CHAPTER 3

Linear Equations and Matrices

In this chapter we introduce matrices via the theory of simultaneous linear equations. This method has the advantage of leading in a natural way to the concept of the reduced row-echelon form of a matrix. In addition, we will formulate some of the basic results dealing with the existence and uniqueness of systems of linear equations. In Chapter 5 we will arrive at the same matrix algebra from the viewpoint of linear transformations.

3.1 SYSTEMS OF LINEAR EQUATIONS

Let $a_1, \ldots, a_n, y$ be elements of a field $\mathcal{F}$, and let $x_1, \ldots, x_n$ be **unknowns** (also called **variables** or **indeterminates**). Then an equation of the form

$$a_1 x_1 + \cdots + a_n x_n = y$$

is called a **linear equation in n unknowns** (over $\mathcal{F}$). The scalars a_i are called the **coefficients** of the unknowns, and y is called the **constant term** of the equation. A vector $(c_1, \ldots, c_n) \in \mathcal{F}^n$ is called a **solution vector** of this equation if and only if

$$a_1 c_1 + \cdots + a_n c_n = y$$

in which case we say that $(c_1, \ldots, c_n)$ **satisfies** the equation. The set of all such solutions is called the **solution set** (or the **general solution**).

Now consider the following **system of m linear equations in n unknowns**:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= y_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= y_2 \\ \vdots \qquad\quad & \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= y_m \end{aligned}$$

We abbreviate this system by

$$\sum_{j=1}^{n} a_{ij}x_j = y_i, \qquad i = 1, \ldots, m \ .$$

If we let S_i denote the solution set of the equation $\sum_j a_{ij}x_j = y_i$ for each i, then the solution set S of the system is given by the intersection $S = \cap S_i$. In other words, if $(c_1, \ldots, c_n) \in \mathcal{F}^n$ is a solution of the system of equations, then it is a solution of each of the m equations in the system.

Example 3.1 Consider this system of two equations in three unknowns over the real field $\mathbb{R}$:

$$\begin{aligned} 2x_1 - 3x_2 + x_3 &= 6 \\ x_1 + 5x_2 - 2x_3 &= 12 \end{aligned}$$

The vector $(3, 1, 3) \in \mathbb{R}^3$ is not a solution of this system because

$$2(3) - 3(1) + 3 = 6$$

while

$$3 + 5(1) - 2(3) = 2 \neq 12 \ .$$

However, the vector $(5, 1, -1) \in \mathbb{R}^3$ is a solution since

$$2(5) - 3(1) + (-1) = 6$$

and

$$5 + 5(1) - 2(-1) = 12 \ . \ /\!\!/$$

Associated with a system of linear equations are two rectangular arrays of elements of $\mathcal{F}$ that turn out to be of great theoretical as well as practical significance. For the system $\sum_j a_{ij}x_j = y_i$, we define the **matrix of coefficients** A as the array

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

and the **augmented matrix** as the array aug A given by

$$aug\ A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & y_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & y_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & y_n \end{pmatrix}.$$

In general, we will use the term **matrix** to denote any array such as the array A shown above. This matrix has m **rows** and n **columns**, and hence is referred to as an m x n matrix, or a matrix of **size** m x n. By convention, an element $a_{ij} \in \mathcal{F}$ of A is labeled with the first index referring to the row and the second index referring to the column. The scalar a_{ij} is usually called the i, *jth* **entry** (or **element**) of the matrix A. We will frequently denote the matrix A by the symbol (a_{ij}).

Another rather general way to define a matrix is as a mapping from a subset of all ordered pairs of positive integers into the field $\mathcal{F}$. In other words, we define the mapping A by $A(i, j) = a_{ij}$ for every $1 \le i \le m$ and $1 \le j \le n$. In this sense, a matrix is actually a mapping, and the m x n array written above is just a representation of this mapping.

Before proceeding with the general theory, let us give a specific example demonstrating how to solve a system of linear equations.

Example 3.2 Let us attempt to solve the following system of linear equations:

$$\begin{aligned} 2x_1 + \ x_2 - 2x_3 &= -3 \\ x_1 - 3x_2 + \ x_3 &= \ \ 8 \\ 4x_1 - \ x_2 - 2x_3 &= \ \ 3 \end{aligned}$$

That our approach is valid in general will be proved in our first theorem below.

Multiply the first equation by 1/2 to get the coefficient of x_1 equal to 1:

$$\begin{aligned} x_1 + (1/2)x_2 - x_3 &= -3/2 \\ x_1 - 3x_2 + x_3 &= 8 \\ 4x_1 - x_2 - 2x_3 &= 3 \end{aligned}$$

Multiply the first equation by −1 and add it to the second to obtain a new second equation, then multiply the first by −4 and add it to the third to obtain a new third equation:

$$\begin{aligned} x_1 + (1/2)x_2 - x_3 &= -3/2 \\ -(7/2)x_2 + 2x_3 &= 19/2 \\ -3x_2 - 2x_3 &= 9 \end{aligned}$$

Multiply the second by −2/7 to get the coefficient of x_2 equal to 1, then multiply this new second equation by 3 and add to the third:

$$\begin{aligned} x_1 + (1/2)x_2 - x_3 &= -3/2 \\ x_2 - (4/7)x_3 &= -19/7 \\ (2/7)x_3 &= 6/7 \end{aligned}$$

Multiply the third by 7/2, then add 4/7 times this new equation to the second:

$$\begin{aligned} x_1 + (1/2)x_2 - x_3 &= -3/2 \\ x_2 &= -1 \\ x_3 &= 3 \end{aligned}$$

Add the third equation to the first, then add −1/2 times the second equation to the new first to obtain

$$\begin{aligned} x_1 &= 2 \\ x_2 &= -1 \\ x_3 &= 3 \end{aligned}$$

This is now a solution of our system of equations. While this system could have been solved in a more direct manner, we wanted to illustrate the systematic approach that will be needed below. ∕∕

Two systems of linear equations are said to be **equivalent** if they have equal solution sets. That each successive system of equations in Example 3.2 is indeed equivalent to the previous system is guaranteed by the following theorem.

Theorem 3.1 The system of two equations in n unknowns over a field $\mathcal{F}$

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \end{aligned} \tag{1}$$

with $a_{11} \neq 0$ is equivalent to the system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a'_{22}x_2 + \cdots + a'_{2n}x_n &= b'_2 \end{aligned} \tag{2}$$

in which

$$a'_{2i} = a_{21}a_{1i} - a_{11}a_{2i}$$

for each $i = 1, \ldots, n$ and

$$b'_2 = a_{21}b_1 - a_{11}b_2 \;.$$

Proof Let us define

$$L_i = \sum_{j=1}^{n} a_{ij}x_j$$

so that (1) may be written as the system

$$\begin{aligned} L_1 &= b_1 \\ L_2 &= b_2 \end{aligned} \tag{1$'$}$$

while (2) is just

$$\begin{aligned} L_1 &= b_1 \\ a_{21}L_1 - a_{11}L_2 &= a_{21}b_1 - a_{11}b_2 \end{aligned} \tag{2$'$}$$

If $(x_1, \ldots, x_n) \in \mathcal{F}^n$ is a solution of (1′), then the two equations

$$\begin{aligned} a_{21}L_1 &= a_{21}b_1 \\ a_{11}L_2 &= a_{11}b_2 \end{aligned}$$

and hence also

$$a_{21}L_1 - a_{11}L_2 = a_{21}b_1 - a_{11}b_2$$

are all true equations. Therefore every solution of (1′) also satisfies (2′).

Conversely, suppose that we have a solution $(x_1, \ldots, x_n)$ to the system (2′). Then clearly

$$a_{21}L_1 = a_{21}b_1$$

is a true equation. Hence, subtracting the second of (2′) from this gives us

$$a_{21}L_1 - (a_{21}L_1 - a_{11}L_2) = a_{21}b_1 - (a_{21}b_1 - a_{11}b_2)$$

or

$$a_{11}L_2 = a_{11}b_2 \ .$$

Thus $L_2 = b_2$ is also a true equation. This shows that any solution of (2′) is a solution of (1′) also. ∎

We point out that in the proof of Theorem 3.1 (as well as in Example 3.2), it was only the coefficients themselves that were of any direct use to us. The unknowns x_i were never actually used in any of the manipulations. This is the reason that we defined the matrix of coefficients (a_{ij}). What we now proceed to do is to generalize the above method of solving systems of equations in a manner that utilizes this matrix explicitly.

Exercises

1. For each of the following systems of equations, find a solution if it exists:

(a)
$$\begin{aligned} x + 2y - 3z &= -1 \\ 3x - y + 2z &= 7 \\ 5x + 3y - 4z &= 2 \end{aligned}$$

(b)
$$\begin{aligned} 2x + y - 2z &= 10 \\ 3x + 2y + 2z &= 1 \\ 5x + 4y + 3z &= 4 \end{aligned}$$

(c)
$$\begin{aligned} x + 2y - 3z &= 6 \\ 2x - y + 4z &= 2 \\ 4x + 3y - 2z &= 14 \end{aligned}$$

2. Determine whether or not the each of the following two systems is equivalent (over the complex field):

(a)
$$\begin{aligned} x - y &= 0 \\ 2x + y &= 0 \end{aligned} \qquad \text{and} \qquad \begin{aligned} 3x + y &= 0 \\ x + y &= 0 \end{aligned}$$

(b)
$$\begin{aligned} -x + y + 4z &= 0 \\ x + 3y + 8z &= 0 \\ (1/2)x + y + (5/2)z &= 0 \end{aligned} \qquad \text{and} \qquad \begin{aligned} x - z &= 0 \\ y + 3z &= 0 \end{aligned}$$

(c)
$$\begin{aligned} 2x + (-1+i)y + t &= 0 \\ 3y - 2iz + 5t &= 0 \end{aligned}$$

and

$$\begin{aligned}(1+i/2)x + \quad 8y - iz - \ t &= 0\\ (2/3)x - (1/2)y + z + 7t &= 0\end{aligned}$$

3.2 ELEMENTARY ROW OPERATIONS

The important point to realize in Example 3.2 is that we solved a system of linear equations by performing some combination of the following operations:

(a) Change the order in which the equations are written.
(b) Multiply each term in an equation by a nonzero scalar.
(c) Multiply one equation by a nonzero scalar and then add this new equation to another equation in the system.

Note that (a) was not used in Example 3.2, but it would have been necessary if the coefficient of x_1 in the first equation had been 0. The reason for this is that we want the equations put into echelon form as defined below.

