CHAPTER 1

SYSTEMS OF LINEAR EQUATIONS AND MATRICES

The two central problems about which much of the theory of linear algebra revolves are the problem of finding all solutions to a linear system and that of finding an eigensystem for a square matrix. The latter problem requires some background development. By contrast, the former problem is easy to understand and motivate. In this chapter we address the problem of when a linear system has a solution and how to solve such a system for all of its solutions. Examples of linear systems appear in nearly every scientific discipline. We touch on a few of them in this chapter.

1.1. Introduction to Systems of Linear Equations

Any straight line in the xy-plane can be represented algebraically by an equation of the form

$$a_1x + a_2y = b,$$

where a_1, a_2, b are real constants, and a_1 and a_2 are not both zero. An equation of this form is called a linear equation in the variables x and y. More generally,

- **Linear equation**

  A ***linear equation*** *in the n variables $x_1, x_2, \ldots, x_n$ is an equation that can be expressed in the form*

  $$a_1x_1 + a_2x_1 + \cdots + a_nx_n = b,$$

  where $a_1, a_2, \ldots, a_n$ and b are real constants. The variables are sometimes called ***unknowns***.

Example: The equations

$$x + 3y = 7, \quad y = \frac{1}{2}x + 3z + 1, \quad x_1 - 2x_2 - 3x_3 + x_4 = 0$$

are linear. Observe that a linear equation does not involve any product or roots of variables. All variables occur only to the first power and do not appear as arguments for trigonometric, logarithmic, or exponential functions. Thus, the equations

$$x + 3\sqrt{y} = 5, \quad x^2 + y^2 = 1, \quad ax^2 + bx + c = 0, \quad y = \sin x$$

are not linear (called *nonlinear*).■

- **Solution set**

  *A **solution** of a linear equation* $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ *is a sequence of numbers* $s_1, s_2, \ldots, s_n$ *such that the equation is satisfied only when we substitute* $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$. *The set of all solutions of the equation is called the **solution set** or sometimes the **general solution** of the equation.*

Example: Find the general solution of

(1) $4x - 2y = 1$

(2) $x_1 - 4x_2 + 7x_3 = 5$

Solution:

(1) To find solutions, we can assign an arbitrary value to x and solve the equation for y, or choose an arbitrary value for y and solve the equation for x.

If we follow the first approach and assign x an arbitrary value t, then we find

$$x = t, \quad y = 2t - \frac{1}{2}.$$

These formulas describe the general solution in terms of an arbitrary number t, called a *parameter*. Particular numerical values can be obtained by specific values of t. For instance,

$$\begin{aligned} t &= 3 \Rightarrow x = 3, \quad y = \frac{11}{2}, \\ t &= -\frac{1}{2} \Rightarrow x = -\frac{1}{2}, \quad y = -\frac{3}{2}. \end{aligned}$$

If we follow the second approach and assign y the arbitrary value t, then we find

$$x = \frac{1}{2}t + \frac{1}{4}, \quad y = t.$$

Although these formulas are different from those obtained above, they yield the same solution set as t varies over all possible real numbers. For instance, the previous formulas gave the solution $x = 3$, $y = \frac{11}{2}$ when $t = 3$, whereas the formulas immediately above yield that the solution when $t = \frac{11}{2}$.

(2) To find the general solution, we can assign arbitrary values to any two variables and solve for the third one. In particular, if we assign arbitrary values s and t to x_2 and x_3 respectively, and solve for x_1, then we obtain

$$x_1 = 5 + 4s - 7t, \quad x_2 = s, \quad x_3 = t,$$

the general solution.■

- **System of linear equations**

  *A **system of linear equations** of m equations in n variables $x_1, x_2, \ldots, x_n$ is a list of equations of the form*

$$\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1j}x_j & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2j}x_j & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ a_{i1}x_1 & + & a_{i2}x_2 & + & \cdots & + & a_{ij}x_j & \cdots & + & a_{in}x_n & = & b_i \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mj}x_j & \cdots & + & a_{mn}x_n & = & b_m \end{array} \tag{1}$$

  *A sequence of numbers $s_1, s_2, \ldots, s_n$ is called a **solution** of the system if $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ is a solution of all equations in the system.*

Notice that in the i-th row, the coefficient of the j-th variable x_j is a_{ij}, and the right-hand side of the equation is b_i. This systematic way of describing the system is useful when we introduce the matrix concept.

Examples:

(1) The system

$$\begin{array}{ccccccc} 4x_1 & - & 4x_2 & + & 3x_3 & = & -7 \\ 3x_1 & + & x_2 & + & 9x_3 & = & -4 \end{array}$$

has the solution $x_1 = 1, x_2 = 2, x_3 = -1$, since these values satisfy both equations. However, the values $x_1 = -1, x_2 = 8, x_3 = -1$ do not form a solution, since they satisfy only the second equation.

(2) The system

$$\begin{array}{ccccc} x & + & y & = & 4 \\ 2x & + & 2y & = & 6 \end{array}$$

has no solution since if we multiply the second equation by $\frac{1}{2}$, then the resulting equivalent system

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

has contradictory equations.

(3) Consider an open economy consisting of three sectors that supply each other and consumers. Suppose that the three sectors are **E**nergy, **M**aterials, and **S**ervices, and suppose further that the demands of a sector are proportional to its output. The system is said to be in **equilibrium** if the total output of each sector equals the amounts consumed by all sectors and consumers.

For instance, suppose that the following input-output table of demand constants of proportionality and consumer **D**emand (a fixed number) for the output of each sector.

	Consumed by				
		E	M	S	D
Produced by	E	0.2	0.3	0.1	2
	M	0.1	0.3	0.2	1
	S	0.4	0.2	0.1	3

To express the equilibrium of the system we write

$$\begin{array}{rcrcrcrcr} E & = & 0.2E & + & 0.3M & + & 0.1S & + & 2 \\ M & = & 0.1E & + & 0.3M & + & 0.2S & + & 1 \\ S & = & 0.4E & + & 0.2M & + & 0.1S & + & 3 \end{array}$$

or equivalently

$$\begin{array}{rcrcrcr} 0.8E & - & 0.3M & - & 0.1S & = & 2 \\ -0.1E & + & 0.7M & - & 0.2S & = & 1 \\ -0.4E & - & 0.2M & + & 0.9S & = & 3 \end{array}$$

The question that interest economists are whether this system has solutions, and if so, what they are.■

- **Consistent and inconsistent systems**

  A system of linear equations that has at least one solution is said to be ***consistent****. If there is no solution of the system, then the system is called* ***inconsistent****.*

To illustrate the possibilities that can occur in solving a linear system, we consider a general system of two linear equations in the unknowns x and y:

$$\begin{array}{rcrcl} a_1x & + & b_1y & = & c_1 \\ a_2x & + & b_2y & = & c_2 \end{array}$$

The graphs of these equations are lines in the xy-plane, call them l_1 and l_2. There are three possibilities for these lines:

1. The lines l_1 and l_2 may be parallel, in which case there is no intersection, and consequently no solution to the system.

2. The lines l_1 and l_2 may intersect at only one point, in which case the system has exactly one solution.

3. The lines l_1 and l_2 may coincide, in which case there are infinitely many points of intersection, and consequently infinitely many solutions to the system.

We will show later that the same three possibilities hold for arbitrary linear systems.

- **Solutions of linear systems**

  Every system of linear equations has no solutions, or has exactly one solution, or has infinitely many solutions.

1.2. Gaussian Elimination

In this section we consider the main theme of this chapter, which is the systematic solution of linear systems, as defined in equation (1). The principal methodology is the method of *Gaussian elimination* and its variants. The idea of this process is to reduce a system of equations by certain admissible and reversible algebraic operations, called *elementary operations*, to form in which we can easily see what the solutions to the system are, if there are any.

We start with some terminology and notation.

- **Matrices and vectors**

  A ***matrix*** *is a rectangular array of numbers. The numbers that occur in the matrix are called* ***entries****. If a matrix has m* ***rows*** *(horizontal lines) and n* **columns** *(vertical lines), then the* ***size*** *of the matrix is said to be $m \times n$ (read m by n). If the matrix is of size $1 \times n$ or $m \times 1$, then it is called a (row or column, respectively)* ***vector****. If $m = n$, then it is called a* ***square matrix of size*** *n.*

  We write

$$A = [a_{ij}]_{m \times n}$$

  to indicate an $m \times n$ matrix A, and write

$$b = [b_i]$$

  to indicate an n-vector.

For example,

$$\begin{bmatrix} 2 & -1 & 1 \\ 4 & 4 & 20 \end{bmatrix}_{2\times 3} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{3\times 3}$$

are matrices and

$$\begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}_{3\times 1} \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}_{1\times 3}$$

are column and row vectors, respectively.

- **Augmented matrix of a linear system**

  The system (1) *of* m *equations in* n *unknowns can be abbreviated by writing the matrix*

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} & b_i \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} & b_m \end{bmatrix}$$

  This matrix is called the ***augmented matrix*** *corresponding to the system* (1)*.*

For example, the augmented matrix for the system

$$\begin{array}{rcrcrcl} x_1 & + & x_2 & + & 2x_3 & = & 9 \\ 2x_1 & + & 4x_2 & - & 3x_3 & = & 1 \\ 3x_1 & + & 6x_2 & - & 5x_3 & = & 0 \end{array}$$

is

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

When constructing an augmented matrix we must write the coefficients of the unknowns in the same order in each equation, and the constants must be on the right.

The basic method for solving a system of linear equations is to replace the given system by a new system that has the same solution set but is easier to solve. This new system is generally obtained in a series of steps by applying the following three types of operations to eliminate unknowns systematically:

1. Multiply an equation through by a nonzero constant.

2. Interchange two equations.

3. Add a multiple of one equation to another.

Since the rows of an augmented matrix correspond to equations in the associated system, these three operations correspond to following operations on the rows of the augmented matrix:

1. Multiply a row through by a nonzero constant.

2. Interchange two rows.

3. Add a multiple of one row to another.

These operations are called **elementary row operations**.

