{12}x_2\right) - a_{12}\left(a_{21}x_1 + a_{22}x_2\right) \\ &= \left(a_{11}a_{22} - a_{12}a_{21}\right)x_1 + \left(a_{12}a_{22} - a_{12}a_{22}\right)x_2 \\ &= x_1 \det A. \end{aligned}$$

Therefore,

$$x_1 \det A = \det\begin{bmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{bmatrix}$$

which gives us the expression for x_1:

$$x_1 = \frac{\det\begin{bmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{bmatrix}}{\det A}.$$

Observe that the definition of determinant of a matrix A involves the first row of A. It turns out that it is possible to use any row or any column of A to compute its determinant. To state the next theorem, we introduce the following notation. The (i,j)-*cofactor* of A is the number C_{ij} given by

$$C_{ij} = (-1)^{i+j} \det A_{ij}.$$

We omit the proof of the following fundamental theorem to avoid a lenthy digression.

Theorem 1.17

The determinant of an $n \times n$ matrix $A = \left(a_{ij}\right)$ can be computed by a cofactor expansion across any row or down any column. The expansion across the i -th row is

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}.$$

The cofactor expansion down the j -th column is

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

Example 1.20

Use a cofactor expansion to compute $\det A$, where

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

It is always desirable to minimize computational effort. If we choose to expand the determinant down the second column, then the cofactor expansion involves only one term because $a_{12} = a_{32} = 0$:

$$\det A = 4C_{22} = 4 \cdot (-1)^{2+2} \det \begin{bmatrix} 1 & 5 \\ -2 & 2 \end{bmatrix} = 4(12) = 48.$$

Let $A = (a_{ij})$ be an $m \times n$ matrix. We say that a_{ii} are elements in the *main diagonal*. A matrix is said to be *lower triangular* if $a_{ij} = 0$ when $i > j$, that is, when all the elements above the main diagonal are all zero. A matrix is said to be *upper triangular* if $a_{ij} = 0$ when $i < j$, that is, when all the elements below the main diagonal are all zero. Some examples of lower triangular matrices are

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

Some examples of upper triangular matrices are

$$\begin{bmatrix} 1 & 3 & 4 \\ 0 & 2 & 6 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 3 & 5 \\ 0 & 1 & 4 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Matrices in echelon form are upper triangular matrices.

Theorem 1.18

If A is an $n \times n$ triangular matrix, then $\det A$ is the product of the entries on the main diagonal of A.

Proof. We present the proof only for upper triangular matrices. The proof for lower triangular matrix is similar.

We prove this by induction on the size of the matrix. We start by verifying the first non trivial case. Then, for a 2×2 upper triangular matrix, we have

$$\det\begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} = a_{11}a_{12} - 0 \cdot a_{22} = a_{11}a_{12} \cdot$$

Suppose that the theorem has been established for $(k-1)\times(k-1)$ upper triangular matrices. For a $k \times k$ upper triangular matrix, expand its determinant across its last row:

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ 0 & a_{22} & a_{23} & \cdots & a_{2k} \\ 0 & 0 & a_{33} & \cdots & a_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{kk} \end{bmatrix} = a_{kk}C_{kk} = a_{kk}(-1)^{2k}\det\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1(k-1)} \\ 0 & a_{22} & a_{23} & \cdots & a_{2(k-1)} \\ 0 & 0 & a_{33} & \cdots & a_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{(k-1)(k-1)} \end{bmatrix}$$

The last matrix is a $(k-1)\times(k-1)$ upper triangular matrix. So by the induction hypothesis, we obtain

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ 0 & a_{22} & a_{23} & \cdots & a_{2k} \\ 0 & 0 & a_{33} & \cdots & a_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{kk} \end{bmatrix} = a_{kk}a_{11}a_{22}\cdots a_{(k-1)(k-1)} = a_{11}a_{22}\cdots a_{kk}.$$

Then, by mathematical induction, the theorem is proved for all $n \times n$ upper triangular matrices. ■

One of the most important properties of determinants is they "behave well" under the applications of elementary row operations.

Theorem 1.19

Let A be a square matrix.

a. If the matrix B is obtained from A by interchanging two rows, then $\det A = -\det B$.

b. If the matrix B is obtained from A by multiplying one row by a scalar k, then $\det A = \frac{1}{k}\det B \cdot$

c. If the matrix B is obtained from A by adding a multiple of a row to another row, then $\det A = \det B$.

A common strategy in hand calculations is to reduce A to echelon form and then to use the fact that the determinant of a triangular matrix is the product of the entries in its main diagonal.

Example 1.21

Compute $\det A$, where

$$A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}.$$

First, we interchange the first and the last row, which changes the sign of the determinant:

$$\det A = -\det\begin{bmatrix} 1 & -4 & 0 & 6 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 2 & -8 & 6 & 8 \end{bmatrix}.$$

Next, we replace rows in order to produce 0 ' under the pivot in the first column. These replacements do not change the determinant:

$$\det A = -\det\begin{bmatrix} 1 & -4 & 0 & 6 \\ 0 & 3 & 5 & -8 \\ 0 & -12 & 1 & 16 \\ 0 & 0 & 6 & -4 \end{bmatrix}.$$

We continue to replace rows to produce 0 's under $a_{22} = 3$. This does not change the determinant:

$$\det A = -\det\begin{bmatrix} 1 & -4 & 0 & 6 \\ 0 & 3 & 5 & -8 \\ 0 & 0 & 21 & -16 \\ 0 & 0 & 6 & -4 \end{bmatrix}.$$

If we rescale the last row by a factor of $\frac{1}{2}$, the determinant changes as follows:

$$\det A = -2\det\begin{bmatrix} 1 & -4 & 0 & 6 \\ 0 & 3 & 5 & -8 \\ 0 & 0 & 21 & -16 \\ 0 & 0 & 3 & -2 \end{bmatrix}.$$

The third row can be replaced by itself plus -7 times the last row, which does not change the determinant:

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

Finally, we obtain a triangular matrix if we interchange the last two rows, which changes the sign of the determinant:

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

Therefore,

$$\det A = 2(1)(3)(3)(-2) = -36 .$$

We are ready to state the main theorem of this section.

Theorem 1.20

A square matrix A is invertible if and only if $\det A \neq 0$.

Proof. Suppose that A has been reduced to an echelon form U by row replacements and row interchanges (no rescaling). Clearly, this is always possible. Each row interchange produces a change of sign; so if there are r row interchanges, by Theorem 1.19, we have

$$\det A = (-1)^r \det U .$$

Since U is in echelon form, it is an upper triangular matrix, and so $\det U$ is the product of its diagonal entries $u_{11}, u_{22}, \ldots, u_{nn}$.

If A is invertible, then all the entries u_{ii} are pivots because $A \sim I$, which means that $u_{ii} \neq 0$ for all $i = 1, 2, \ldots, n$. Therefore,

$$\det A = (-1)^r u_{11} u_{22} \ldots u_{nn} \neq 0 .$$

If A is not invertible, the at least one entry, say u_{kk}, is not a pivot, which means that $u_{kk} = 0$. Consequently, $\det A = 0$.

We conclude this section by stating two more properties of the determinant.

Theorem 1.21

If A and B are $n \times n$ matrices, then

a. $\det A^\top = \det A$

b. $\det AB = (\det A)(\det B)$