We now see how to use the matrix aug A as a tool in solving a system of linear equations. In particular, we define the following so-called **elementary row operations** (or **transformations**) as applied to the augmented matrix:

(α) Interchange two rows.
(β) Multiply one row by a nonzero scalar.
(γ) Add a scalar multiple of one row to another.

It should be clear that operations (α) and (β) have no effect on the solution set of the system and, in view of Theorem 3.1, that operation (γ) also has no effect.

The next two examples show what happens both in the case where there is no solution to a system of linear equations, and in the case of an infinite number of solutions. In performing these operations on a matrix, we will let R_i denote the *ith* row. We leave it to the reader to repeat Example 3.2 using this notation.

Example 3.3 Consider this system of linear equations over the field $\mathbb{R}$:

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

The augmented matrix is

$$\begin{pmatrix} 1 & 3 & 2 & 7 \\ 2 & 1 & -1 & 5 \\ -1 & 2 & 3 & 4 \end{pmatrix}$$

and the reduction proceeds as follows. We first perform the following elementary row operations:

$$\begin{matrix} \\ R_2 - 2R_1 \to \\ R_3 + R_1 \to \end{matrix} \begin{pmatrix} 1 & 3 & 2 & 7 \\ 0 & -5 & -5 & -9 \\ 0 & 5 & 5 & 11 \end{pmatrix}$$

Now, using this matrix, we obtain

$$\begin{matrix} \\ -R_2 \to \\ R_3 + R_2 \to \end{matrix} \begin{pmatrix} 1 & 3 & 2 & 7 \\ 0 & 5 & 5 & 9 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

It is clear that the equation 0z = 2 has no solution, and hence this system has no solution. ∕∕

Example 3.4 Let us solve the following system over the field ℝ:

$$\begin{aligned} x_1 - 2x_2 + 2x_3 - x_4 &= -14 \\ 3x_1 + 2x_2 - x_3 + 2x_4 &= 17 \\ 2x_1 + 3x_2 - x_3 - x_4 &= 18 \\ -2x_1 + 5x_2 - 3x_3 - 3x_4 &= 26 \end{aligned}$$

We have the matrix aug A given by

$$\begin{pmatrix} 1 & -2 & 2 & -1 & -14 \\ 3 & 2 & -1 & 2 & 17 \\ 2 & 3 & -1 & -1 & 18 \\ -2 & 5 & -3 & -3 & 26 \end{pmatrix}$$

and hence we obtain the sequence

$$\begin{matrix} \\ R_2 - 3R_1 \to \\ R_3 - 2R_1 \to \\ R_4 + 2R_1 \to \end{matrix} \begin{pmatrix} 1 & -2 & 2 & -1 & -14 \\ 0 & 8 & -7 & 5 & 59 \\ 0 & 7 & -5 & 1 & 46 \\ 0 & 1 & 1 & -5 & -2 \end{pmatrix}$$

$$\begin{array}{r} \\ R_4 \to \\ R_2 - 8R_4 \to \\ R_3 - 7R_4 \to \end{array}\begin{pmatrix} 1 & -2 & 2 & -1 & -14 \\ 0 & 1 & 1 & -5 & -2 \\ 0 & 0 & -15 & 45 & 75 \\ 0 & 0 & -12 & 36 & 60 \end{pmatrix}$$

$$\begin{array}{r} \\ \\ (-1/15)R_3 \to \\ (-1/12)R_4 \to \end{array}\begin{pmatrix} 1 & -2 & 2 & -1 & -14 \\ 0 & 1 & 1 & -5 & -2 \\ 0 & 0 & 1 & -3 & -5 \\ 0 & 0 & 1 & -3 & -5 \end{pmatrix}$$

We see that the third and fourth equations are identical, and hence we have three equations in four unknowns:

$$\begin{aligned} x_1 - 2x_2 + 2x_3 - x_4 &= -14 \\ x_2 + x_3 - 5x_4 &= -2 \\ x_3 - 3x_4 &= -5 \end{aligned}$$

It is now apparent that there are an infinite number of solutions because, if we let $c \in \mathbb{R}$ be any real number, then our solution set is given by $x_4 = c$, $x_3 = 3c - 5$, $x_2 = 2c + 3$ and $x_1 = -c + 2$. //

Two m x n matrices are said to be **row equivalent** if one can be transformed into the other by a finite number of elementary row operations. We leave it to the reader to show that this defines an equivalence relation on the set of all m x n matrices (see Exercise 3.2.1).

Our next theorem is nothing more than a formalization of an earlier remark.

Theorem 3.2 Let A and B be the augmented matrices of two systems of m linear equations in n unknowns. If A is row equivalent to B, then both systems have the same solution set.

Proof If A is row equivalent to B, then we can go from the system represented by A to the system represented by B by a succession of the operations (a), (b) and (c) described above. It is clear that operations (a) and (b) have no effect on the solutions, and the method of Theorem 3.1 shows that operation (c) also has no effect. ∎

A matrix is said to be in **row-echelon form** if successive rows of the matrix start out (from the left) with more and more zeros. In particular, a matrix is said to be in **reduced row-echelon** form if it has the following properties (which are more difficult to state precisely than they are to understand):

(1) All zero rows (if any) occur below all nonzero rows.
(2) The first nonzero entry (reading from the left) in each row is equal to 1.
(3) If the first nonzero entry in the i*th* row is in the j_i*th* column, then every other entry in the j_i*th* column is 0.
(4) If the first nonzero entry in the i*th* row is in the j_i*th* column, then $j_1 < j_2 < \cdots$.

We will call the first (or **leading**) nonzero entries in each row of a row-echelon matrix the **distinguished elements** of the matrix. Thus, a matrix is in reduced row-echelon form if the distinguished elements are each equal to 1, and they are the only nonzero entries in their respective columns.

Example 3.5 The matrix

$$\begin{pmatrix} 1 & 2 & -3 & 0 & 1 \\ 0 & 0 & 5 & 2 & -4 \\ 0 & 0 & 0 & 7 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

is in row-echelon form but not in reduced row-echelon form. However, the matrix

$$\begin{pmatrix} 1 & 0 & 5 & 0 & 2 \\ 0 & 1 & 2 & 0 & 4 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

is in reduced row-echelon form. Note that the distinguished elements of the first matrix are the numbers 1, 5 and 7, and the distinguished elements of the second matrix are the numbers 1, 1 and 1. ∕∕

The algorithm detailed in the proof of our next theorem introduces a technique generally known as **Gaussian elimination**.

Theorem 3.3 Every m x n matrix A is row equivalent to a reduced row-echelon matrix.

Proof This is essentially obvious from Example 3.4. The detailed description which follows is an algorithm for determining the reduced row-echelon form of a matrix.

Suppose that we first put A into the form where the leading entry in each nonzero row is equal to 1, and where every other entry in the column containing this first nonzero entry is equal to 0. (This is called simply the **row-reduced** form of A.) If this can be done, then all that remains is to perform a finite number of row interchanges to achieve the final desired reduced row-echelon form.

To obtain the row-reduced form we proceed as follows. First consider row 1. If every entry in row 1 is equal to 0, then we do nothing with this row. If row 1 is nonzero, then let j_1 be the smallest positive integer for which $a_{1j_1} \neq 0$ and multiply row 1 by $(a_{1j_1})^{-1}$. Next, for each $i \neq 1$ we add $-a_{ij_1}$ times row 1 to row i. This leaves us with the leading entry a_{1j_1} of row 1 equal to 1, and every other entry in the j_1*th* column equal to 0.

Now consider row 2 of the matrix we are left with. Again, if row 2 is equal to 0 there is nothing to do. If row 2 is nonzero, assume that the first nonzero entry occurs in column j_2 (where $j_2 \neq j_1$ by the results of the previous paragraph). Multiply row 2 by $(a_{2j_2})^{-1}$ so that the leading entry in row 2 is equal to 1, and then add $-a_{ij_2}$ times row 2 to row i for each $i \neq 2$. Note that these operations have no effect on either column j_1, or on columns $1, \dots, j_1$ of row 1.

It should now be clear that we can continue this process a finite number of times to achieve the final row-reduced form. We leave it to the reader to take an arbitrary matrix (a_{ij}) and apply successive elementary row transformations to achieve the desired final form. ∎

While we have shown that every matrix is row equivalent to at least one reduced row-echelon matrix, it is not obvious that this equivalence is unique. However, we shall show in the next section that this reduced row-echelon matrix is in fact unique. Because of this, the reduced row-echelon form of a matrix is often called the **row canonical form**.

Exercises

1. Show that row equivalence defines an equivalence relation on the set of all matrices.

2. For each of the following matrices, first reduce to row-echelon form, and then to row canonical form:

$$(a)\ \begin{pmatrix} 1 & -2 & 3 & -1 \\ 2 & -1 & 2 & 2 \\ 3 & 1 & 2 & 3 \end{pmatrix} \qquad (b)\ \begin{pmatrix} 1 & 2 & -1 & 2 & 1 \\ 2 & 4 & 1 & -2 & 3 \\ 3 & 6 & 2 & -6 & 5 \end{pmatrix}$$

$$(c)\ \begin{pmatrix} 1 & 3 & -1 & 2 \\ 0 & 1 & -5 & 3 \\ 2 & -5 & 3 & 1 \\ 4 & 1 & 1 & 5 \end{pmatrix}$$

3. For each of the following systems, find a solution or show that no solution exists:

$$(a)\ \begin{aligned} x + y + z &= 1 \\ 2x - 3y + 7z &= 0 \\ 3x - 2y + 8z &= 4 \end{aligned} \qquad (b)\ \begin{aligned} x - y + 2z &= 1 \\ x + y + z &= 2 \\ 2x - y + z &= 5 \end{aligned}$$

$$(c)\ \begin{aligned} x - y + 2z &= 4 \\ 3x + y + 4z &= 6 \\ x + y + z &= 1 \end{aligned} \qquad (d)\ \begin{aligned} x + 3y + z &= 2 \\ 2x + 7y + 4z &= 6 \\ x + y - 4z &= 1 \end{aligned}$$

$$(e)\ \begin{aligned} x + 3y + z &= 0 \\ 2x + 7y + 4z &= 0 \\ x + y - 4z &= 0 \end{aligned} \qquad (f)\ \begin{aligned} 2x - y + 5z &= 19 \\ x + 5y - 3z &= 4 \\ 3x + 2y + 4z &= 5 \end{aligned}$$

$$(g)\ \begin{aligned} 2x - y + 5z &= 19 \\ x + 5y - 3z &= 4 \\ 3x + 2y + 4z &= 25 \end{aligned}$$

4. Let f_1, f_2 and f_3 be elements of $F[\mathbb{R}]$ (i.e., the space of all real-valued functions defined on $\mathbb{R}$).
(a) Given a set $\{x_1, x_2, x_3\}$ of real numbers, define the 3 x 3 matrix $F(x) = (f_i(x_j))$ where the rows are labelled by i and the columns are labelled by j. Prove that the set $\{f_i\}$ is linearly independent if the rows of the matrix $F(x)$ are linearly independent.
(b) Now assume that each f_i has first and second derivatives defined on some interval $(a, b) \subset \mathbb{R}$, and let $f_i^{(j)}$ denote the jth derivative of f_i (where $f_i^{(0)}$ is just f_i). Define the matrix $W(x) = (f_i^{(j-1)}(x))$ where $1 \le i, j \le 3$. Prove that $\{f_i\}$ is linearly independent if the rows of $W(x)$ are independent

for some $x \in (a, b)$. (The determinant of $W(x)$ is called the **Wronskian** of the set of functions $\{f_i\}$.)