Example: Consider the system

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ 2x & + & 4y & - & 3z & = & 1 \\ 3x & + & 6y & - & 5z & = & 0 \end{array}$$

Adding -2 times the first equation to the second and -3 times first equation to the third give

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ & & 2y & - & 7z & = & -17 \\ & & 3y & - & 11z & = & -27 \end{array}$$

Multiplying the second equation by $\frac{1}{2}$ gives

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ & & y & - & \frac{7}{2}z & = & -\frac{17}{2} \\ & & 3y & - & 11z & = & -27 \end{array}$$

Adding -3 times the second equation to the third gives

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ & & y & - & \frac{7}{2}z & = & -\frac{17}{2} \\ & & & - & \frac{1}{2}z & = & -\frac{3}{2} \end{array}$$

Multiplying the third equation by -2 gives

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ & & y & - & \frac{7}{2}z & = & -\frac{17}{2} \\ & & & & z & = & 3 \end{array}$$

Adding -1 times the second equation to the first gives

$$\begin{array}{rcrcrcr} x & & & + & \frac{11}{2}z & = & \frac{35}{2} \\ & & y & - & \frac{7}{2}z & = & -\frac{17}{2} \\ & & & & z & = & 3 \end{array}$$

Adding $-\frac{11}{2}$ times the third equation to the first and $\frac{7}{2}$ times the third equation to the second give

$$\begin{aligned} x \qquad\quad &= 1 \\ y \quad &= 2 \\ z &= 3 \end{aligned}$$

Hence, the solution is evidently

$$x = 1, y = 2, z = 3$$

Consider now the corresponding augmented matrix

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{bmatrix}$$

Then following elementary row operations yield

$$\begin{bmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{bmatrix} \xrightarrow[III-3I]{II-2I} \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 3 & -11 & -27 \end{bmatrix} \xrightarrow{(\frac{1}{2})II} \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 3 & -11 & -27 \end{bmatrix} \xrightarrow{III-II} \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & -\frac{1}{2} & -\frac{3}{2} \end{bmatrix}$$

$$\xrightarrow{(-2)III} \begin{bmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & 1 & 3 \end{bmatrix} \xrightarrow{I-II} \begin{bmatrix} 1 & 0 & \frac{11}{2} & \frac{35}{2} \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & 1 & 3 \end{bmatrix} \xrightarrow[II+\frac{7}{2}III]{I-\frac{11}{2}III} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Hence, the solution is

$$x = 1, y = 2, z = 3,$$

as we have verified before.■

In the last example we solved a linear system in the unknowns x, y, and z by reducing the augmented matrix to the form

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

from which the solution $x = 1$, $y = 2$, $z = 3$ became evident. This is an example of a matrix that is in reduced row-echelon form.

- **Row-echelon forms**

  A matrix is in ***reduced row-echelon form*** *if it has the following properties:*

1. *If a row does not consist entirely of zeros, then the first nonzero number in the row is 1. We call this a leading 1.*

2. *If there are any rows that consist entirely of zeros, then they are all grouped together at the bottom of the matrix.*

3. *If there are two rows that do not consist entirely of zeros, then the leading 1 in the lower row occurs farther to the right than the leading 1 in the higher row.*

4. *Each column that contains a leading 1 has zeros everywhere else in that column.*

*A matrix that has the first three properties is said to be in **row-echelon form**. Thus, a matrix in the reduced row-echelon form is of necessity in row-echelon form, but not conversely.*

For example, the following matrices are in reduced row-echelon form.

$$\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & -2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The following matrices are in row-echelon form.

$$\begin{bmatrix} 1 & 4 & 3 & 7 \\ 0 & 1 & 6 & 2 \\ 0 & 0 & 1 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 2 & 6 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Example: Suppose that the augmented matrix for a system of linear equations has been reduced by row operations to the given reduced row-echelon form. Solve the system.

(1) $\begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 4 \end{bmatrix}$ **(2)** $\begin{bmatrix} 1 & 0 & 0 & 4 & -1 \\ 0 & 1 & 0 & 2 & 6 \\ 0 & 0 & 1 & 3 & 2 \end{bmatrix}$ **(3)** $\begin{bmatrix} 1 & 6 & 0 & 0 & 4 & -2 \\ 0 & 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & 5 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ **(4)** $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Solution:

(1) The corresponding linear system is

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

so the solution is $x_1 = 5, x_2 = -2, x_3 = 4$.

(2) The corresponding linear system is

$$\begin{array}{ccccccccc} x_1 & & & & & + & 4x_4 & = & -1 \\ & & x_2 & & & + & 2x_4 & = & 6 \\ & & & & x_3 & + & 3x_4 & = & 2 \end{array}$$

Since x_1, x_2, and x_3 correspond to leading 1's in the augmented matrix, we call them **leading variables**. The nonleading variables (in this case x_4) are called **free variables**. Solving for the leading variables in terms of the free variables gives

$$\begin{array}{ccccccc} x_1 & = & - & 1 & - & 4x_4 \\ x_2 & = & & 6 & - & 2x_4 \\ x_3 & = & & 2 & - & 3x_4 \end{array}$$

From this form of the equations we see that the free variable x_4 can be assigned an arbitrary value, say t, which then determines the values of the leading variables x_1, x_2, and x_3. Thus, there are infinitely many solutions, and the solution set is determined by the formulas

$$x_1 = -1 - 4t, x_2 = 6 - 2t, x_3 = 2 - 3t, x_4 = t$$

(3) The row of zeros leads to the equation

$$0x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 = 0,$$

which places no restrictions on the solutions. Thus, we can omit this equation and write the corresponding system as

$$\begin{array}{ccccccccc} x_1 & + & 6x_2 & & & & + & 4x_5 & = & -2 \\ & & & & x_3 & & + & 3x_5 & = & 1 \\ & & & & & x_4 & + & 5x_5 & = & 2 \end{array}$$

Here the leading variables are x_1, x_3, and x_4, and free variables are x_2 and x_5. Solving for the leading variables in terms of the free variables gives

$$\begin{array}{ccccccccc} x_1 & = & - & 2 & - & 4x_5 & - & 6x_2 \\ x_3 & = & & 1 & - & 3x_5 & & \\ x_4 & = & & 2 & - & 5x_5 & & \end{array}$$

Letting $x_2 = t$ and $x_5 = s$, arbitrary values, there are infinitely many solutions, and the solution set is determined by the formulas

$$x_1 = -2 - 4t - 6s, x_2 = s, x_3 = 1 - 3t, x_4 = 2 - 5t, x_5 = t$$

(4) The last equation in the corresponding system is

$$0x_1 + 0x_2 + 0x_3 = 1$$

Since this equation cannot be satisfied, there is no solution to the system.■

As we have just seen, it is easy to solve a system of linear equations once its augmented matrix is in reduced row-echelon form. If a linear system is solved by reducing its augmented matrix in reduced row-echelon form, then the procedure is called **Gauss-Jordan elimination**. If a linear system is solved by reducing its augmented matrix in row-echelon form, then the procedure is called **Gaussian elimination**.

Example: Using Gauss-Jordan elimination solve the system

$$\begin{array}{rcrcrcrcrcrcr} x_1 & + & 3x_2 & - & 2x_3 & & & + & 2x_5 & & & = & 0 \\ 2x_1 & + & 6x_2 & - & 5x_3 & - & 2x_4 & + & 4x_5 & - & 3x_6 & = & -1 \\ & & & & 5x_3 & + & 10x_4 & & & + & 15x_6 & = & 5 \\ 2x_1 & + & 6x_2 & & & + & 8x_4 & + & 4x_5 & + & 18x_6 & = & 6 \end{array}$$

Solution: The augmented matrix corresponding to the system is

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 & -1 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 2 & 6 & 0 & 8 & 4 & 18 & 6 \end{bmatrix}$$

We have

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 & -1 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 2 & 6 & 0 & 8 & 4 & 18 & 6 \end{bmatrix} \xrightarrow[IV-2I]{II-2I} \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & -2 & 0 & -3 & -1 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 0 & 0 & 4 & 8 & 0 & 18 & 6 \end{bmatrix}$$

$$\xrightarrow{(-1)II} \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 0 & 0 & 4 & 8 & 0 & 18 & 6 \end{bmatrix} \xrightarrow[\substack{III-5II \\ IV-4II}]{I+2II} \begin{bmatrix} 1 & 3 & 0 & 4 & 2 & 6 & 2 \\ 0 & 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 2 \end{bmatrix} \xrightarrow{III \longleftrightarrow IV} \begin{bmatrix} 1 & 3 & 0 & 4 & 2 & 6 & 2 \\ 0 & 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 6 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{(\frac{1}{6})III} \begin{bmatrix} 1 & 3 & 0 & 4 & 2 & 6 & 2 \\ 0 & 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[II-3III]{I-6III} \begin{bmatrix} 1 & 3 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So the corresponding system of equations is

$$\begin{array}{rcrcrcrcrcrcl} x_1 & + & 3x_2 & & & + & 4x_4 & + & 2x_5 & & & = & 0 \\ & & & & x_3 & + & 2x_4 & & & & & = & 0 \\ & & & & & & & & & & x_6 & = & \frac{1}{3} \end{array}$$

Solving for the leading variables in terms of free variables, we find that

$$\begin{array}{rclcrcrcr} x_1 & = & - & 3x_2 & - & 4x_4 & - & 2x_5 \\ x_3 & = & - & 2x_4 & & & & \\ x_6 & = & \frac{1}{3} & & & & & \end{array}$$

Assigning $x_2 = r$, $x_4 = s$, and $x_5 = t$, arbitrary values, we obtain the solution set in the form

$$x_1 = -3r - 4s - 2t, x_2 = r, x_3 = -2s, x_4 = s, x_5 = t, x_6 = \frac{1}{3},$$

implying that there are infinitely many solutions.■

We will show later that every matrix has a unique reduced row-echelon form, that is, one will arrive at the same reduced row-echelon form for a given matrix no matter how the row operations are varied. In contrast, a row-echelon form of a given matrix is not unique, that is, different sequences of row operations can produce different row echelon forms.

Since a matrix has a unique reduced row-echelon form, we can introduce the notion of rank of a matrix.

- **Rank of a matrix**

  The ***rank*** *of a matrix A is the number of nonzero rows of the reduced row-echelon form of A. This number is written as* $\operatorname{rank}(A)$*.*

There are other ways to describe the rank of a matrix. For example, rank can be also defined as the number of columns of the reduced row-echelon form with leading 1's in them, since each leading 1 of a reduced row echelon form occupies a unique column.

In the case that A is a coefficient matrix of a linear system, we can interpret $\operatorname{rank}(A)$ as the number of leading variables.

Example: Find the rank of the matrix

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

Solution: Elementary row operations give

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 2 & 5 \\ 3 & 3 & 2 \end{bmatrix} \xrightarrow[III-3I]{II-2I} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & -4 \end{bmatrix} \overset{(-\frac{1}{4})III}{\longrightarrow} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[III-II]{I-2II} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

the reduced row-echelon form. Thus, $\text{rank}(A) = 2.$■

The rank of a matrix is nonnegative number but it could be 0. This happens if the matrix has only zero rows. Here are some simple limits on the size of $\text{rank}(A)$.

- **Bounds for rank**

  For a matrix A of size $m \times n$, we have

$$0 \leq \text{rank}\,(A) \leq \min\{m, n\}\,.$$

Here is an application of rank concept to the systems of equations.

- **Consistency in terms of rank**

  The general system (1) *with $m \times n$ coefficient matrix A, right-hand side vector b, and augmented matrix $[A|b]$ is consistent if and only if* $\text{rank}(A) = \text{rank}([A|b])$*, in which case either*

  1. $\text{rank}(A) = n$*, in which case the system has a unique solution*

  or

  2. $\text{rank}(A) < n$*, in which case the system has infinitely many solutions depending on* $n - \text{rank}(A)$ *free variables.*

Example: Solve the following linear systems.

(1) $\begin{array}{rcrcrcr} x & + & y & + & z & = & 1 \\ x & - & y & + & 2z & = & 3 \\ x & + & 3y & & & = & 5 \end{array}$ **(2)** $\begin{array}{rcrcrcr} x & + & y & + & z & = & 1 \\ x & - & y & + & 2z & = & 3 \\ x & + & 3y & & & = & -1 \end{array}$ **(3)** $\begin{array}{rcrcrcr} x & + & y & + & z & = & 1 \\ x & - & y & + & 2z & = & 3 \\ x & + & 3y & + & z & = & 5 \end{array}$

Solution:

(1) The augmented matrix is equivalent to

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 2 & 3 \\ 1 & 3 & 0 & 5 \end{bmatrix} \xrightarrow[III-I]{II-I} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 2 \\ 0 & 2 & -1 & 4 \end{bmatrix} \xrightarrow{III+II} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 2 \\ 0 & 0 & 0 & 6 \end{bmatrix} \xrightarrow{(-\frac{1}{2})II} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -\frac{1}{2} & -1 \\ 0 & 0 & 0 & 6 \end{bmatrix}$$
$$\xrightarrow{I-II} \begin{bmatrix} 1 & 0 & \frac{3}{2} & 2 \\ 0 & 1 & -\frac{1}{2} & -1 \\ 0 & 0 & 0 & 6 \end{bmatrix}$$

Last row corresponds to the equation

$$0x + 0y + 0z = 6$$

which is not possible. So the system has no solution. In fact, we observe that $\text{rank}(A) = 2$ while $\text{rank}([A|b]) = 3$.

(2) The augmented matrix is equivalent to

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 2 & 3 \\ 1 & 3 & 0 & -1 \end{bmatrix} \xrightarrow[III-I]{II-I} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 2 \\ 0 & 2 & -1 & -2 \end{bmatrix} \xrightarrow{III+II} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{(-\frac{1}{2})II} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -\frac{1}{2} & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\xrightarrow{I-II} \begin{bmatrix} 1 & 0 & \frac{3}{2} & 2 \\ 0 & 1 & -\frac{1}{2} & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since $\text{rank}(A) = 2 = \text{rank}([A|b])$, the system is consistent. Since there are three unknowns, there are infinitely many solutions depending on one parameter. Since the corresponding system is

$$\begin{array}{rcrcrcl} x & + & y & + & z & = & 1 \\ & - & 2y & + & z & = & 2 \end{array}$$

the solution set is determined by the formulas

$$x = 2 - \frac{3t}{2}, y = \frac{t}{2} - 1, z = t,$$

where t is a parameter.

(3) The augmented matrix is equivalent to

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 2 & 3 \\ 1 & 3 & 1 & 5 \end{bmatrix} \xrightarrow[III-I]{II-I} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 2 \\ 0 & 2 & 0 & 4 \end{bmatrix} \xrightarrow{III+II} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 2 \\ 0 & 0 & 1 & 6 \end{bmatrix} \xrightarrow{\left(-\frac{1}{2}\right)II} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -\frac{1}{2} & -1 \\ 0 & 0 & 1 & 6 \end{bmatrix}$$
$$\xrightarrow{I-II} \begin{bmatrix} 1 & 0 & \frac{3}{2} & 2 \\ 0 & 1 & -\frac{1}{2} & -1 \\ 0 & 0 & 1 & 6 \end{bmatrix} \xrightarrow[II+\frac{1}{2}III]{I-\left(\frac{3}{2}\right)III} \begin{bmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 6 \end{bmatrix}$$

Since $\text{rank}(A) = 3 = \text{rank}([A|b])$, and there are three unknowns, the system has exactly one solution, which is $x = -7, y = 2, z = 6$.■

Example: Determine the solution of the system

$$\begin{array}{ccccccc} kx & + & y & + & z & = & 1 \\ x & + & ky & + & z & = & 1 \\ x & + & y & + & kz & = & 1 \end{array}$$

for the values of k.

Solution: The augmented matrix is equivalent to

$$\begin{bmatrix} k & 1 & 1 & 1 \\ 1 & k & 1 & 1 \\ 1 & 1 & k & 1 \end{bmatrix} \xrightarrow{I \leftrightarrow III} \begin{bmatrix} 1 & 1 & k & 1 \\ 1 & k & 1 & 1 \\ k & 1 & 1 & 1 \end{bmatrix} \xrightarrow[III-kI]{II-I} \begin{bmatrix} 1 & 1 & k & 1 \\ 0 & k-1 & 1-k & 0 \\ 0 & 1-k & 1-k^2 & 1-k \end{bmatrix}$$
$$\xrightarrow{III+II} \begin{bmatrix} 1 & 0 & k & 1 \\ 0 & k-1 & 1-k & 0 \\ 0 & 0 & -(k+2)(k-1) & 1-k \end{bmatrix}$$

Now

·If $k = 1$, then $\text{rank}(A) = 1 = \text{rank}([A|b])$, and there are three unknowns, there are infinitely many solutions depending on two parameters.

·If $k = -2$, then $\text{rank}(A) = 2$, $\text{rank}([A|b]) = 3$, so there is no solution.

·If $k \neq 1$ and $k \neq -2$, then $\text{rank}(A) = 3 = \text{rank}([A|b])$, and there is exactly one solution.■

In general, there is not an easy way to see when a system is consistent outside explicitly solving the system. However, in special cases, there is an easy answer. One such important special case is given by the following.

- **Homogeneous systems**

  A system of linear equations of m equations in n variables $x_1, x_2, \ldots, x_n$ is said to be ***homogeneous*** *if the constant terms are all zero, that is, the system has the form*

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}$$

  Otherwise, the system is said to be ***inhomogeneous****.*

The nice feature of homogeneous systems is that they are always consistent. This is because all such systems have $x_1 = 0, x_2 = 0, \ldots, x_n = 0$ as a solution. This solution is called the **trivial solution**. If there are other solutions, they are called **nontrivial solutions**.

Example: Solve and describe the solution set of the homogeneous system

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

Solution: The augmented matrix is equivalent to

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[III-I]{II-I} \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix} \xrightarrow[(-1)III]{(\frac{1}{2})II} \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \xrightarrow[II+\frac{1}{2}III]{I-III} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

We have $\text{rank}(A) = 3 = \text{rank}([A|b])$, and since there are four unknowns, there are infinitely many solutions depending on one parameter. We have

$$\begin{array}{ccccccc} x_1 & + & x_2 & & & = & 0 \\ & & & x_3 & & = & 0 \\ & & & & x_4 & = & 0 \end{array}$$

so setting $x_2 = t$, the solution set is obtained as $x_1 = -t, x_2 = t, x_3 = 0, x_4 = 0$.■

It is in fact easy to determine whether a homogeneous system has infinitely many solutions.

- *A homogeneous system with more unknown than equations has infinitely many solutions.*

Proof: If the system has more unknowns than equations, then $m < n$, where m and n refer to the size $m \times n$ of the corresponding coefficient matrix. But we know that $\text{rank}(A) \leq \min\{m, n\}$, so $\text{rank}(A) < n$, which implies that there are infinitely many solutions.■

1.3. Applications of Linear Systems

In this section we will discuss some brief applications of the linear system.

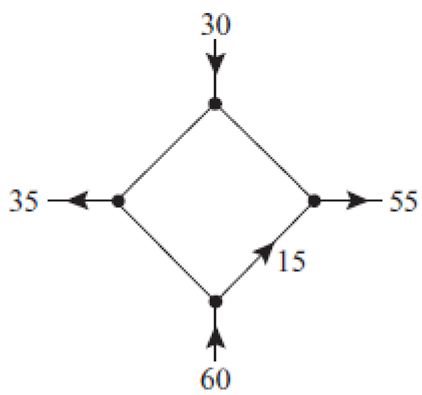
Network Analysis

The concept of a network appears in variety of applications. Roughly speaking, a **network** is a set of branches through which something flows. For example, the branches might be electrical wires through which electricity flows, pipes through which water flows, traffic lanes through which vehicular traffic flows, or economic linkages through which money flows.

In most networks, the branches meet at points, called **nodes** or **junctions**, where the flow divides. For example, in an electrical network, nodes occur where three or more wires join, in a traffic network they occur at street intersections, and in a financial network they occur at banking centers where incoming money is distributed to individuals or other institutions.

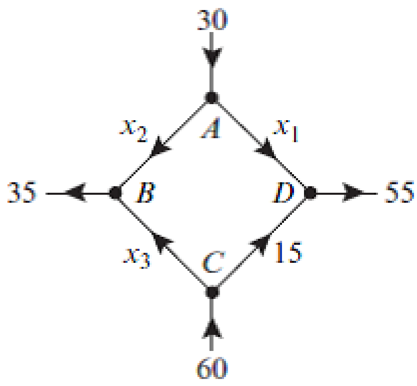
In the study of networks, there is generally some numerical measure of the rate at which the medium flows through a branch. We will restrict our attention to networks in which there is **flow conservation** at each node, by which we mean that the rate of flow into any node is equal to the rate of flow out of that node. This ensures that the flow medium does not build up at the nodes and block the free movement of the medium through the network.

Example: The following figure shows a network.



Find the flow rates and directions of flow in the remaining directions.

Solution: We assign arbitrary directions to the unknown flow rates x_1, x_2, and x_3. We need not be concerned if some of the directions are incorrect since an incorrect direction will be signaled by a negative value for the flow rate when we solve for the unknowns.



By the flow conservation, we have

$$\begin{aligned} &\text{node A}: && x_1 + x_2 = 30 \\ &\text{node B}: && x_2 + x_3 = 35 \\ &\text{node C}: && x_3 + 15 = 60 \\ &\text{node D}: && x_1 + 15 = 55 \end{aligned}$$

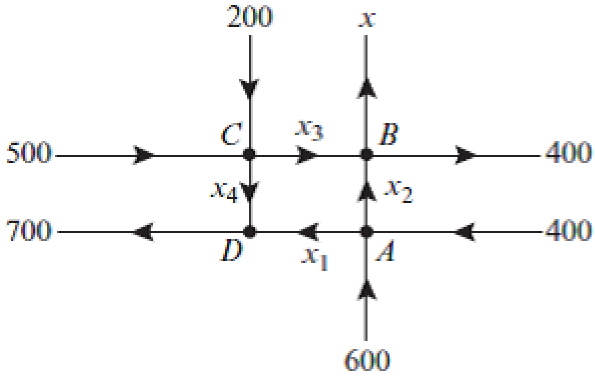
These produce the linear system

$$\begin{array}{ccccccc} x_1 & + & x_2 & & & = & 30 \\ & & x_2 & + & x_3 & = & 35 \\ x_1 & & & & & = & 40 \\ & & & & x_3 & = & 45 \end{array}$$

so $x_1 = 40, x_2 = -10$, and $x_3 = 45$. The fact that x_2 is negative tells us that the direction of the corresponding flow is incorrect.■

Example: The network in the following figure shows a proposed plan for the traffic flow around a new park.

The plan calls for a computerized traffic light at the north exit, and the diagram indicates the average number of vehicles per hour that are expected to flow in and out of the streets that border the complex. All streets are one-way. Find the maximum and minimum number of vehicles through AB route.