Show that each of the following sets of functions is linearly independent:
(c) $f_1(x) = -x^2 + x + 1$, $f_2(x) = x^2 + 2x$, $f_3(x) = x^2 - 1$.
(d) $f_1(x) = \exp(-x)$, $f_2(x) = x$, $f_3(x) = \exp(2x)$.
(e) $f_1(x) = \exp(x)$, $f_2(x) = \sin x$, $f_3(x) = \cos x$.

5. Let

$$A = \begin{pmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{pmatrix} .$$

Determine the values of $Y = (y_1, y_2, y_3)$ for which the system $\sum_i a_{ij} x_j = y_i$ has a solution.

6. Repeat the previous problem with the matrix

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

3.3 ROW AND COLUMN SPACES

We will find it extremely useful to consider the rows and columns of an arbitrary $m \times n$ matrix as vectors in their own right. In particular, the rows of A are to be viewed as vector n-tuples $A_1, \ldots, A_m$ where each $A_i = (a_{i1}, \ldots, a_{in}) \in \mathcal{F}^n$. Similarly, the columns of A are to be viewed as vector m-tuples $A^1, \ldots, A^n$ where each $A^j = (a_{1j}, \ldots, a_{mj}) \in \mathcal{F}^m$. For notational clarity, we should write A^j as the column vector

$$\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

but it is typographically easier to write this horizontally whenever possible. Note that we label the row vectors of A by subscripts, and the columns of A by superscripts.

Since each row A_i is an element of $\mathcal{F}^n$, the set of all rows of a matrix can be used to generate a new vector space V over $\mathcal{F}$. In other words, V is the space spanned by the rows A_i, and hence any $v \in V$ may be written as

$$v = \sum_{i=1}^{m} c_i A_i$$

where each $c_i \in \mathcal{F}$. The space V (which is apparently a subspace of $\mathcal{F}^n$) is called the **row space** of A. The dimension of V is called the **row rank** of A, and will be denoted by rr(A). Since V is a subspace of $\mathcal{F}^n$ and dim $\mathcal{F}^n$ = n, it follows that rr(A) = dim V ≤ n. On the other hand, V is spanned by the m vectors A_i, so that we must have dim V ≤ m. It then follows that rr(A) ≤ min{m, n}.

In an exactly analogous manner, we define the column space W of a matrix A as that subspace of $\mathcal{F}^m$ spanned by the n column vectors A^j. Thus any $w \in W$ is given by

$$w = \sum_{j=1}^{n} b_j A^j$$

The **column rank** of A, denoted by cr(A), is given by cr(A) = dim W and, as above, we must have cr(A) ≤ min{m, n}.

An obvious question is whether a sequence of elementary row operations changes either the row space or the column space of a matrix. A moments thought should convince you that the row space should not change, but it may not be clear exactly what happens to the column space. These questions are answered in our next theorem. While the following proof appears to be rather long, it is actually quite simple to understand.

Theorem 3.4 Let A and $\tilde{A}$ be row equivalent m x n matrices. Then the row space of A is equal to the row space of $\tilde{A}$, and hence rr(A) = rr($\tilde{A}$). Furthermore, we also have cr(A) = cr($\tilde{A}$). (However, note that the column space of A is not necessarily the same as the column space of $\tilde{A}$.)

Proof Let V be the row space of A and $\tilde{V}$ the row space of $\tilde{A}$. Since A and $\tilde{A}$ are row equivalent, A may be obtained from $\tilde{A}$ by applying successive elementary row operations. But then each row of A is a linear combination of rows of $\tilde{A}$, and hence $V \subset \tilde{V}$. On the other hand, $\tilde{A}$ may be obtained from A in a similar manner so that $\tilde{V} \subset V$. Therefore $V = \tilde{V}$ and hence rr(A) = rr($\tilde{A}$).

Now let W be the column space of A and $\tilde{W}$ the column space of $\tilde{A}$. Under elementary row operations, it will not be true in general that $W = \tilde{W}$, but we will show it is still always true that dim W = dim $\tilde{W}$. Let us define the mapping $f\colon W \to \tilde{W}$ by

$$f\left(\sum_{i=1}^{n} c_i A^i\right) = \sum_{i=1}^{n} c_i \tilde{A}^i \;.$$

In other words, if we are given any linear combination of the columns of A, then we look at the same linear combination of columns of $\tilde{A}$. In order to show that this is well-defined, we must show that if $\sum a_i A^i = \sum b_i A^i$, then $f(\sum a_i A^i) = f(\sum b_i A^i)$. This equivalent to showing that if $\sum c_i A^i = 0$ then $f(\sum c_i A^i) = 0$ because if $\sum(a_i - b_i)A^i = 0$ and $f(\sum(a_i - b_i)A^i) = 0$, then

$$\begin{aligned} 0 &= f\left(\Sigma(a_i - b_i)A^i\right) = \Sigma(a_i - b_i)\tilde{A}^i = \Sigma a_i\tilde{A}^i - \Sigma b_i\tilde{A}^i \\ &= f\left(\Sigma a_i A^i\right) - f\left(\Sigma b_i A^i\right) \end{aligned}$$

and therefore

$$f(\Sigma a_i A^i) = f(\Sigma b_i A^i) \;.$$

Now note that

$$\begin{aligned} f\left(\Sigma\left(b_i A^i + c_i A^i\right)\right) &= f\left(\Sigma(b_i + c_i)A^i\right) = \Sigma(b_i + c_i)\tilde{A}^i \\ &= \Sigma b_i\tilde{A}^i + \Sigma c_i\tilde{A}^i = f\left(\Sigma b_i A^i\right) + f\left(\Sigma c_i A^i\right) \end{aligned}$$

and

$$f(k(\Sigma c_i A^i)) = f(\Sigma(kc_i)A^i) = \Sigma(kc_i)\,\tilde{A}^i = kf(\Sigma c_i A^i)$$

so that f is a vector space homomorphism. If we can show that W and $\tilde{W}$ are isomorphic, then we will have $cr(A) = \dim W = \dim \tilde{W} = cr(\tilde{A})$. Since f is clearly surjective, we need only show that Ker f = {0} for each of the three elementary row transformations.

In the calculations to follow, it must be remembered that

$$A_i = (a_{i1}, \dots, a_{in})$$

and

$$A^i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix} .$$

Since

$$\Sigma c_i A_i = (\Sigma c_i a_{i1}, \dots, \Sigma c_i a_{in})$$

we see that $\Sigma c_i A_i = 0$ if and only if $\Sigma c_i a_{ij} = 0$ for every $j = 1, \dots, n$. Similarly,

$$\Sigma c_i A^i = (\Sigma c_i a_{1i}, \dots, \Sigma c_i a_{mi})$$

so that $\Sigma c_i A^i = 0$ if and only if $\Sigma c_i a_{ji} = 0$ for every $j = 1, \dots, m$ (remember that we usually write a column vector in the form of a row vector).

We first consider a transformation of type α. For definiteness, we interchange rows 1 and 2, although it will be obvious that any pair of rows will work. In other words, we define $\tilde{A}_1 = A_2$, $\tilde{A}_2 = A_1$ and $\tilde{A}_j = A_j$ for $j = 3, \ldots, n$. Therefore

$$f(\Sigma c_i A^i) = \Sigma c_i \tilde{A}^i = (\Sigma c_i a_{2i}, \Sigma c_i a_{1i}, \Sigma c_i a_{3i}, \ldots, \Sigma c_i a_{mi}) .$$

If

$$\Sigma c_i A^i = 0$$

then

$$\Sigma c_i a_{ji} = 0$$

for every $j = 1, \ldots, m$ and hence we see that $f(\Sigma c_i A^i) = 0$. This shows that f is well-defined for type α transformations. Conversely, if

$$f(\Sigma c_i A^i) = 0$$

then we see that again

$$\Sigma c_i a_{ji} = 0$$

for every $j = 1, \ldots, m$ since each component in the expression $\Sigma c_i \tilde{A}^i = 0$ must equal 0. Hence $\Sigma c_i A^i = 0$ if and only if $f(\Sigma c_i A^i) = 0$, and hence Ker f = {0} for type α transformations (which also shows that f is well-defined).

We leave it to the reader (see Exercise 3.3.1) to show that f is well-defined and Ker f = {0} for transformations of type β, and we go on to those of type γ. Again for definiteness, we consider the particular transformation $\tilde{A}_1 = A_1 + kA_2$ and $\tilde{A}_j = A_j$ for $j = 2, \ldots, m$. Then

$$\begin{aligned} f\left(\Sigma c_i A^i\right) &= \Sigma c_i \tilde{A}^i = \Sigma c_i \left(a_{1i} + ka_{2i}, a_{2i}, \ldots, a_{mi}\right) \\ &= \left(\Sigma c_i a_{1i} + \Sigma k c_i a_{2i}, \Sigma c_i a_{2i}, \ldots, \Sigma c_i a_{mi}\right) \end{aligned}$$

If

$$\Sigma c_i A^i = 0$$

then

$$\Sigma c_i a_{ji} = 0$$

for every $j = 1, \ldots, m$ so that $\Sigma c_i \tilde{A}^i = 0$ and f is well-defined for type γ transformations. Conversely, if

$$\Sigma c_i \tilde{A}^i = 0$$

then

$$\Sigma c_i a_{ji} = 0$$

for j = 2, . . . , m, and this then shows that $\sum c_i a_{1i} = 0$ also. Thus $\sum c_i \tilde{A}^i = 0$ implies that $\sum c_i A^i = 0$, and hence $\sum c_i A^i = 0$ if and only if $f(\sum c_i A^i) = 0$. This shows that Ker f = {0} for type γ transformations also, and f is well-defined.