Solution: By the flow conservation, we have

$$\begin{aligned}
\text{node A}: &\quad x_1 + x_2 = 400 + 600 \\
\text{node B}: &\quad x_2 + x_3 = 400 + x \\
\text{node C}: &\quad x_3 + x_4 = 500 + 200 \\
\text{node D}: &\quad x_1 + x_4 = 700 \\
\text{the network}: &\quad x + 700 + 400 = 500 + 400 + 600 + 200
\end{aligned}$$

Thus, $x = 600$, and we obtain the following linear system

$$\begin{array}{ccccccccc}
x_1 & + & x_2 & & & & & = & 1000 \\
 & & x_2 & + & x_3 & & & = & 1000 \\
 & & & & x_3 & + & x_4 & = & 700 \\
x_1 & + & & & & & x_4 & = & 700
\end{array}$$

The corresponding augmented matrix is

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1000 \\ 0 & 1 & 1 & 0 & 1000 \\ 0 & 0 & 1 & 1 & 700 \\ 1 & 0 & 0 & 1 & 700 \end{bmatrix} \xrightarrow{IV-I} \begin{bmatrix} 1 & 1 & 0 & 0 & 1000 \\ 0 & 1 & 1 & 0 & 1000 \\ 0 & 0 & 1 & 1 & 700 \\ 0 & -1 & 0 & 1 & -300 \end{bmatrix} \xrightarrow[IV+II]{I-II} \begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1000 \\ 0 & 0 & 1 & 1 & 700 \\ 0 & 0 & 1 & 1 & 700 \end{bmatrix}$$

$$\xrightarrow{IV-III} \begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1000 \\ 0 & 0 & 1 & 1 & 700 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[II-III]{I+III} \begin{bmatrix} 1 & 0 & 0 & 1 & 700 \\ 0 & 1 & 0 & -1 & 300 \\ 0 & 0 & 1 & 1 & 700 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus the corresponding system is

$$\begin{array}{ccccc}
x_1 & + & x_4 & = & 700 \\
x_2 & - & x_4 & = & 300 \\
x_3 & + & x_4 & = & 700
\end{array}$$

which yields

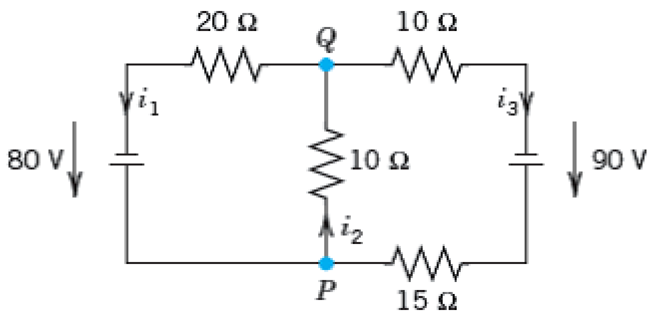
$$x_1 = 700 - t, \quad x_2 = 300 + t, \quad x_3 = 700 - t, \quad x_4 = t.$$

Since the streets are one-way and a negative flow rate indicates a flow in the wrong direction, the average flow rates are all nonnegative. Hence, t satisfies $0 \leq t \leq 700$, and we find that

$$0 \leq x_1 \leq 700, \quad 300 \leq x_2 \leq 1000, \quad 0 \leq x_3 \leq 700, \quad 0 \leq x_4 \leq 700.$$

These imply that the maximum number of vehicles through AB route is 1000, and the minimum number of them is 300.■

Example: Consider the following electrical network in which the currents i_1, i_2, and i_3 are unknowns.



The equations for the currents result from Kirchhoff's laws:

Kirchhoff's Current Law states that at any point of a circuit, the sum of the inflowing currents equals the sum of the outflowing currents.

Kirchhoff's Voltage Law states that in any closed loop, the sum of all voltage drops equals the impressed electromotive force.

By **Ohm's Law**, the voltage drop across a resistor is Ri, where R is the resistance in ohms. Find the currents i_1, i_2, and i_3.

Solution: By Kirchhoff's laws, we have

$$\begin{aligned} \text{node P and Q}: &\quad i_2 = i_1 + i_3 \\ \text{right loop}: &\quad 10i_2 + 25i_3 = 90 \\ \text{left loop}: &\quad 20i_1 + 10i_2 = 80 \end{aligned}$$

Thus, we obtain the following linear system

$$\begin{array}{rcrcrcr} i_1 & - & i_2 & + & i_3 & = & 0 \\ & & 10i_2 & + & 25i_3 & = & 90 \\ 20i_1 & + & 10i_2 & & & = & 80 \end{array}$$

The corresponding augmented matrix is

$$\begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{bmatrix} \xrightarrow{III-20I} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 30 & -20 & 80 \end{bmatrix} \xrightarrow{(\frac{1}{10})II} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & \frac{5}{2} & 9 \\ 0 & 30 & -20 & 80 \end{bmatrix}$$
$$\xrightarrow[III-30II]{I+II} \begin{bmatrix} 1 & 0 & \frac{7}{2} & 9 \\ 0 & 1 & \frac{5}{2} & 9 \\ 0 & 0 & -95 & -190 \end{bmatrix} \xrightarrow{(-\frac{1}{95})III} \begin{bmatrix} 1 & 0 & \frac{7}{2} & 9 \\ 0 & 1 & \frac{5}{2} & 9 \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow[II-(\frac{5}{2})III]{I-(\frac{7}{2})III} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

which gives $i_1 = 2$, $i_2 = 4$, and $i_3 = 2$■

Balancing Chemical Equations

Chemical compounds are represented by chemical formulas that describe the atomic makeup of their molecules. For example, water is composed of two hydrogen atoms and one oxygen atom, so its chemical formula is H_2O.

When chemical compounds are combined under the right conditions, the atoms in their molecules rearrange to form new compounds. For example, when methane burns, the methane (CH_4) and stable oxygen (O_2) react to form carbon dioxide (CO_2) and water (H_2O). This is indicated by the chemical equation

$$CH_4 + O_2 \to CO_2 + H_2O.$$

The molecules to the left of the arrow are called the **reactants** and those to the right the **products**. In this equation the plus signs serve to separate the molecules and are not intended as algebraic operations. However, this equation does not tell the whole story, since it fails to account for the proportions of molecules required for a complete reaction (no reactants left over).

A chemical equation is said to be **balanced** if for each type of atom in the reaction, the same number of atoms appears on each side of the arrow. For example,

$$CH_4 + O_2 \to CO_2 + H_2O$$

is balanced as

$$CH_4 + 2O_2 \to CO_2 + 2H_2O.$$

The equation $CH_4+O_2 \to CO_2+H_2O$ is sufficiently simple that it could have been balanced by trial and error, but more complicated chemical equations can be balanced using systems of linear equations.

Example: Balance the chemical equation

$$\mathrm{HCl} + \mathrm{Na_3PO_4} \to \mathrm{H_3PO_4} + \mathrm{NaCl}.$$

Solution: Let x_1, x_2, x_3, and x_4 be positive integers that balance the equation:

$$x_1\,(\mathrm{HCl}) + x_2\,(\mathrm{Na_3PO_4}) \to x_3\,(\mathrm{H_3PO_4}) + x_4\,(\mathrm{NaCl})\,.$$

Equating the number of atoms of each type on the two sides yields

$$\begin{aligned} \mathrm{H}: &\quad x_1 = 3x_3 \\ \mathrm{Cl}: &\quad x_1 = x_4 \\ \mathrm{Na}: &\quad 3x_2 = x_4 \\ \mathrm{P}: &\quad x_2 = x_3 \\ \mathrm{O}: &\quad 4x_2 = 4x_3 \end{aligned}$$

from which we obtain the homogeneous linear system

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

The corresponding augmented matrix is

$$\begin{bmatrix} 1 & 0 & -3 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 3 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \end{bmatrix} \xrightarrow[III-3IV]{II-I} \begin{bmatrix} 1 & 0 & -3 & 0 & 0 \\ 0 & 0 & 3 & -1 & 0 \\ 0 & 0 & 3 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \end{bmatrix} \xrightarrow{III-II} \begin{bmatrix} 1 & 0 & -3 & 0 & 0 \\ 0 & 0 & 3 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} 1 & 0 & -3 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 3 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\frac{1}{3}III} \begin{bmatrix} 1 & 0 & -3 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -\frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[II+III]{I+3III} \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & -\frac{1}{3} & 0 \\ 0 & 0 & 1 & -\frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

from which we conclude that the general solution of the system is

$$x_1 = t, \quad x_2 = \frac{t}{3}, \quad x_3 = \frac{t}{3}, \quad x_4 = t,$$

where t is arbitrary. To obtain the smallest positive integer that balance the equation, we set $t = 3$, which gives $x_1 = 3$, $x_2 = 1$, $x_3 = 1$, $x_4 = 3$, and the balanced equation is of the form $3HCl + Na_3PO_4 \rightarrow H_3PO_4 + 3NaCl$.■

Polynomial Interpolation

An important problem in various applications is to find a polynomial whose graph passes through a specified set of points in the plane. This polynomial is called an **interpolating polynomial** for the points.

We consider the problem of finding a polynomial whose graph passes through n points with distinct x-coordinates

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

Since there are n conditions to be satisfied, intuition suggests that we should begin by looking for a polynomial of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}.$$

The following is the basic result on polynomial interpolation.

- *Given any n points in the xy-plane that have distinct x-coordinates, there is a unique polynomial of degree $n-1$ or less whose graph passes through those points.*

Since the graph of the polynomial $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}$ is the graph of the equation

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1},$$

it follows that the coordinates of the points must satisfy

$$\begin{array}{ccccccccccc} a_0 & + & a_1 x_1 & + & a_2 x_1^2 & + & \cdots & + & a_{n-1} x_1^{n-1} & = & y_1 \\ a_0 & + & a_1 x_2 & + & a_2 x_2^2 & + & \cdots & + & a_{n-1} x_2^{n-1} & = & y_2 \\ \vdots & & \vdots & & \vdots & & & & \vdots & & \vdots \\ a_0 & + & a_1 x_n & + & a_2 x_n^2 & + & \cdots & + & a_{n-1} x_n^{n-1} & = & y_n \end{array}$$

Since x's and y's are known, we can view this as a linear system in the unknowns $a_0, a_1, \ldots, a_{n-1}$. The augmented matrix for the system is

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} & y_1 \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} & y_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} & y_n \end{bmatrix}$$

Hence the interpolating polynomial can be found by reducing this matrix to reduced row-echelon form.

Example: Find a cubic polynomial whose graph passes through the points $(1,3), (2,-2), (3,-5)$, and $(4,0)$.

Solution: Let the desired polynomial be $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$. The augmented matrix for the linear system in the unknowns a_0, a_1, a_2, a_3 is

$$\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & y_1 \\ 1 & x_2 & x_2^2 & x_2^3 & y_2 \\ 1 & x_3 & x_3^2 & x_3^3 & y_3 \\ 1 & x_4 & x_4^2 & x_4^3 & y_4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 3 \\ 1 & 2 & 4 & 8 & -2 \\ 1 & 3 & 9 & 27 & -5 \\ 1 & 4 & 16 & 64 & 0 \end{bmatrix}$$

The reduced row-echelon form of this matrix is

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

so the interpolating polynomial is $p(x) = 4 + 3x - 5x^2 + x^3$.■

1.4. Matrices and Matrix Operations

Matrices arise in many contexts other than as augmented matrices for systems of linear equations. In this section we begin our study of matrix theory by giving some of the fundamental definitions and results of the subject.

We recall the following definition from Section 1.2.

- **Matrices and vectors**

  A ***matrix*** *is a rectangular array of numbers. The numbers that occur in the matrix are called* ***entries****. If a matrix has m* ***rows*** *(horizontal lines) and n* ***columns*** *(vertical lines), then the* ***size*** *of the matrix is said to be $m \times n$ (read m by n). If the matrix is of size $1 \times n$ or $m \times 1$, then it is called a (row or column, respectively)* ***vector****. If $m = n$, then it is called a* ***square matrix of size*** *n.*

  We write

$$A = [a_{ij}]_{m \times n}$$

to indicate an $m \times n$ matrix A, and write

$$b = [b_i]$$

to indicate an n-vector.

The notational style of writing a matrix in the form $A = [a_{ij}]$ hints that a matrix could be treated like a single number.

- **Equality of matrices**

  Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are said to be ***equal*** *if these matrices have the same size, and for each pair (i, j), $a_{ij} = b_{ij}$, that is, corresponding entries of A and B are equal.*

Note that matrices of different sizes are not equal.

- **Matrix addition and substraction**

  Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times n$ matrices. The ***sum*** *of the matrices, denoted by $A + B$, is the $m \times n$ matrix defined by the formula*

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

  The ***negative*** *of the matrix A, denoted by $-A$, is defined by the formula*

$$-A = [-a_{ij}] .$$

  The ***difference*** *of A and B, denoted by $A - B$, is defined by the formula*

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

We again note that matrices of different sized cannot be added or subtracted.

- **Scalar multiplication**

  Let $A = [a_{ij}]$ be an $m \times n$ matrix and c be a scalar. The ***product of the scalar*** *c* ***with the matrix*** *A, denoted by cA, is defined by the formula*

$$cA = [ca_{ij}] .$$

Combining the ideas of scalar multiplication and addition, we introduce a very fundamental notion in linear algebra.

- **Linear combination**

  *A **linear combination** of the matrices $A_1, A_2, \ldots, A_n$ of the same size is an expression of the form*

  $$c_1A_1 + c_2A_2 + \cdots + c_nA_n,$$

  where $c_1, c_2, \ldots, c_n$ are scalars.

The linear combination idea has a really useful application to linear systems, namely, it gives us another way to express the solution set of a linear system that clearly identifies the role of free variables.

Example: Suppose that a linear system in the unknowns x_1, x_2, x_3, x_4 has general solution

$$(x_2 + 3x_4, x_2, 2x_2 - x_4, x_4),$$

where the variables x_2 and x_4 are free. Describe the solution set of this linear system in terms of linear combinations with free variables as coefficients.

Solution: We use only the parts of the general solution involving x_2 for one vector and the parts involving x_4 as the other vector in a such way that these vectors add up to general solution. We have

$$\begin{bmatrix} x_2 + 3x_4 \\ x_2 \\ 2x_2 - x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_2 \\ 2x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 3x_4 \\ 0 \\ -x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} 1 \\ 1 \\ 2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 3 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

Now we simply define the vectors $A_1 = (1, 1, 2, 0)$, $A_2 = (3, 0, -1, 1)$, and we see that since x_2 and x_4 are arbitrary, the solution set is

$$S = \{x_2A_1 + x_4A_2 : x_2, x_4 \in \mathbb{R}\}.$$

In other words, the solution set to the system is the set of all possible linear combinations of the vectors A_1 and A_2.■

- **Zero matrix**

  *A **zero matrix** is a matrix whose every entry is 0. Such matrices are denoted by 0.*

The laws for addition and scalar multiplication are pretty much what we would expect them to be.

- **Laws of matrix addition and scalar multiplication**

  Let A, B, C be matrices of the same size $m \times n$, 0 be the $m \times n$ zero matrix, and c and d be scalars.

  (1) *$A + B$ is an $m \times n$ matrix (closure law)*

  (2) *$(A + B) + C = A + (B + C)$ (associative law)*

  (3) *$A + B = B + A$ (commutative law)*

  (4) *$A + 0 = A$ (identity law)*

  (5) *$A + (-A) = 0$ (inverse law)*

  (6) *cA is an $m \times n$ matrix (closure law)*

  (7) *$c\,(dA) = (cd)\,A$ (associative law)*

  (8) *$(c + d)\,A = cA + dA$ (distributive law)*

  (9) *$c\,(A + B) = cA + cB$ (distributive law)*

  (10) *$1A = A$*

Matrix multiplication is somewhat not obvious comparing to matrix addition and scalar multiplication. To motivate the definition, we consider a single linear equation

$$2x - 3y + 4z = 5.$$

We will find it useful to think of the left-hand side of the equation as a product of the coefficient matrix $\begin{bmatrix} 2 & -3 & 4 \end{bmatrix}$ and the column matrix of unknowns $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$. Thus

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

This motivates the following definition.

- **Row-column product**

  The ***product*** *of the $1 \times n$ row $\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$ with the $n \times 1$ column $\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ is defined to be the 1×1 matrix $\begin{bmatrix} a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \end{bmatrix}$.*

This row-column product strategy guides us to the general definition. We note that the column number of the first matrix matches the row number of the second, and this number disappears in the size of the resulting product. This is exactly what happens in general.

- **Matrix product**

  *Let $A = [a_{ij}]$ be an $m \times r$ matrix and $B = [b_{ij}]$ be an $r \times n$ matrix. The **product** of the matrices A and B, denoted by AB, is the $m \times n$ matrix $C = [c_{ij}]$ with entries*

$$c_{ij} = \sum_{k=1}^{r} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ir}b_{rj},$$

  where $1 \le i \le m$ and $1 \le j \le n$.

Notice that, in contrast to the case of addition, two matrices may be of different sizes when we can multiply them together. It is also worth noticing that if A and B are square and of the same size, then the products AB and BA are always defined.

Example: Compute, if possible, the products of the following pairs of matrices

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

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

Solution:

(1) A is of size 2×3 and B is of size 3×2, so AB and BA are defined:

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

$$BA = \begin{bmatrix} 4 & -2 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 6 \\ 2 & 3 & -1 \\ 4 & 7 & 1 \end{bmatrix}_{3\times 3}$$

(2) A is of size 2×3 and B is of size 2×1, so neither AB nor BA is defined.

(3) A is of size 2×2 and B is of size 2×3, so AB is defined but BA is not defined:

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix}_{2\times 3}$$

(4) A and B are of size 2×2, so AB and BA are defined:

$$\begin{aligned} AB &= \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}_{2\times 2} \\ BA &= \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ -2 & -2 \end{bmatrix}_{2\times 2} \end{aligned} \quad \blacksquare$$

Matrix multiplication is not commutative, that is $AB \neq BA$ in general (check (1) and (4) of the above example). It is also interesting to note that $AB = 0$ does not necessarily imply $BA = 0$ or $A = 0$ or $B = 0$ (check (4) of the above example).

The calculation in (3) of the above example inspires some more notation.

- **Identity matrix**

  A matrix of the form

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} = [\delta_{ij}]$$

  is called an $n \times n$ ***identity matrix****. The (i,j)th entry of I_n is designated by the Kronecker symbol δ_{ij}:*

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Identity matrices arise naturally in studying reduced row-echelon forms of square matrices.

- *If R is the reduced row-echelon form of an $n \times n$ matrix A, then either R has a row of zeros or R is the identity matrix I_n.*

Proof: Suppose that the reduced row-echelon form of A is R, then either the last row in R consists entirely of zeros or it does not. If not, the matrix contains no zero rows, and consequently each of the n rows has a leading entry of 1. Since these leading 1's occur progressively farther to the right as we move down the matrix, each of these 1's must occur on the main diagonal. Since the other entries in the same column as one of these 1's are zero, R must be I_n. Thus, either R has a row of zeros or $R = I_n$.$\blacksquare$

An identity matrix is in fact a special form of some particular matrix.

- **Triangular, diagonal, and scalar matrices**

  *A square matrix in which all the entries above the main diagonal are zero is called **lower triangular**, and a square matrix in which all the entries below the main diagonal are zero is called **upper triangular**. A matrix that is either upper triangular or lower triangular is called **triangular**. A square matrix in which all the entries off the main diagonal are zero is called a **diagonal matrix**.*

  *If all diagonal entries of a diagonal matrix S are equal, say c, then we call S a **scalar matrix** because multiplication of any square matrix A of the same size by S has the same effect as the multiplication by a scalar, that is,*

$$AS = SA = cA.$$

  *An **identity matrix** is a scalar matrix with scalar 1.*

We note that the product of lower triangular matrices is lower triangular, and the product of upper triangular matrices is upper triangular.

Another type of matrix that occurs frequently enough to be discussed is a **block matrix**. Actually, we already used the idea of blocks when we described the augmented matrix $[A|b]$ of a linear system. We say that the augmented matrix has the block or partitioned form $[A|b]$. To partitioning a matrix, we insert vertical and horizontal lines between entries.

Example: The matrix

$$A = \left[\begin{array}{rrr|rr|r} 3 & 0 & -1 & 5 & 9 & -2 \\ -5 & 2 & 4 & 0 & -3 & 1 \\ \hline -8 & -6 & 3 & 1 & 7 & -4 \end{array}\right]$$

can be also written as the 2×3 block matrix

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}$$

whose entries are the blocks (or submatrices)

$$\begin{aligned} A_{11} &= \begin{bmatrix} 3 & 0 & -1 \\ -5 & 2 & 4 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 5 & 9 \\ 0 & -3 \end{bmatrix}, \quad A_{13} = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \\ A_{21} &= \begin{bmatrix} -8 & -6 & 3 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 1 & 7 \end{bmatrix}, \quad A_{23} = \begin{bmatrix} -4 \end{bmatrix} \blacksquare \end{aligned}$$

If matrices A and B are the same size and are partitioned in exactly same way, then it is natural to make the same partition of the ordinary matrix sum $A + B$. In this case, each block of $A + B$ is the matrix sum of the corresponding blocks of A and B. Multiplication of a partitioned matrix by a scalar is also computed block by block.

Partitioned matrices can be multiplied by the usual row-column rule as if the block entries we scalar, provided that for a product AB, the column partition of A matches the row partition of B.