In summary, by constructing an explicit isomorphism in each case, we have shown that the column spaces W and $\tilde{W}$ are isomorphic under all three types of elementary row operations, and hence it follows that the column spaces of row equivalent matrices must have the same dimension. ∎

Corollary If $\tilde{A}$ is the row-echelon form of A, then $\sum c_i A^i = 0$ if and only if $\sum c_i \tilde{A}^i = 0$.

Proof This was shown explicitly in the proof of Theorem 3.4 for each type of elementary row operation. ∎

In Theorem 3.3 we showed that every matrix is row equivalent to a reduced row-echelon matrix, and hence (by Theorem 3.4) any matrix and its row canonical form have the same row space. Note though, that if the original matrix has more rows than the dimension of its row space, then the rows obviously can not all be linearly independent. However, we now show that the nonzero rows of the row canonical form are in fact linearly independent, and hence form a basis for the row space.

Theorem 3.5 The nonzero row vectors of an m x n reduced row-echelon matrix R form a basis for the row space of R.

Proof From the four properties of a reduced row-echelon matrix, we see that if R has r nonzero rows, then there exist integers $j_1, \ldots, j_r$ with each $j_i \le n$ and $j_1 < \cdots < j_r$ such that R has a 1 in the i*th* row and j_i*th* column, and every other entry in the j_i*th* column is 0 (it may help to refer to Example 3.5 for visualization). If we denote these nonzero row vectors by $R_1, \ldots, R_r$ then any arbitrary vector

$$v = \sum_{i=1}^{r} c_i R_i$$

has c_i as its j_i*th* coordinate (note that v may have more than r coordinates if r < n). Therefore, if v = 0 we must have each coordinate of v equal to 0, and hence $c_i = 0$ for each i = 1, . . . , r. But this means that the R_i are linearly independent, and since $\{R_i\}$ spans the row space by definition, we see that they must in fact form a basis. ∎

Corollary If A is any matrix and R is a reduced row-echelon matrix row equivalent to A, then the nonzero row vectors of R form a basis for the row space of A.

Proof In Theorem 3.4 we showed that A and R have the same row space. The corollary now follows from Theorem 3.5. ∎

Example 3.6 Let us determine whether or not the following matrices have the same row space:

$$A = \begin{pmatrix} 1 & 2 & -1 & 3 \\ 2 & 4 & 1 & -2 \\ 3 & 6 & 3 & -7 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 2 & -4 & 11 \\ 2 & 4 & -5 & 14 \end{pmatrix} .$$

We leave it to the reader to show (and you really should do it) that the reduced row-echelon form of these matrices is

$$A = \begin{pmatrix} 1 & 2 & 0 & 1/3 \\ 0 & 0 & 1 & -8/3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 2 & 0 & 1/3 \\ 0 & 0 & 1 & -8/3 \end{pmatrix} .$$

Since the the nonzero rows of the reduced row-echelon form of A and B are identical, the row spaces must be the same. ⫽

Now that we have a better understanding of the row space of a matrix, let us go back and show that the reduced row-echelon form of a given matrix is unique. We first prove a preliminary result dealing with the row-echelon form of two matrices having the same row space.

Theorem 3.6 Let $A = (a_{ij})$ be a row-echelon matrix with distinguished elements $a_{1j_1}, a_{2j_2}, \ldots, a_{rj_r}$ and let $B = (b_{ij})$ be another row-echelon matrix with distinguished elements $b_{1k_1}, b_{2k_2}, \ldots, b_{sk_s}$. Assume that A and B have the same row space (and therefore the same number of columns). Then the distinguished elements of A are in the same position as those of B, i.e., $r = s$ and $j_1 = k_1, j_2 = k_2, \ldots, j_r = k_r$.

Proof Since $A = 0$ if and only if $B = 0$, we need only consider the nontrivial case where $r \geq 1$ and $s \geq 1$. First suppose that $j_1 < k_1$. This means that the j_1th column of B is equal to $(0, \ldots, 0)$. Since A and B have the same row space, the first row of A is just a linear combination of the rows of B. In particular, we then have $a_{1j_1} = \sum_i c_i b_{ij_1}$ for some set of scalars c_i (see the proof of

Theorem 3.4). But $b_{ij_1} = 0$ for every i, and hence $a_{1j_1} = 0$ which contradicts the assumption that a_{1j_1} is a distinguished element of A (and must be nonzero by definition). We are thus forced to conclude that $j_1 \ge k_1$. However, we could clearly have started with the assumption that $k_1 < j_1$, in which case we would have been led to conclude that $k_1 \ge j_1$. This shows that we must actually have $j_1 = k_1$.

Now let A′ and B′ be the matrices which result from deleting the first row of A and B respectively. If we can show that A′ and B′ have the same row space, then they will also satisfy the hypotheses of the theorem, and our conclusion follows at once by induction.

Let $R = (a_1, a_2, \dots, a_n)$ be any row of A′ (and hence a row of A), and let $B_1, \dots, B_m$ be the rows of B. Since A and B have the same row space, we again have

$$R = \sum_{i=1}^{m} d_i B_i$$

for some set of scalars d_i. Since R is not the first row of A and A′ is in row-echelon form, it follows that $a_i = 0$ for $i = j_1 = k_1$. In addition, the fact that B is in row-echelon form means that every entry in the k_1*th* column of B must be 0 except for the first, i.e., $b_{1k_1} \ne 0$, $b_{2k_1} = \cdots = b_{mk_1} = 0$. But then

$$0 = a_{k_1} = d_1 b_{1k_1} + d_2 b_{2k_1} + \cdots + d_m b_{mk_1} = d_1 b_{1k_1}$$

which implies that $d_1 = 0$ since $b_{1k_1} \ne 0$. This shows that R is actually a linear combination of the rows of B′, and hence (since R was arbitrary) the row space of A′ must be a subspace of the row space of B′. This argument can clearly be repeated to show that the row space of B′ is a subspace of the row space of A′, and hence we have shown that A′ and B′ have the same row space. ∎

Theorem 3.7 Let $A = (a_{ij})$ and $B = (b_{ij})$ be reduced row-echelon matrices. Then A and B have the same row space if and only if they have the same nonzero rows.

Proof Since it is obvious that A and B have the same row space if they have the same nonzero rows, we need only prove the converse. So, suppose that A and B have the same row space. Then if A_i is an arbitrary nonzero row of A, we may write

$$A_i = \Sigma_r c_r B_r \tag{1}$$

where the B_r are the nonzero rows of B. The proof will be finished if we can show that $c_r = 0$ for $r \ne i$ and $c_i = 1$.

To show that $c_i = 1$, let a_{ij_i} be the first nonzero entry in A_i, i.e., a_{ij_i} is the distinguished element of the i*th* row of A. Looking at the j_i*th* component of (1) we see that

$$a_{ij_i} = \Sigma_r c_r b_{rj_i} \tag{2}$$

(see the proof of Theorem 3.4). From Theorem 3.6 we know that b_{ij_i} is the distinguished element of the i*th* row of B, and hence it is the only nonzero entry in the j_i*th* column of B (by definition of a reduced row-echelon matrix). This means that (2) implies $a_{ij_i} = c_i\, b_{ij_i}$. In fact, it must be true that $a_{ij_i} = b_{ij_i} = 1$ since A and B are reduced row-echelon matrices, and therefore $c_i = 1$.

Now let b_{kj_k} be the first nonzero entry of B_k (where $k \neq i$). From (1) again we have

$$a_{ij_k} = \Sigma_r c_r b_{rj_k} \quad . \tag{3}$$

Since B is a reduced row-echelon matrix, $b_{kj_k} = 1$ is the only nonzero entry in the j_k*th* column of B, and hence (3) shows us that $a_{ij_k} = c_k b_{kj_k}$. But from Theorem 3.6, a_{kj_k} is a distinguished element of A, and hence the fact that A is row-reduced means that $a_{ij_k} = 0$ for $i \neq k$. This forces us to conclude that $c_k = 0$ for $k \neq i$ as claimed. ∎

Suppose that two people are given the same matrix A and asked to transform it to reduced row-echelon form R. The chances are quite good that they will each perform a different sequence of elementary row operations to achieve the desired result. Let R and R′ be the reduced row-echelon matrices that our two students obtain. We claim that R = R′. Indeed, since row equivalence defines an equivalence relation, we see from Theorem 3.4 that the row spaces of R and R′ will be the same. Therefore Theorem 3.7 shows us that the rows of R must equal the rows of R′. Hence we are justified in calling the reduced row-echelon form of a matrix *the* row canonical form as mentioned earlier.

Exercises

1. In the proof of Theorem 3.4, show that Ker f = {0} for a type β operation.

2. Determine whether or not the following matrices have the same row space:

$$A = \begin{pmatrix} 1 & -2 & -1 \\ 3 & -4 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & -1 & 2 \\ 3 & 3 & -1 \end{pmatrix} \qquad C = \begin{pmatrix} 1 & -1 & 3 \\ 2 & -1 & 10 \\ 3 & -5 & 1 \end{pmatrix} .$$

3. Show that the subspace of $\mathbb{R}^3$ spanned by the vectors $u_1 = (1, 1, -1)$, $u_2 = (2, 3, -1)$ and $u_3 = (3, 1, -5)$ is the same as the subspace spanned by the vectors $v_1 = (1, -1, -3)$, $v_2 = (3, -2, -8)$ and $v_3 = (2, 1, -3)$.

4. Determine whether or not each of the following sets of vectors is linearly independent:
   (a) $u_1 = (1, -2, 1)$, $u_2 = (2, 1, -1)$ and $u_3 = (7, -4, 1)$.
   (b) $u_1 = (1, 2, -3)$, $u_2 = (1, -3, 2)$ and $u_3 = (2, -1, 5)$.

5. (a) Suppose we are given an m x n matrix $A = (a_{ij})$, and suppose that one of the columns of A, say A^i, is a linear combination of the others. Show that under any elementary row operation resulting in a new matrix $\tilde{A}$, the column $\tilde{A}^i$ is the same linear combination of the columns of $\tilde{A}$ that A^i is of the columns of A. In other words, show that all linear relations between columns are preserved by elementary row operations.
   (b) Use this result to give another proof of Theorem 3.4.
   (c) Use this result to give another proof of Theorem 3.7.

3.4 THE RANK OF A MATRIX

It is important for the reader to realize that there is nothing special about the rows of a matrix. Everything that we have done up to this point in discussing elementary row operations could just as easily have been done with columns instead. In particular, this means that Theorems 3.4 and 3.5 are equally valid for column spaces if we apply our elementary transformations to columns instead of rows. This observation leads us to our next fundamental result.

Theorem 3.8 If $A = (a_{ij})$ is any m x n matrix over a field $\mathcal{F}$, then $rr(A) = cr(A)$.

Proof Let $\tilde{A}$ be the reduced row-echelon form of A. By Theorem 3.4 it is sufficient to show that $rr(\tilde{A}) = cr(\tilde{A})$. If $j_1 < \cdots < j_r$ are the columns containing the distinguished elements of $\tilde{A}$, then $\{A^{j_1}, \ldots, A^{j_r}\}$ is a basis for the column space of $\tilde{A}$, and hence $cr(\tilde{A}) = r$. (In fact, these columns are just the first r standard basis vectors in $\mathcal{F}^n$.) But from the corollary to Theorem 3.5, we see that rows $\tilde{A}_1, \ldots, \tilde{A}_r$ form a basis for the row space of $\tilde{A}$, and thus $rr(\tilde{A}) = r$ also. ∎

In view of this result, we define the **rank** r(A) of a matrix A as

$$r(A) = rr(A) = cr(A) .$$

Our next theorem forms the basis for a practical method of finding the rank of a matrix.

Theorem 3.9 If A is any matrix, then r(A) is equal to the number of nonzero rows in the (reduced) row-echelon matrix row equivalent to A. (Alternatively, r(A) is the number of nonzero columns in the (reduced) column-echelon matrix column equivalent to A.)

Proof Noting that the number of nonzero rows in the row-echelon form is the same as the number of nonzero rows in the reduced row-echelon form, we see that this theorem follows directly from the corollary to Theorem 3.5. ∎

If A is an n x n matrix such that $a_{ij} = 0$ for $i \neq j$ and $a_{ii} = 1$, then we say that A is the **identity matrix** of size n and write this matrix as I_n. Since the size is usually understood, we will generally simply write I. If $I = (I_{ij})$, then another useful way of writing this is in terms of the Kronecker delta as $I_{ij} = \delta_{ij}$. Written out, I has the form

$$I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} .$$

Theorem 3.10 If A is an n x n matrix of rank n, then the reduced row-echelon matrix row equivalent to A is the identity matrix I_n.

Proof This follows from the definition of a reduced row-echelon matrix and Theorem 3.9. ∎

Example 3.7 Let us find the rank of the matrix A given by

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

To do this, we will apply Theorem 3.9 to columns instead of rows (just for variety). Proceeding with the elementary transformations, we obtain the following sequence of matrices:

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & -3 & 6 \\ -2 & 3 & -3 \\ -1 & 6 & -5 \end{pmatrix}$$
$$\qquad\quad \uparrow \qquad \uparrow$$
$$A^2 - 2A^1 \quad A^3 + 3A^1$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & -1 & 0 \\ -2 & 1 & 1 \\ -1 & 2 & 7/3 \end{pmatrix}$$
$$\qquad\quad \uparrow \qquad \uparrow$$
$$(1/3)A^2 \quad (1/3)(A^3 + 2A^2)$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 3 & 1/3 & 7/3 \end{pmatrix}$$
$$\qquad\quad \uparrow \qquad \uparrow$$
$$A^1 + 2A^2 \quad -(A^2 - A^3)$$

Thus the reduced column-echelon form of A has three nonzero columns, so that r(A) = cr(A) = 3. We leave it to the reader (see Exercise 3.4.1) to show that the row canonical form of A is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

and hence r(A) = cr(A) = rr(A) as it should. ⁄⁄

Exercises

1. Verify the row-canonical form of the matrix in Example 3.7.

2. Let A and B be arbitrary m x n matrices. Show that r(A + B) ≤ r(A) + r(B).

3. Using elementary row operations, find the rank of each of the following matrices:

$$(a)\ \begin{pmatrix} 1 & 3 & 1 & -2 & -3 \\ 1 & 4 & 3 & -1 & -4 \\ 2 & 3 & -4 & -7 & -3 \\ 3 & 8 & 1 & -7 & -8 \end{pmatrix} \qquad (b)\ \begin{pmatrix} 1 & 2 & -3 \\ 2 & 1 & 0 \\ -2 & -1 & 3 \\ -1 & 4 & -2 \end{pmatrix}$$

$$(c)\ \begin{pmatrix} 1 & 3 \\ 0 & -2 \\ 5 & -1 \\ -2 & 3 \end{pmatrix} \qquad (d)\ \begin{pmatrix} 5 & -1 & 1 \\ 2 & 1 & -2 \\ 0 & -7 & 12 \end{pmatrix}$$

4. Repeat the previous problem using elementary column operations.

3.5 SOLUTIONS TO SYSTEMS OF LINEAR EQUATIONS

We now apply the results of the previous section to the determination of some general characteristics of the solution set to systems of linear equations. We will have more to say on this subject after we have discussed determinants in the next chapter.

To begin with, a system of linear equations of the form

$$\sum_{j=1}^{n} a_{ij} x_j = 0, \qquad i = 1, \dots, m$$

is called a **homogeneous system** of m linear equations in n unknowns. It is obvious that choosing $x_1 = x_2 = \cdots = x_n = 0$ will satisfy this system, but this is not a very interesting solution. It is called the **trivial** (or **zero**) **solution**. Any other solution, if it exists, is referred to as a **nontrivial solution**.

A more general type of system of linear equations is of the form

$$\sum_{j=1}^{n} a_{ij} x_j = y_i, \qquad i = 1, \dots, m$$

where each y_i is a given scalar. This is then called a **nonhomogeneous system** of linear equations. Let us define the column vector

$$Y = (y_1, \ldots, y_m) \in \mathcal{F}^m .$$

From our discussion in the proof of Theorem 3.4, we see that $a_{ij}x_j$ is just x_j times the i*th* component of the j*th* column $A^j \in \mathcal{F}^m$. Thus our system of nonhomogeneous equations may be written in the form

$$\sum_{j=1}^n A^j x_j = x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = Y$$

where this vector equation is to be interpreted in terms of its components. (In the next section, we shall see how to write this as a product of matrices.) It should also be obvious that a homogeneous system may be written in this notation as

$$\sum_{j=1}^n A^j x_j = 0 \ .$$

Let us now look at some elementary properties of the solution set of a homogeneous system of equations.

Theorem 3.11 The solution set S of a homogeneous system of m equations in n unknowns is a subspace of $\mathcal{F}^n$.

Proof Let us write our system as $\sum_j a_{ij}x_j = 0$. We first note that $S \neq \varnothing$ since $(0, \ldots, 0) \in \mathcal{F}^n$ is the trivial solution of our system. If $u = (u_1, \ldots, u_n) \in \mathcal{F}^n$ and $v = (v_1, \ldots, v_n) \in \mathcal{F}^n$ are both solutions, then

$$\sum_j a_{ij}(u_j + v_j) = \sum_j a_{ij}u_j + \sum_j a_{ij}v_j = 0$$

so that $u + v \in S$. Finally, if $c \in \mathcal{F}$ then we also have

$$\sum_j a_{ij}(cu_j) = c\sum_j a_{ij}u_j = 0$$

so that $cu \in S$. ∎

If we look back at Example 3.4, we see that a system of m equations in n > m unknowns will necessarily result in a nonunique, and hence nontrivial, solution. The formal statement of this fact is contained in our next theorem.

Theorem 3.12 Let a homogeneous system of m equations in n unknowns have the m x n matrix of coefficients A. Then the system has a nontrivial solution if and only if $r(A) < n$.

Proof By writing the system in the form $\sum_j x_j A^j = 0$, it is clear that a nontrivial solution exists if and only if the n column vectors $A^j \in \mathcal{F}^m$ are linearly dependent. Since the rank of A is equal to the dimension of its column space, we must therefore have $r(A) < n$. ∎

It should now be clear that if an n x n (i.e., square) matrix of coefficients A (of a homogeneous system) has rank equal to n, then the only solution will be the trivial solution since reducing the augmented matrix (which then has the last column equal to the zero vector) to reduced row-echelon form will result in each variable being set equal to zero (see Theorem 3.10).

Theorem 3.13 Let a homogeneous system of linear equations in n unknowns have a matrix of coefficients A. Then the solution set S of this system is a subspace of $\mathcal{F}^n$ with dimension $n - r(A)$.

Proof Assume that S is a nontrivial solution set, so that by Theorem 3.12 we have $r(A) < n$. Assume also that the unknowns $x_1, \ldots, x_n$ have been ordered in such a way that the first $k = r(A)$ columns of A span the column space (this is guaranteed by Theorem 3.4). Then the remaining columns $A^{k+1}, \ldots, A^n$ may be written as

$$A^i = \sum_{j=1}^{k} b_{ij} A^j, \qquad i = k+1, \ldots, n$$

and where each $b_{ij} \in \mathcal{F}$. If we define $b_{ii} = -1$ and $b_{ij} = 0$ for $j \neq i$ and $j > k$, then we may write this as

$$\sum_{j=1}^{n} b_{ij} A^j = 0, \qquad i = k+1, \ldots, n$$

(note the upper limit on this sum differs from the previous equation). Next we observe that the solution set S consists of all vectors $x \in \mathcal{F}^n$ such that

$$\sum_{j=1}^{n} x_j A^j = 0$$

and hence in particular, the $n - k$ vectors

$$b^{(i)} = (b_{i1}, \ldots, b_{in})$$

for each i = k + 1, . . . , n must belong to S. We show that they in fact form a basis for S, which is then of dimension n - k.

To see this, we first write out each of the $b^{(i)}$:

$$\begin{aligned} b^{(k+1)} &= (b_{k+1\,1}, \dots, b_{k+1\,k}, -1, 0, 0, \dots, 0) \\ b^{(k+2)} &= (b_{k+2\,1}, \dots, b_{k+2\,k}, 0, -1, 0, \dots, 0) \\ &\vdots \\ b^{(n)} &= (b_{n1}, \dots, b_{nk}, 0, 0, \dots, 0, -1) \ . \end{aligned}$$

Hence for any set $\{c_i\}$ of n - k scalars we have

$$\sum_{i=k+1}^{n} c_i b^{(i)} = \left(\sum_{i=k+1}^{n} c_i b_{i1}, \dots, \sum_{i=k+1}^{n} c_i b_{in}, -c_{k+1}, \dots, -c_n \right)$$

and therefore

$$\sum_{i=k+1}^{n} c_i b^{(i)} = 0$$

if and only if $c_{k+1} = \cdots = c_n = 0$. This shows that the $b^{(i)}$ are linearly independent. (This should have been obvious from their form shown above.)