Example: Find AB where

$$A = \begin{bmatrix} 2 & -3 & 1 & 0 & -4 \\ 1 & 5 & -2 & 3 & -1 \\ 0 & -4 & -2 & 7 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 6 & 4 \\ -2 & 1 \\ -3 & 7 \\ -1 & 3 \\ 5 & 2 \end{bmatrix}$$

Solution: We partitioning A and B as follows.

$$A = \left[\begin{array}{ccc|cc} 2 & -3 & 1 & 0 & 4 \\ 1 & 5 & -2 & 3 & -1 \\ \hline 0 & -4 & -2 & 7 & 1 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$B = \left[\begin{array}{cc} 6 & 4 \\ -2 & 1 \\ -3 & 7 \\ \hline -1 & 3 \\ 5 & 2 \end{array}\right] = \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}$$

Thus

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} \\ A_{21}B_{11} + A_{22}B_{21} \end{bmatrix},$$

where

$$A_{11}B_{11} = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 5 & -2 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ -2 & 1 \\ -3 & 7 \end{bmatrix} = \begin{bmatrix} 15 & 12 \\ 2 & -5 \end{bmatrix}$$

$$A_{12}B_{21} = \begin{bmatrix} 0 & -4 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} -20 & -8 \\ -8 & 7 \end{bmatrix}$$

$$A_{21}B_{11} = \begin{bmatrix} 0 & -4 & -2 \end{bmatrix} \begin{bmatrix} 6 & 4 \\ -2 & 1 \\ -3 & 7 \end{bmatrix} = \begin{bmatrix} 14 & -18 \end{bmatrix}$$

$$A_{22}B_{21} = \begin{bmatrix} 7 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} -2 & 23 \end{bmatrix},$$

which give

$$AB = \begin{bmatrix} -5 & 4 \\ -6 & 2 \\ 12 & 5 \end{bmatrix} \blacksquare$$

Example: Find the matrix product

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

Solution: We have

$$\left[\begin{array}{cc|cc} 1 & 2 & 0 & 0 \\ 3 & 4 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \end{array}\right] \left[\begin{array}{cc|cc} 0 & 0 & 2 & 1 \\ 0 & 0 & -1 & 1 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} 0 & B_{12} \\ 0 & B_{22} \end{bmatrix} = \begin{bmatrix} 0 & A_{11}B_{12} \\ 0 & A_{22}B_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 3 \\ 0 & 0 & 2 & 7 \\ 0 & 0 & 1 & 0 \end{bmatrix} \blacksquare$$

Block matrices provide another way to view matrix multiplication. Suppose that an $m \times r$ matrix A is partitioned into its r column vectors $c_1, c_2, \ldots, c_r$ (each of size $m \times 1$) and an $r \times n$ matrix B is partitioned into its r row vectors $r_1, r_2, \ldots, r_r$ (each of size $1 \times n$). Each term in the sum

$$c_1 r_1 + c_2 r_2 + \cdots + c_r r_r$$

has size $m \times n$ so the sum itself is an $m \times n$ matrix. The entry in row i and column j of the sum is given by the expression

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ir}b_{rj},$$

from which it follows that

$$AB = c_1 r_1 + c_2 r_2 + \cdots + c_r r_r.$$

We call this as the **column-row expansion** of AB.

Example: Find AB where

$$A = \begin{bmatrix} 2 & -3 & 1 & 0 & -4 \\ 1 & 5 & -2 & 3 & -1 \\ 0 & -4 & -2 & 7 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 6 & 4 \\ -2 & 1 \\ -3 & 7 \\ -1 & 3 \\ 5 & 2 \end{bmatrix}$$

Solution: Using the column-row expansion, we find that

$$\begin{aligned} AB &= \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 6 & 4 \end{bmatrix} + \begin{bmatrix} -3 \\ 5 \\ -4 \end{bmatrix} \begin{bmatrix} -2 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \\ -2 \end{bmatrix} \begin{bmatrix} -3 & 7 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \\ 7 \end{bmatrix} \begin{bmatrix} -1 & 3 \end{bmatrix} + \begin{bmatrix} -4 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} 5 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 12 & 8 \\ 6 & 4 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 6 & -3 \\ -10 & 5 \\ 8 & -4 \end{bmatrix} + \begin{bmatrix} -3 & 7 \\ 6 & -14 \\ 6 & -14 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -3 & 9 \\ -7 & 21 \end{bmatrix} + \begin{bmatrix} -20 & -8 \\ -5 & -2 \\ 5 & 2 \end{bmatrix} = \begin{bmatrix} -5 & 4 \\ -6 & 2 \\ 12 & 5 \end{bmatrix} \blacksquare \end{aligned}$$

Matrix multiplication can be viewed as linear combinations. Let A be an $m \times n$ matrix and x be an $n \times 1$ matrix, say

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Then

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

This gives the following result.

- **Matrix product as linear combinations**

  If A is an $m \times n$ matrix, and if x is an $n \times 1$ matrix, then the product Ax can be expressed as a linear combination of the column vectors of A in which the coefficients are the entries of x.

Matrix multiplication has important application to linear systems. Consider the following system of m equations in n unknowns

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

Since two matrices are equal if and only if their corresponding entries are equal, we can replace the

m equations by the single matrix equation

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

The $m \times 1$ matrix on the left side can be written as as product to give

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

If we denote these matrices by A, x, and b, respectively, then we can replace the original system by the single matrix equation

$$Ax = b.$$

Matrix multiplication enables us to prove the uniqueness of a reduced row-echelon form of a matrix.

- **Uniqueness of a reduced row-echelon form**

  The reduced row-echelon form of a matrix is unique.

Proof: Let A be an $m \times n$ matrix. We will proceed by induction on n. For $n = 1$, the proof is obvious. Suppose that $n > 1$. Let A' be the matrix obtained from A by deleting the nth column. We observe that any sequence of elementary row operations which converts A in reduced row-echelon form also converts A' in reduced row-echelon form. Thus, by induction, if B and C are reduced row-echelon forms of A, then they can differ in the nth column only. Assume that $B \neq C$. Then there is a row, say rth row, of B which is not equal to the rth row of C. Let u be any column vector such that $Bu = \mathbf{0}$. Then $Cu = \mathbf{0}$ and hence $(B - C)\,u = \mathbf{0}$. We observe that the first $n-1$ columns of $B - C$ are zero columns. Thus the rth coordinate of $(B - C)\,u$ is $(b_{rn} - c_{rn})\,u_n = 0$. Since $b_{rn} \neq c_{rn}$, we must have $u_n = 0$. Thus the nth columns of B and C must involve a leading 1 otherwise they are free columns and u_n takes on arbitrary values, not 0. But since the first $n-1$ columns of B and C are identical, the row in which this leading 1 must appear must be the same for both B and C, namely, the row which is the first zero row of reduced row-echelon form of A'. Since the remaining entries in the nth columns of B and C must be all zero, we have $B = C$, which is a contradiction.■

Matrix multiplication can be used to carry out an elementary row operation for a special type of matrix.

- **Elementary matrix**

  *An $n \times n$ matrix is called an **elementary matrix** if it can be obtained from the $n \times n$ identity matrix I_n by performing a single elementary row operation.*

For example, the following matrices are elementary.

$\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$ multiply the second row of I_2 by -3

$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ interchange the second and fourth rows of I_4

$\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ add 3 times of the third row of I_3 to the first row

The following result shows that when a matrix A is multiplied on the left by an elementary matrix E, the effect is to perform an elementary row operation on A.

- **Elementary matrices and row operations**

  If an elementary matrix E results from performing a certain row operation on I_m and if A is an $m \times n$ matrix, then the product EA is the matrix that results when this same row operation is performed on A.

Example: Consider the matrix

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

and the elementary matrix

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

which results from adding 3 times the first row of I_3 to the third row. The product EA is

$$EA = \begin{bmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 4 & 4 & 10 & 6 \end{bmatrix}$$

which is precisely same matrix that results when we add 3 times of the first row of A to the third row.■

Sometimes we prefer to work with a different form of a given matrix that contains the same information. Transposes are operations that allow us to do that.

- **Transpose of a matrix**

  *If A is any $m \times n$ matrix, then the **transpose** of A, denoted by A^T, is defined to be the $n \times m$ matrix that results by interchanging rows and columns of A, that is, the first row of A^T is the first column of A, the second column of A^T is the second row of A, and so forth.*

For example, we have

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

$$B = \begin{bmatrix} 2 & 3 \\ 1 & 4 \\ 5 & 6 \end{bmatrix} \Rightarrow B^T = \begin{bmatrix} 2 & 1 & 5 \\ 3 & 4 & 6 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 3 & 5 \end{bmatrix} \Rightarrow C^T = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$$

We note that the transpose of a lower triangular matrix is upper triangular, and the transpose of an upper triangular matrix is lower triangular.

Even when dealing with vectors alone, the transpose notation is useful.

- **Inner and outer products**

  *Let u and v be column vectors of the same size, say $n \times 1$. The **inner product** of u and v is the scalar quantity $u^T v$, and the **outer product** of u and v is the $n \times n$ matrix uv^T.*

For example, if

$$u = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \quad \text{and} \quad v = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$$

are 3×1 vectors, then the inner product of them is

$$u^T v = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} = 3$$

while the outer product is

$$uv^T = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 8 & 2 \\ -3 & -4 & -1 \\ 3 & 4 & 1 \end{bmatrix}.$$

Here are a few basic laws relating transposes to other matrix arithmetic.

- **Laws of matrix transpose**

  Assuming that the sizes of A and B are such that the indicated operations can be performed, the following rules are valid.

  (1) $(A+B)^T = A^T + B^T$

  (2) $(AB)^T = B^T A^T$

  (3) $(cA)^T = cA^T$, *c constant*

  (4) $\left(A^T\right)^T = A$

A very important type of special matrix is one that is invariant under operation of transposing.

- **Symmetric and skew-symmetric matrices**

  *The matrix A is said to be **symmetric** if $A^T = A$, and **skew-symmetric** if $A^T = -A$.*

Example: Let A be a square matrix. Show the following.

(a) The matrix $B = \frac{1}{2}\left(A + A^T\right)$ is symmetric.

(b) The matrix $C = \frac{1}{2}\left(A - A^T\right)$ is skew-symmetric.

(c) The matrix A can be expressed as a sum of a symmetric matrix and a skew-symmetric matrix.

(d) Express the matrix

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

as a sum of a symmetric and a skew-symmetric matrix.

Solution:

(a) Since

$$B^T = \left[\frac{1}{2}\left(A + A^T\right)\right]^T = \frac{1}{2}\left(A^T + A\right) = B,$$

B is symmetric.

(b) Since

$$C^T = \left[\frac{1}{2}\left(A - A^T\right)\right]^T = \frac{1}{2}\left(A^T - A\right) = -\frac{1}{2}\left(A - A^T\right) = -C,$$

C is skew-symmetric.

(c) We have

$$\begin{aligned} B &= \tfrac{1}{2}\left(A + A^T\right) \\ C &= \tfrac{1}{2}\left(A - A^T\right) \end{aligned} \Rightarrow A = B + C.$$

(d) Since

$$B = \frac{1}{2}\left(A + A^T\right) = \frac{1}{2}\begin{bmatrix} 2 & 2 & -6 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 2 & 0 & 0 \\ 2 & 1 & 0 \\ -6 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -3 \\ 1 & 1 & 2 \\ -3 & 2 & 1 \end{bmatrix}$$

and

$$C = \frac{1}{2}\left(A - A^T\right) = \frac{1}{2}\begin{bmatrix} 2 & 2 & -6 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 2 & 0 & 0 \\ 2 & 1 & 0 \\ -6 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & -3 \\ -1 & 0 & 2 \\ 3 & -2 & 0 \end{bmatrix}$$

we have

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

The following result lists the main algebraic properties of symmetric matrices.