Now suppose that $d = (d_1, \dots, d_n)$ is any solution of

$$\sum_{j=1}^{n} x_j A^j = 0 \ .$$

Since S is a vector space (Theorem 3.11), any linear combination of solutions is a solution, and hence the vector

$$y = d + \sum_{i=k+1}^{n} d_i b^{(i)}$$

must also be a solution. In particular, writing out each component of this expression shows that

$$y_j = d_j + \sum_{i=k+1}^{n} d_i b_{ij}$$

and hence the definition of the b_{ij} shows that $y = (y_1, \dots, y_k, 0, \dots, 0)$ for some set of scalars y_i. Therefore, we have

$$0 = \sum_{j=1}^{n} y_j A^j = \sum_{j=1}^{k} y_j A^j$$

and since $\{A^1, \dots, A^k\}$ is linearly independent, this implies that $y_j = 0$ for each j = 1, . . . , k. Hence y = 0 so that

$$d = -\sum_{i=k+1}^{n} d_i b^{(i)}$$

and we see that any solution may be expressed as a linear combination of the $b^{(i)}$.

Since the $b^{(i)}$ are linearly independent and we just showed that they span S, they must form a basis for S. ∎

Suppose that we have a homogeneous system of m equations in $n > m$ unknowns, and suppose that the coefficient matrix A is in row-echelon form and has rank m. Then each of the m successive equations contains fewer and fewer unknowns, and since there are more unknowns than equations, there will be $n - m = n - r(A)$ unknowns that do not appear as the first entry in any of the rows of A. These $n - r(A)$ unknowns are called **free variables**. We may arbitrarily assign any value we please to the free variables to obtain a solution of the system.

Let the free variables of our system be $x_{i_1}, \dots, x_{i_k}$ where $k = n - m = n - r(A)$, and let v_s be the solution vector obtained by setting x_{i_s} equal to 1 and each of the remaining free variables equal to 0. (This is essentially what was done in the proof of Theorem 3.13.) We claim that $v_1, \dots, v_k$ are linearly independent and hence form a basis for the solution space of the (homogeneous) system (which is of dimension $n - r(A)$ by Theorem 3.13).

To see this, we basically follow the proof of Theorem 3.13 and let B be the k x n matrix whose rows consist of the solution vectors v_s. For each s, our construction is such that we have $x_{i_s} = 1$ and $x_{i_r} = 0$ for $r \neq s$ (and the remaining $m = n - k$ unknowns are in general nonzero). In other words, the solution vector v_s has a 1 in the position of x_{i_s}, while for $r \neq s$ the vector v_r has a 0 in this same position. This means that each of the k columns corresponding to the free variables in the matrix B contains a single 1 and the rest zeros. We now interchange column 1 and column i_1, then column 2 and column $i_2, \dots,$ and finally column k and column i_k. This yields the matrix

$$C = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & b_{1\,k+1} & \cdots & b_{1n} \\ 0 & 1 & 0 & \cdots & 0 & 0 & b_{2\,k+1} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & 0 & 1 & b_{k\,k+1} & \cdots & b_{kn} \end{pmatrix}$$

where the entries $b_{i\,k+1}, \dots, b_{in}$ are the values of the remaining m unknowns in the solution vector v_i. Since the matrix C is in row-echelon form, its rows are independent and hence $r(C) = k$. However, C is column-equivalent to B, and therefore $r(B) = k$ also (by Theorem 3.4 applied to columns). But the rows of B consist precisely of the k solution vectors v_s, and thus these solution vectors must be independent as claimed.

Example 3.8 Consider the homogeneous system of linear equations

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

If we reduce this system to row-echelon form, we obtain

$$\begin{aligned} x + 2y - 4z + 3w - t &= 0 \\ 2z - w + 2t &= 0 \end{aligned} \qquad (*)$$

It is obvious that the rank of the matrix of coefficients is 2, and hence the dimension of the solution space is 5 - 2 = 3. The free variables are clearly y, w and t. The solution vectors v_s are obtained by choosing (y = 1, w = 0, t = 0), (y = 0, w = 1, t = 0) and (y = 0, w = 0, t = 1). Using each of the these in the system (*), we obtain the solutions

$$\begin{aligned} v_1 &= (-2, 1, 0, 0, 0) \\ v_2 &= (-1, 0, 1/2, 1, 0) \\ v_3 &= (-3, 0, -1, 0, 1) \end{aligned}$$

Thus the vectors v_1, v_2 and v_3 form a basis for the solution space of the homogeneous system. ∕∕

We emphasize that the corollary to Theorem 3.4 shows us that the solution set of a homogeneous system of equations is unchanged by elementary row operations. It is this fact that allows us to proceed as we did in Example 3.8.

We now turn our attention to the solutions of a nonhomogeneous system of equations $\sum_j a_{ij}x_j = y_i$.

Theorem 3.14 Let a nonhomogeneous system of linear equations have matrix of coefficients A. Then the system has a solution if and only if r(A) = r(aug A).

Proof Let $c = (c_1, \dots, c_n)$ be a solution of $\sum_j a_{ij}x_j = y_i$. Then writing this as

$$\sum_j c_j A^j = Y$$

shows us that Y is in the column space of A, and hence

$$r(\text{aug } A) = cr(\text{aug } A) = cr(A) = r(A) .$$

Conversely, if cr(aug A) = r(aug A) = r(A) = cr(A), then Y is in the column space of A, and hence $Y = \sum c_j A^j$ for some set of scalars c_j . But then the vector $c = (c_1, \dots, c_n)$ is a solution since it obviously satisfies $\sum_j a_{ij} x_j = y_i$. ∎

Using Theorem 3.13, it is easy to describe the general solution to a nonhomogeneous system of equations.

Theorem 3.15 Let

$$\sum_{j=1}^{n} a_{ij} x_j = y_j$$

be a system of nonhomogeneous linear equations. If $u = (u_1, \dots, u_n) \in \mathcal{F}^n$ is a solution of this system, and if S is the solution space of the associated homogeneous system, then the set

$$u + S = \{u + v \colon v \in S\}$$

is the solution set of the nonhomogeneous system.

Proof If $w = (w_1, \dots, w_n) \in \mathcal{F}^n$ is any other solution of $\sum_j a_{ij} x_j = y_i$, then

$$\sum_j a_{ij}(w_j - u_j) = \sum_j a_{ij} w_j - \sum_j a_{ij} u_j = y_i - y_i = 0$$

so that $w - u \in S$, and hence $w = u + v$ for some $v \in S$. Conversely, if $v \in S$ then

$$\sum_j a_{ij}(u_j + v_j) = \sum_j a_{ij} u_j + \sum_j a_{ij} v_j = y_j + 0 = y_j$$

so that u + v is a solution of the nonhomogeneous system. ∎

Theorem 3.16 Let A be an n x n matrix of rank n. Then the system

$$\sum_{j=1}^{n} A^j x_j = Y$$

has a unique solution for arbitrary vectors $Y \in \mathcal{F}^n$.

Proof Since $Y = \sum A^j x_j$, we see that $Y \in \mathcal{F}^n$ is just a linear combination of the columns of A. Since r(A) = n, it follows that the columns of A are linearly independent and hence form a basis for $\mathcal{F}^n$. But then any $Y \in \mathcal{F}^n$ has a unique expansion in terms of this basis (Theorem 2.4, Corollary 2) so that the vector X with components x_j must be unique. ∎

Example 3.9 Let us find the complete solution set over the real numbers of the nonhomogeneous system

$$\begin{aligned} 3x_1 + \ x_2 + \ 2x_3 + \ 4x_4 &= \ 1 \\ x_1 - \ x_2 + \ 3x_3 - \ x_4 &= \ 3 \\ x_1 + 7x_2 - 11x_3 + 13x_4 &= -13 \\ 11x_1 + \ x_2 + 12x_3 + 10x_4 &= \ 9 \end{aligned}$$

We assume that we somehow found a particular solution u = (2, 5, 1, −3) ∈ $\mathbb{R}^4$, and hence we seek the solution set S of the associated homogeneous system. The matrix of coefficients A of the homogeneous system is given by

$$A = \begin{pmatrix} 3 & 1 & 2 & 4 \\ 1 & -1 & 3 & -1 \\ 1 & 7 & -11 & 13 \\ 11 & 1 & 12 & 10 \end{pmatrix} .$$

The first thing we must do is determine r(A). Since the proof of Theorem 3.13 dealt with columns, we choose to construct a new matrix B by applying elementary column operations to A. Thus we define

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 4 & 5 & 3 \\ 7 & -20 & -25 & -15 \\ 1 & 8 & 10 & 6 \end{pmatrix}$$

where the columns of B are given in terms of those of A by $B^1 = A^2$, $B^2 = A^1 - 3A^2$, $B^3 = A^3 - 2A^2$ and $B^4 = A^4 - 4A^2$. It is obvious that B^1 and B^2 are independent, and we also note that $B^3 = (5/4)B^2$ and $B^4 = (3/4)B^2$. Then r(A) = r(B) = 2, and hence we have dim S = 4 − 2 = 2.

(An alternative method of finding r(A) is as follows. If we interchange the first two rows of A and then add a suitable multiple the new first row to eliminate the first entry in each of the remaining three rows, we obtain

$$\begin{pmatrix} 1 & -1 & 3 & -1 \\ 0 & 4 & -7 & 7 \\ 0 & 8 & -14 & 14 \\ 0 & 12 & -21 & 21 \end{pmatrix} .$$

It is now clear that the first two rows of this matrix are independent, and that the third and fourth rows are each multiples of the second. Therefore r(A) = 2 as above.)