- **Operations for symmetric matrices**

  If A and B are symmetric matrices with the same size, and if k is any scalar, then

  (1) A^T *is symmetric.*

  (2) $A + B$ *and* $A - B$ *are symmetric.*

  (3) kA *is symmetric.*

It is not true, in general, that the product of symmetric matrices is symmetric. To see why this is so, let A and B be symmetric matrices with the same size. Then it follows that

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

Thus, $(AB)^T = AB$ if and only if $AB = BA$, that is, if and only if A and B commute. In summary, we have the following result.

- **Product of symmetric matrices**

  The product of two symmetric matrices is symmetric if and only if the matrices commute.

We also note that if A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix, so the products AA^T and $A^T A$ are both square matrices. Such products are always symmetric since

$$\left(AA^T\right)^T = \left(A^T\right)^T A^T = AA^T \quad \text{and} \quad \left(A^T A\right)^T = A^T \left(A^T\right)^T = A^T A.$$

- **Trace of a matrix**

  *If A is a square matrix, then the **trace** of A, denoted by $tr(A)$, is defined to be the sum of the entries on the main diagonal of A. The trace of A is undefined if A is not a square matrix.*

For example, if

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

then $\operatorname{tr}(A) = (-1) + 5 + 7 + 0 = 11$.

We have the following properties of trace concept.

- **Properties of trace**

  For matrices A and B of size n, we have

  (1) $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.

  (2) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Example: Show that there are no square matrices A and B such that $AB - BA = I$.

Solution: Let the matrices A and B be square matrices of the same size, say n. Then

$$\operatorname{tr}(A - B) = \operatorname{tr}(A) - \operatorname{tr}(B) \quad \text{and} \quad \operatorname{tr}(AB) = \operatorname{tr}(BA).$$

Now, if $AB - BA = I_n$, then

$$n = \operatorname{tr}(I_n) = \operatorname{tr}(AB - BA) = \operatorname{tr}(AB) - \operatorname{tr}(BA) = 0,$$

which is impossible. Hence, no square matrices for which $AB - BA = I$ is satisfy exist.■

1.5. Inverses of Matrices

In real arithmetic every nonzero number a has a reciprocal $a^{-1} = \frac{1}{a}$ with the property $aa^{-1} = 1$. The number a^{-1} is sometimes called the **multiplicative inverse** of a. Our objective in this section is to develop an analog of this result for matrix arithmetic.

- **Invertible matrix**

  *If A is a square matrix, and if a matrix B of the same size can be found such that $AB = BA = I$, then A is said to be **invertible** and B is called an **inverse** of A. If no such matrix B can be found, then A is said to be **singular**.*

We note that the relationship $AB = BA = I$ is not changed by interchanging A and B, so if A is invertible and B is an inverse of A, then it is also true that B is invertible, and A is an inverse of B. Thus, when $AB = BA = I$, we say that A and B are inverses of one another.

For example, the matrix $B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ is an inverse for $A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ since

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

and

$$BA = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

On the other hand, the matrix

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

has no inverse. To see why, let

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

be any 3×3 matrix. The third column of BA is

$$\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

so BA cannot be equal to I_3.

We can immediately give two classes of invertible matrices.

Example: Any elementary matrix is invertible, and its inverse is also elementary. To see this, recall that the left multiplication by an elementary matrix is the same as performing the corresponding elementary row operation. Thus

· If E is obtained from I by interchanging two rows, then $EE = I$, so that E works as its own inverse.

· If E is obtained from I by multiplying a row by a nonzero constant, and if F is obtained from I by dividing a row by that constant, then $EF = I$, so F is an inverse of E.

· If E is obtained from I by adding a multiple of a row to another row, and if F is obtained from I by subtracting that multiple of this row to that row, then $EF = I$, so F is an inverse of E.■

Example: A diagonal matrix with nonzero diagonal entries is invertible. To see this suppose that

$$D = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & d_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & d_n \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & a_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & a_n \end{bmatrix}.$$

Then it is obvious that

$$DA = AD = \begin{bmatrix} d_1a_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & d_2a_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & d_na_n \end{bmatrix}.$$

Therefore, if $d_k \neq 0$ for $k = 1, 2, \ldots, n$, then letting $a_k = \frac{1}{d_k}$, we see that $DA = AD = I$.■

Here are some of the basic laws of inverse calculations.

- **Laws of matrix inverses**

  Let A, B, C be matrices of the appropriate sizes so that the following multiplications make sense, I be a suitably sized identity matrix, and c be a nonzero scalar. Then

  (1) *If the matrix A is invertible, then it has only one inverse, which is denoted by A^{-1} (uniqueness).*

  (2) *If the matrix A is invertible, then $(A^{-1})^{-1} = A$ (double inverse).*

  (3) *If any two of the three matrices A, B, and AB are invertible, then so is the third, and moreover, $(AB)^{-1} = B^{-1}A^{-1}$ (2/3 rule).*

(4) *If the matrix A is invertible, then $(cA)^{-1} = \frac{1}{c}A^{-1}$.*

(5) *If the matrix A is invertible, then $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$ (transpose/inverse).*

(6) *If the matrix A is invertible and if $AB = AC$ or $BA = CA$, then $B = C$ (cancelation).*

Proof:

(1) Suppose that both B and C work as inverses to the matrix A. Then

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

Hence, we shall write A^{-1} for the unique inverse of the square matrix A provided that there is an inverse.

(2) Since $AA^{-1} = A^{-1}A = I$, A^{-1} is invertible and $\left(A^{-1}\right)^{-1} = A$.

(3) Suppose that A and B are both invertible and of same size. Then

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

and

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

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

Suppose now that A and AB are both invertible. Then

$$\begin{aligned} BB^{-1} &= A^{-1}A\left(BB^{-1}\right)A^{-1}A = A^{-1}(AB)\left(B^{-1}A^{-1}\right)A = A^{-1}(AB)(AB)^{-1}A \\ &= A^{-1}A = I, \end{aligned}$$

so B is invertible. The remaining case can be proved in the same way.

(4) If A is invertible, then

$$(cA)\left(\frac{1}{c}A^{-1}\right) = c\frac{1}{c}AA^{-1} = AA^{-1} = I,$$

and

$$\left(\frac{1}{c}A^{-1}\right)(cA) = \frac{1}{c}cA^{-1}A = A^{-1}A = I,$$

so $(cA)^{-1} = \frac{1}{c}A^{-1}$.

(5) If A is invertible, then

$$\left(A^{-1}\right)^T A^T = \left(AA^{-1}\right)^T = I^T = I,$$

and

$$A^T\left(A^{-1}\right)^T = \left(A^{-1}A\right)^T = I^T = I,$$

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

(6) If A is invertible and $AB = AC$, then

$$AB = AC \Rightarrow A^{-1}AB = A^{-1}AC \Rightarrow B = C,$$

and if $BA = CA$, then

$$BA = CA \Rightarrow BAA^{-1} = CAA^{-1} \Rightarrow B = C,$$

so cancelation holds.■

We can introduce integer powers of a square matrix by inverse notation.

- **Powers of a matrix**

  If A is a square matrix, then we define the nonnegative integer powers of A to be

$$A^0 = I, \quad A^k = \underbrace{AA\cdots A}_{k\text{-terms}}.$$

  Moreover, if A is invertible, then we define the negative integer powers to be

$$A^{-k} = \left(A^{-1}\right)^k = \underbrace{A^{-1}A^{-1}\cdots A^{-1}}_{k\text{-terms}}.$$

We can use the various laws of arithmetic and inverse to carry out an inverse calculation.

Example: Let

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

Show that $(I - A)^3 = \mathbf{0}$ and use this to find A^{-1}.

Solution: We have

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

and

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

Next we use the laws of matrix arithmetic to obtain

$$\mathbf{0} = (I - A)^3 = (I - A)\left(I^2 - 2A + A^2\right) = I - 3A + 3A^2 - A^3$$

which implies that

$$I = 3A - 3A^2 + A^3 = A\left(3I - 3A + A^2\right),$$

so $A^{-1} = 3I - 3A + A^2$. Thus

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

the inverse of A.■

The next result establishes some fundamental relationships among invertibility, homogeneous linear systems, reduced row echelon forms, and elementary matrices.

- **Equivalent statements I**

  If A is an $n \times n$ matrix, then the following statements are equivalent:

  (1) *A is invertible.*

  (2) *$Ax = \mathbf{0}$ has only the trivial solution.*

  (3) *The reduced row echelon form of A is I_n.*

  (4) *A is expressible as a product of elementary matrices.*

Proof:

(1)⇒(2) Let A be invertible and x_0 be any solution of $Ax = \mathbf{0}$. Then

$$Ax_0 = \mathbf{0} \Rightarrow A^{-1}Ax_0 = \mathbf{0} \Rightarrow x_0 = \mathbf{0}.$$

(2)⇒(3) Let $Ax = \mathbf{0}$ be the matrix form of the system

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \cdots & + & a_{nn}x_n & = & 0 \end{array}$$

and assume the system has only the trivial solution. If we solve the system by Gauss-Jordan elimination, then the system of equations corresponding to the reduced row echelon form of the augmented matrix will be

$$\begin{array}{ccccccc} x_1 & & & & & = & 0 \\ & & x_2 & & & = & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & x_n & = & 0 \end{array}$$

so the augmented matrix is

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

If we disregard the last column of zeros, we conclude that the reduced row echelon form of A is I_n.

(3)⇒(4) Assume that the reduced row echelon form of A is I_n, so that A can be reduced to I_n by a finite sequence of elementary row operations. Each of these operations can be accomplished by multiplying on the left by an appropriate elementary matrix. Thus we can find elementary matrices $E_1, E_2, \ldots, E_k$ such that

$$E_1 E_2 \cdots E_k A = I_n$$

which implies that

$$A = E_k^{-1} E_{k-1}^{-1} \cdots E_2^{-1} E_1^{-1} I_n.$$

Since inverses of elementary matrices are elementary, this equation expresses A as a product of elementary matrices.

(4)⇒(1) If A is a product of elementary matrices, then the matrix A is a product of invertible matrices and hence is invertible since inverses of elementary matrices are invertible.■

As an application of above result, we can develop a procedure that can be used to tell whether a given matrix is invertible, and if so, produce its inverse. To derive this procedure, assume for the moment, that A is an invertible $n \times n$ matrix. If we multiply both sides of the equation

$$E_1 E_2 \cdots E_k A = I_n$$

on the right by A^{-1} and simplify, we obtain

$$A^{-1} = E_1 E_2 \cdots E_k I_n,$$

which tells us that the same sequence of row operations that reduces A to I_n will transform I_n to A^{-1}.

- **Inverse procedure**

  Given an $n \times n$ matrix A, to compute A^{-1}:

  (1) *Form the superaugmented matrix $A = [A|I_n]$.*

  (2) *Reduce the first n columns of A to reduced row echelon form by performing elementary operations on the matrix A resulting in the matrix $[R|B]$.*

  (3) *If $R = I_n$, then set $A^{-1} = B$, otherwise, A is singular and A^{-1} does not exist.*

Example: Find the inverse of

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

Solution: We have

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 1 & 0 & 8 & 0 & 0 & 1 \end{array}\right] \underset{III-I}{\overset{II-2I}{\longrightarrow}} \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & -2 & 5 & -1 & 0 & 1 \end{array}\right] \underset{III+2II}{\overset{I-2II}{\longrightarrow}} \left[\begin{array}{ccc|ccc} 1 & 0 & 9 & 5 & -2 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & -1 & -5 & 2 & 1 \end{array}\right]$$

$$\overset{(-1)III}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 9 & 5 & -2 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right] \underset{II+3III}{\overset{I-9III}{\longrightarrow}} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -40 & 16 & 9 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right]$$

Thus

$$A^{-1} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}$$

is the inverse of A.■

Example: Consider the matrix

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

We have

$$\left[\begin{array}{ccc|ccc} 1 & 6 & 4 & 1 & 0 & 0 \\ 2 & 4 & -1 & 0 & 1 & 0 \\ -1 & 2 & 5 & 0 & 0 & 1 \end{array}\right] \underset{III+I}{\overset{II-2I}{\longrightarrow}} \left[\begin{array}{ccc|ccc} 1 & 6 & 4 & 1 & 0 & 0 \\ 0 & -8 & -9 & -2 & 1 & 0 \\ 0 & 8 & 9 & 1 & 0 & 1 \end{array}\right] \overset{III+II}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 6 & 4 & 1 & 0 & 0 \\ 0 & -8 & -9 & -2 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 \end{array}\right]$$

Since we have obtained a row of zeros on the left side, A is not invertible.■

We know from earlier work that invertible matrix factors produce an invertible product. Conversely, the following result shows that if the product of square matrices is invertible, then the factors themselves must be invertible.

- **Matrix product and invertibility**

  Let A and B be square matrices of the same size. If AB is invertible, then A and B must also be invertible.

Proof: We will show first that B is invertible by showing that the homogeneous system $Bx = \mathbf{0}$ has only the trivial solution. If we assume that x_0 is any solution of this system, then $(AB)\,x_0 = A\,(Bx_0) = A\mathbf{0} = \mathbf{0}$, so $x_0 = \mathbf{0}$. But the invertibility of B implies the invertibility of B^{-1}, which in turn implies that

$$(AB)\,B^{-1} = A\left(BB^{-1}\right) = AI = A$$

is invertible since the left side is a product of invertible matrices. This completes the proof.■

In Section 1.1 we made the statement that every linear system either has no solution, or has exactly one solution, or has infinitely many solutions. We are now in a position to prove this fundamental result.

- **Solutions of linear systems**

  A system of linear equations has no solutions, or has exactly one solution, or infinitely many solutions.

Proof: For a system of linear equations exactly one of the following is true:

(a) the system has no solution.

(b) the system has exactly one solution.

(c) the system has more than one solution.

The proof will be complete if we can show that the system has infinitely many solutions in case (c).

Let $Ax = b$ be the matrix representation of the given system. Assume that $Ax = b$ has more than one solution, and let $x_0 = x_1 - x_2$, where x_1 and x_2 are any two distinct solutions. Because x_1 and x_2 are distinct, the matrix x_0 is nonzero, and moreover,

$$Ax_0 = A\,(x_1 - x_2) = Ax_1 - Ax_2 = b - b = \mathbf{0}.$$

If we now let k be any scalar, then

$$A\left(x_1 + kx_0\right) = Ax_1 + kAx_0 = b + k\mathbf{0} = b,$$

which says that $x_1 + kx_0$ is a solution of $Ax = b$. Since x_0 is nonzero and there are infinitely many choices for k, the system $Ax = b$ has infinitely many solutions.■

Thus far we have studied two procedures for solving linear systems, namely, Gauss–Jordan elimination and Gaussian elimination. The following result provides an actual formula for the solution of a linear system of n equations in n unknowns in the case where the coefficient matrix is invertible.

- **Consistency of linear systems**

  If A is an invertible $n \times n$ matrix, then for each $n \times 1$ matrix b, the system of equations $Ax = b$ has exactly one solution, namely, $x = A^{-1}b$.

Proof: Since $A\left(A^{-1}b\right) = b$, it follows that $x = A^{-1}b$ is a solution of $Ax = b$. To show that this is the only solution, we will assume that x_0 is an arbitrary solution and then show that x_0 must be the solution $A^{-1}b$. If x_0 is any solution of $Ax = b$, then $Ax_0 = b$. Multiplying both sides of this equation by A^{-1}, we obtain $x_0 = A^{-1}b$.■

Example: Consider the system of equations

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 3x_3 & = & 5 \\ 2x_1 & + & 5x_2 & + & 3x_3 & = & 3 \\ x_1 & & & + & 8x_3 & = & 17 \end{array}$$

In the matrix form this system can be written as $Ax = b$, where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad b = \begin{bmatrix} 5 \\ 3 \\ 17 \end{bmatrix}$$

Since

$$A^{-1} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}$$

the solution of the system is

$$x = A^{-1}b = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 17 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix},$$

or $x_1 = 1, x_2 = -1, x_3 = 2.\blacksquare$

We are now in a position to add two more statements to the given four.

- **Equivalent statements II**

  If A is an $n \times n$ matrix, then the following statements are equivalent:

  (1) *A is invertible.*

  (2) *$Ax = \mathbf{0}$ has only the trivial solution.*

  (3) *The reduced row echelon form of A is I_n.*

  (4) *A is expressible as a product of elementary matrices.*

  (5) *$Ax = b$ is consistent for every $n \times 1$ matrix b.*

  (6) *$Ax = b$ has exactly one solution for every $n \times 1$ matrix b.*

1.6. Applications in Cryptography

We conclude this chapter by noting that matrix multiplication and inverse matrices can be used in cryptography. The study of encoding and decoding secret messages is called **cryptography**. Although secret codes date to the earliest days of written communication, there has been a recent surge of interest in the subject because of the need to maintain the privacy of information transmitted over public lines of communication. In the language of cryptography, codes are called ciphers, uncoded messages are called plaintext, and coded messages are called ciphertext. The process of converting from plaintext to ciphertext is called enciphering, and the reverse process of converting from ciphertext to plaintext is called deciphering. Matrix multiplication can be used to code messages and uncode them.

Example: *Using the table*

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**
10	23	2	0	13	1	22	21	11	12	6	26	18	9
O	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**	**Space**	
16	24	7	20	15	4	25	3	8	19	5	17	14	

and the matrix

$$\begin{bmatrix} 1 & -2 & 3 \\ 2 & -5 & 10 \\ -1 & 2 & -2 \end{bmatrix}$$

(a) *encrypt* **THE BROWN FOX IS QUICK**

(b) *decrypt* **ACDQEVJENXMH**

Solution:

(a) We obtain the 3×8 plaintext matrix $\mathbb{K}$ as

$$\mathbb{K} = \begin{bmatrix} \mathrm{T} & & \mathrm{O} & & \mathrm{X} & \mathrm{S} & \mathrm{U} & \mathrm{K} \\ \mathrm{H} & \mathrm{B} & \mathrm{W} & \mathrm{F} & & & \mathrm{I} & \mathrm{X} \\ \mathrm{E} & \mathrm{R} & \mathrm{N} & \mathrm{O} & \mathrm{I} & \mathrm{Q} & \mathrm{C} & \mathrm{X} \end{bmatrix} \to \mathbb{K} = \begin{bmatrix} 4 & 14 & 16 & 14 & 19 & 15 & 25 & 6 \\ 21 & 23 & 8 & 1 & 14 & 14 & 11 & 19 \\ 13 & 20 & 9 & 16 & 11 & 7 & 2 & 19 \end{bmatrix}$$

where we have used the dummy letter X to complete the matrix. Now, we have

$$\begin{aligned} \mathbb{A}\mathbb{K} &= \begin{bmatrix} 1 & -2 & 3 \\ 2 & -5 & 10 \\ -1 & 2 & -2 \end{bmatrix} \begin{bmatrix} 4 & 14 & 16 & 14 & 19 & 15 & 25 & 6 \\ 21 & 23 & 8 & 1 & 14 & 14 & 11 & 19 \\ 13 & 20 & 9 & 16 & 11 & 7 & 2 & 19 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 28 & 27 & 60 & 24 & 8 & 9 & 25 \\ 33 & 113 & 82 & 183 & 78 & 30 & 15 & 107 \\ 12 & -8 & -18 & -44 & -13 & -1 & -7 & -6 \end{bmatrix} \\ &\equiv \begin{bmatrix} 1 & 1 & 0 & 6 & 24 & 8 & 9 & 25 \\ 6 & 5 & 1 & 21 & 24 & 3 & 15 & 26 \\ 12 & 19 & 9 & 10 & 14 & 26 & 20 & 21 \end{bmatrix} (\mathrm{mod} 27) \end{aligned}$$

Thus the ciphertext is

$$\begin{bmatrix} \mathrm{F} & \mathrm{F} & \mathrm{D} & \mathrm{K} & \mathrm{P} & \mathrm{W} & \mathrm{N} & \mathrm{U} \\ \mathrm{K} & \mathrm{Y} & \mathrm{F} & \mathrm{H} & \mathrm{P} & \mathrm{V} & \mathrm{S} & \mathrm{L} \\ \mathrm{J} & \mathrm{X} & \mathrm{N} & \mathrm{A} & & \mathrm{L} & \mathrm{R} & \mathrm{H} \end{bmatrix} \Rightarrow \textbf{FKJFYXDFNKHAPP WVLNSRULH}$$

(b) We set the 3×4 ciphertext matrix $\mathbb{B}$ as

$$\mathbb{B} = \begin{bmatrix} \mathrm{A} & \mathrm{Q} & \mathrm{J} & \mathrm{X} \\ \mathrm{C} & \mathrm{E} & \mathrm{E} & \mathrm{M} \\ \mathrm{D} & \mathrm{V} & \mathrm{N} & \mathrm{H} \end{bmatrix} \to \begin{bmatrix} 10 & 7 & 12 & 19 \\ 2 & 13 & 13 & 18 \\ 0 & 3 & 9 & 21 \end{bmatrix}$$

Then the plaintext is

$$\begin{aligned}
\mathbb{K} &= \mathbb{A}^{-1}\mathbb{B} = \begin{bmatrix} 10 & -2 & 5 \\ 6 & -1 & 4 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 10 & 7 & 12 & 19 \\ 2 & 13 & 13 & 18 \\ 0 & 3 & 9 & 21 \end{bmatrix} = \begin{bmatrix} 96 & 59 & 139 & 259 \\ 58 & 41 & 95 & 180 \\ 10 & 10 & 21 & 40 \end{bmatrix} \\
&\equiv \begin{bmatrix} 15 & 5 & 4 & 16 \\ 4 & 14 & 14 & 18 \\ 10 & 21 & 21 & 13 \end{bmatrix} (\mathrm{mod} 27) \to \begin{bmatrix} \mathrm{S} & \mathrm{Y} & \mathrm{T} & \mathrm{O} \\ \mathrm{T} & & & \mathrm{M} \\ \mathrm{A} & \mathrm{A} & \mathrm{H} & \mathrm{E} \end{bmatrix}
\end{aligned}$$

that is, **STAY AT HOME**■