We now follow the first part of the proof of Theorem 3.13. Observe that since r(A) = 2 and the first two columns of A are independent, we may write

$$A^3 = (5/4)A^1 - (7/4)A^2$$

and

$$A^4 = (3/4)A^1 + (7/4)A^2 \quad .$$

We therefore define the vectors

$$b^{(3)} = (5/4, -7/4, -1, 0)$$

and

$$b^{(4)} = (3/4, 7/4, 0, -1)$$

which are independent solutions of the homogeneous system and span the solution space S. Therefore the general solution set to the nonhomogeneous system is given by

$$\begin{aligned} u + S &= \{u + \alpha b^{(3)} + \beta b^{(4)}\} \\ &= \{(2, 5, 1, -3) + \alpha(5/4, -7/4, -1, 0) + \beta(3/4, 7/4, 0, 1)\} \end{aligned}$$

where $\alpha, \beta \in \mathbb{R}$ are arbitrary. ⁄⁄

Exercises

1. Find the dimension and a basis for the solution space of each of the following systems of linear equations over $\mathbb{R}$:

(a)
$$\begin{aligned} x + 4y + 2z &= 0 \\ 2x + y + 5z &= 0 \end{aligned}$$

(b)
$$\begin{aligned} x + 3y + 2z &= 0 \\ x + 5y + z &= 0 \\ 3x + 5y + 8z &= 0 \end{aligned}$$

(c)
$$\begin{aligned} x + 2y + 2z - w + 3t &= 0 \\ x + 2y + 3z + w + t &= 0 \\ 3x + 6y + 8z + w + t &= 0 \end{aligned}$$

(d)
$$\begin{aligned} x + 2y - 2z - 2w - t &= 0 \\ x + 2y - z + 3w - 2t &= 0 \\ 2x + 4y - 7z + w + t &= 0 \end{aligned}$$

2. Consider the subspaces U and V of $\mathbb{R}^4$ given by

$$U = \{(a, b, c, d) \in \mathbb{R}^4 : b + c + d = 0\}$$
$$V = \{(a, b, c, d) \in \mathbb{R}^4 : a + b = 0 \text{ and } c = 2d\} \ .$$

(a) Find the dimension and a basis for U.
(b) Find the dimension and a basis for V.
(c) Find the dimension and a basis for U ∩ V.

3. Find the complete solution set of each of the following systems of linear equations over $\mathbb{R}$:

$$(a)\quad \begin{aligned} 3x - y &= 7 \\ 2x + y &= 1 \end{aligned}$$

$$(b)\quad \begin{aligned} 2x - y + 3z &= 5 \\ 3x + 2y - 2z &= 1 \\ 7x + \phantom{2y - {}} 4z &= 11 \end{aligned}$$

$$(c)\quad \begin{aligned} 5x + 2y - z &= 0 \\ 3x + 5y + 3z &= 0 \\ x + 8y + 7z &= 0 \end{aligned}$$

$$(d)\quad \begin{aligned} x - y + 2z + w &= 3 \\ 2x + y - z - w &= 1 \\ 3x + y + z - 3w &= 2 \\ 3x - 2y + 6z \phantom{{}- 3w} &= 7 \end{aligned}$$

3.6 MATRIX ALGEBRA

We now introduce the elementary algebraic operations on matrices. These operations will be of the utmost importance throughout the remainder of this text. In Chapter 5 we will see how these definitions arise in a natural way from the algebra of linear transformations.

Given two m x n matrices $A = (a_{ij})$ and $B = (b_{ij})$, we define their **sum** A + B to be the matrix with entries

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

obtained by adding the corresponding entries of each matrix. Note that both A and B must be of the same size. We also say that A **equals** B if $a_{ij} = b_{ij}$ for all i and j. It is obvious that

$$A + B = B + A$$

and that

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

for any other m x n matrix C. We also define the **zero matrix** 0 as that matrix for which A + 0 = A. In other words, $(0)_{ij} = 0$ for every i and j. Given a matrix $A = (a_{ij})$, we define its **negative** (or **additive inverse**)

$$-A = (-a_{ij})$$

such that A + (−A) = 0. Finally, for any scalar c we define the product of c and A to be the matrix

$$cA = (ca_{ij}) .$$

Since in general the entries a_{ij} in a matrix $A = (a_{ij})$ are independent of each other, it should now be clear that the set of all m x n matrices forms a vector space of dimension mn over a field $\mathcal{F}$ of scalars. In other words, any m x n matrix A with entries a_{ij} can be written in the form

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}E_{ij}$$

where the m x n matrix E_{ij} is defined as having a 1 in the (i, j)*th* position and 0's elsewhere, and there are clearly mn such matrices. We denote the space of all m x n matrices over the field $\mathcal{F}$ by $M_{m\times n}(\mathcal{F})$. The particular case of m = n defines the space $M_n(\mathcal{F})$ of all **square** matrices of size n. We will often refer to a matrix in $M_n(\mathcal{F})$ as an **n-square** matrix.

Now let $A \in M_{m\times n}(\mathcal{F})$ be an m x n matrix, $B \in M_{r\times m}(\mathcal{F})$ be an r x m matrix, and consider the two systems of linear equations

$$\sum_{j=1}^{n} a_{ij}x_j = y_i, \qquad i = 1, \dots, m$$

and

$$\sum_{j=1}^{m} b_{ij}y_j = z_i, \qquad i = 1, \dots, r$$

where $X = (x_1, \dots, x_n) \in \mathcal{F}^n$, $Y = (y_1, \dots, y_m) \in \mathcal{F}^m$ and $Z = (z_1, \dots, z_r) \in \mathcal{F}^r$. Substituting the first of these equations into the second yields

$$z_i = \textstyle\sum_j b_{ij}y_j = \sum_j b_{ij}\sum_k a_{jk}x_k = \sum_k c_{ik}x_k$$

where we defined the **product** of the r x m matrix B and the m x n matrix A to be the r x n matrix C = BA whose entries are given by

$$c_{ik} = \sum_{j=1}^{m} b_{ij}a_{jk} \quad .$$

Thus the (i, k)*th* entry of C = BA is given by the standard scalar product

$$(BA)_{ik} = B_i \bullet A^k$$

of the i*th* row of B with the k*th* column of A (where both are considered as vectors in $\mathcal{F}^m$). Note that matrix multiplication is generally not commutative, i.e., AB ≠ BA. Indeed, the product AB may not even be defined.

Example 3.10 Let A and B be given by

$$A = \begin{pmatrix} 1 & 6 & -2 \\ 3 & 4 & 5 \\ 7 & 0 & 8 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & -9 \\ 6 & 1 \\ 1 & -3 \end{pmatrix} .$$

Then the product of A and B is given by

$$C = AB = \begin{pmatrix} 1 & 6 & -2 \\ 3 & 4 & 5 \\ 7 & 0 & 8 \end{pmatrix} \begin{pmatrix} 2 & -9 \\ 6 & 1 \\ 1 & -3 \end{pmatrix} = \begin{pmatrix} 1\cdot 2 + 6\cdot 6 - 2\cdot 1 & -1\cdot 9 + 6\cdot 1 + 2\cdot 3 \\ 3\cdot 2 + 4\cdot 6 + 5\cdot 1 & -3\cdot 9 + 4\cdot 1 - 5\cdot 3 \\ 7\cdot 2 + 0\cdot 6 + 8\cdot 1 & -7\cdot 9 + 0\cdot 1 - 8\cdot 3 \end{pmatrix}$$

$$= \begin{pmatrix} 36 & 3 \\ 35 & -38 \\ 22 & -87 \end{pmatrix} .$$

Note that it makes no sense to evaluate the product BA.

It is also easy to see that if we have the matrices

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

then

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

while

$$BA = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} \neq AB \ . \ /\!/$$

Example 3.11 Two other special cases of matrix multiplication are worth explicitly mentioning. Let $X \in \mathcal{F}^n$ be the column vector

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} .$$

If A is an m x n matrix, we may consider X to be an n x 1 matrix and form the product AX:

$$AX = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} A_1 \bullet X \\ \vdots \\ A_m \bullet X \end{pmatrix} .$$

As expected, the product AX is an m x 1 matrix with entries given by the standard scalar product $A_i \bullet X$ in $\mathcal{F}^n$ of the i*th* row of A with the vector X. Note that this may also be written in the form

$$AX = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} x_1 + \cdots + \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} x_n$$

which clearly shows that AX is just a linear combination of the columns of A.

Now let $Y \in \mathcal{F}^m$ be the row vector $Y = (y_1, \ldots, y_m)$. If we view this as a 1 x m matrix, then we may form the 1 x n matrix product YA given by

$$\begin{aligned} YA &= (y_1, \ldots, y_m) \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \\ &= (y_1 a_{11} + \cdots + y_m a_{m1}, \ldots, y_1 a_{1n} + \cdots + y_m a_{mn}) \\ &= (Y \bullet A^1, \ldots, Y \bullet A^n) \ . \end{aligned}$$

This again yields the expected form of the product with entries $Y \bullet A^i$. ∕∕

This example suggests the following commonly used notation for systems of linear equations. Consider the system

$$\sum_{j=1}^{n} a_{ij} x_j = y_i$$

where $A = (a_{ij})$ is an m x n matrix. Suppose that we define the column vectors

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathcal{F}^n \quad \text{and} \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathcal{F}^m \ .$$

If we consider X to be an n x 1 matrix and Y to be an m x 1 matrix, then we may write this system in matrix notation as

$$AX = Y \ .$$

Note that the i*th* row vector of A is $A_i = (a_{i1}, \ldots, a_{in})$ so that the expression $\sum_j a_{ij} x_j = y_i$ may be written as the standard scalar product

$$A_i \bullet X = y_i \ .$$

We leave it to the reader to show that if A is an n x n matrix, then

$$A I_n = I_n A = A \ .$$

Even if A and B are both square matrices (i.e., matrices of the form m x m), the product AB will not generally be the same as BA unless A and B are diagonal matrices (see Exercise 3.6.4). However, we do have the following.

Theorem 3.17 For matrices of proper size (so that these operations are defined), we have:

(a) (AB)C = A(BC) (associative law).
(b) A(B + C) = AB + AC (left distributive law).
(c) (B + C)A = BA + CA (right distributive law).
(d) k(AB) = (kA)B = A(kB) for any scalar k.

Proof (a) $[(AB)C]_{ij} = \sum_k (AB)_{ik} c_{kj} = \sum_{r,\,k} (a_{ir} b_{rk}) c_{kj} = \sum_{r,\,k} a_{ir} (b_{rk} c_{kj})$
$= \sum_r a_{ir} (BC)_{rj} = [A(BC)]_{ij}$.

(b) $[A(B+C)]_{ij} = \sum_k a_{ik} (B+C)_{kj} = \sum_k a_{ik} (b_{kj} + c_{kj})$
$= \sum_k a_{ik} b_{kj} + \sum_k a_{ik} c_{kj} = (AB)_{ij} + (AC)_{ij}$
$= [(AB) + (AC)]_{ij}$.

(c) Left to the reader (Exercise 3.6.1).
(d) Left to the reader (Exercise 3.6.1). ∎

Given a matrix $A = (a_{ij})$, we define the **transpose** of A, denoted by $A^T = (a^T{}_{ij})$ to be the matrix with entries given by $a^T{}_{ij} = a_{ji}$. In other words, if A is an m x n matrix, then A^T is an n x m matrix whose columns are just the rows of A. Note in particular that a column vector is just the transpose of a row vector.

Example 3.12 If A is given by

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

then A^T is given by

$$\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} . \;/\!\!/$$

Theorem 3.18 The transpose has the following properties:

(a) $(A + B)^T = A^T + B^T$.

(b) $(A^T)^T = A$.

(c) $(cA)^T = cA^T$ for any scalar c.

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

Proof (a) $[(A + B)^T]_{ij} = [(A + B)]_{ji} = a_{ji} + b_{ji} = a^T{}_{ij} + b^T{}_{ij} = (A^T + B^T)_{ij}$.

(b) $(A^T)^T{}_{ij} = (A^T)_{ji} = a_{ij} = (A)_{ij}$.

(c) $(cA)^T{}_{ij} = (cA)_{ji} = ca_{ji} = c(A^T)_{ij}$.

(d) $(AB)^T{}_{ij} = (AB)_{ji} = \sum_k a_{jk}b_{ki} = \sum_k b^T{}_{ik}a^T{}_{kj} = (B^T A^T)_{ij}$. ∎

We now wish to relate this algebra to our previous results dealing with the rank of a matrix. Before doing so, let us first make some elementary observations dealing with the rows and columns of a matrix product. Assume that $A \in M_{m\times n}(\mathcal{F})$ and $B \in M_{n\times r}(\mathcal{F})$ so that the product AB is defined. Since the (i, j)*th* entry of AB is given by $(AB)_{ij} = \sum_k a_{ik}b_{kj}$, we see that the i*th* row of AB is given by a linear combination of the rows of B:

$$(AB)_i = (\textstyle\sum_k a_{ik}b_{k1}, \ldots, \sum_k a_{ik}b_{kr}) = \sum_k a_{ik}(b_{k1}, \ldots, b_{kr}) = \sum_k a_{ik}B_k .$$

Another way to write this is to observe that

$$(AB)_i = (\Sigma_k a_{ik} b_{k1}, \ldots , \Sigma_k a_{ik} b_{kr})$$
$$= (a_{i1}, \ldots , a_{in}) \begin{pmatrix} b_{11} & \cdots & b_{1r} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nr} \end{pmatrix} = A_i B \ .$$

Similarly, for the columns of a product we find that the *jth* column of AB is a linear combination of the columns of A:

$$(AB)^j = \begin{pmatrix} \Sigma_k a_{1k} b_{kj} \\ \vdots \\ \Sigma_k a_{mk} b_{kj} \end{pmatrix} = \sum_{k=1}^{n} \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix} b_{kj} = \sum_{k=1}^{n} A^k b_{kj}$$

and

$$(AB)^j = \begin{pmatrix} \Sigma_k a_{1k} b_{kj} \\ \vdots \\ \Sigma_k a_{mk} b_{kj} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{pmatrix} = AB^j \ .$$

These formulas will be quite useful to us in several of the following theorems.

Theorem 3.19 For any matrix A we have $r(A^T) = r(A)$.

Proof This is Exercise 3.6.2. ∎

Theorem 3.20 If A and B are any matrices for which the product AB is defined, then the row space of AB is a subspace of the row space of B, and the column space of AB is a subspace of the column space of A.

Proof Using $(AB)_i = \Sigma_k a_{ik} B_k$, we see that the *ith* row of AB is in the space spanned by the rows of B, and hence the row space of AB is a subspace of the row space of B.

Now note that the column space of AB is just the row space of $(AB)^T = B^T A^T$, which is a subspace of the row space of A^T by the first part of the theorem. But the row space of A^T is just the column space of A. ∎

Corollary $r(AB) \le \min\{r(A), r(B)\}$.

Proof Let V_A be the row space of A, and let W_A be the column space of A. Then

$$r(AB) = \dim V_{AB} \le \dim V_B = r(B)$$

while

$$r(AB) = \dim W_{AB} \le \dim W_A = r(A) \;. \;\blacksquare$$

Exercises

1. Complete the proof of Theorem 3.17.

2. Prove Theorem 3.19.

3. Let A be any m x n matrix and let X be any n x 1 matrix, both with entries in $\mathcal{F}$. Define the mapping $f : \mathcal{F}^n \to \mathcal{F}^m$ by f(X) = AX.
(a) Show that f is a linear transformation (i.e., a vector space homomorphism).
(b) Define Im f = $\{AX: X \in \mathcal{F}^n\}$. Show that Im f is a subspace of $\mathcal{F}^m$.
(c) Let U be the column space of A. Show that Im f = U. [*Hint*: Use Example 3.11 to show that Im f $\subset$ U. Next use the equation $(AI)^j = AI^j$ to show that U $\subset$ Im f.]
(d) Let N denote the solution space to the system AX = 0. In other words, $N = \{X \in \mathcal{F}^n: AX = 0\}$. (N is usually called the **null space** of A.) Show that

$$\dim N + \dim U = n \;.$$

[*Hint*: Suppose dim N = r, and extend a basis $\{x_1, \ldots, x_r\}$ for N to a basis $\{x_i\}$ for $\mathcal{F}^n$. Show that U is spanned by the vectors $Ax_{r+1}, \ldots,$ Ax_n, and then that these vectors are linearly independent. Note that this exercise is really just another proof of Theorem 3.13.]

4. A matrix of the form

$$\begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

is called a **diagonal** matrix. In other words, $A = (a_{ij})$ is diagonal if $a_{ij} = 0$ for $i \ne j$. If A and B are both square matrices, we may define the **commutator** [A, B] of A and B to be the matrix [A, B] = AB − BA. If [A, B] = 0, we say that A and B **commute**.
(a) Show that any diagonal matrices A and B commute.

(b) Prove that the only n x n matrices which commute with every n x n diagonal matrix are diagonal matrices.

5. Given the matrices
6.

$$A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \\ -3 & 4 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & -2 & -5 \\ 3 & 4 & 0 \end{pmatrix}$$

compute the following:
(a) AB.
(b) BA.
(c) AA^T.
(d) A^TA.
(e) Verify that $(AB)^T = B^TA^T$.

6. Consider the matrix $A \in M_n(\mathcal{F})$ given by

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} .$$

Thus A has zero entries everywhere except on the **superdiagonal** where the entries are 1's. Let $A^2 = AA$, $A^3 = AAA$, and so on. Show that $A^n = 0$ but $A^{n-1} \neq 0$.

7. Given a matrix $A = (a_{ij}) \in M_n(\mathcal{F})$, the sum of the diagonal elements of A is called the **trace** of A, and is denoted by Tr A. Thus

$$\operatorname{Tr} A = \sum_{i=1}^{n} a_{ii} \quad .$$

(a) Prove that Tr(A + B) = Tr A + Tr B and that $\operatorname{Tr}(\alpha A) = \alpha(\operatorname{Tr} A)$ for any scalar α.
(b) Prove that Tr(AB) = Tr(BA).

8. (a) Prove that it is impossible to find matrices $A, B \in M_n(\mathbb{R})$ such that their commutator $[A, B] = AB - BA$ is equal to 1.

(b) Let $\mathcal{F}$ be a field of characteristic 2 (i.e., a field in which $1 + 1 = 0$; see Exercise 1.5.17). Prove that it is possible to find matrices $A, B \in M_2(\mathcal{F})$ such that $[A, B] = 1$.

9. A matrix $A = (a_{ij})$ is said to be **upper-triangular** if $a_{ij} = 0$ for $i > j$. In other words, every entry of A below the main diagonal is zero. Similarly, A is said to be **lower-triangular** if $a_{ij} = 0$ for $i < j$. Prove that the product of upper (lower) triangular matrices is an upper (lower) triangular matrix.

10. Consider the so-called **Pauli spin matrices**

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and define the permutation symbol ε_{ijk} by

$$\varepsilon_{ijk} = \begin{cases} 1 & \text{if } (i,\ j,\ k) \text{ is an even permutation of } (1,\ 2,\ 3) \\ -1 & \text{if } (i,\ j,\ k) \text{ is an odd permutation of } (1,\ 2,\ 3) \\ 0 & \text{if any two indices are the same} \end{cases} .$$

The **commutator** of two matrices $A, B \in M_n(\mathcal{F})$ is defined by $[A, B] = AB - BA$, and the **anticommutator** is given by $[A, B]_+ = AB + BA$.

(a) Show that $[\sigma_i, \sigma_j] = 2i \sum_k \varepsilon_{ijk} \sigma_k$. In other words, show that $\sigma_i\sigma_j = i\sigma_k$ where (i, j, k) is an even permutation of $(1, 2, 3)$.
(b) Show that $[\sigma_i, \sigma_j]_+ = 2I\delta_{ij}$.
(c) Using part (a), show that $\operatorname{Tr} \sigma_i = 0$.
(d) For notational simplicity, define $\sigma_0 = I$. Show that $\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$ forms a basis for $M_2(\mathbb{C})$. [*Hint*: Show that $\operatorname{Tr}(\sigma_\alpha \sigma_\beta) = 2\delta_{\alpha\beta}$ where $0 \le \alpha, \beta \le 3$. Use this to show that $\{\sigma_\alpha\}$ is linearly independent.]
(e) According to part (d), any $X \in M_2(\mathbb{C})$ may be written in the form $X = \sum_\alpha x_\alpha \sigma_\alpha$. How would you find the coefficients x_α?
(f) Show that $\langle \sigma_\alpha, \sigma_\beta \rangle = (1/2)\operatorname{Tr}(\sigma_\alpha \sigma_\beta)$ defines an inner product on $M_2(\mathbb{C})$.
(g) Show that any matrix $X \in M_2(\mathbb{C})$ that commutes with all of the σ_i (i.e., $[X, \sigma_i] = 0$ for each $i = 1, 2, 3$) must be a multiple of the identity matrix.

11. A square matrix S is said to be **symmetric** if $S^T = S$, and a square matrix A is said to be **skewsymmetric** (or **antisymmetric**) if $A^T = -A$. (We continue to assume as usual that $\mathcal{F}$ is not of characteristic 2.)
    (a) Show that $S \neq 0$ and A are linearly independent in $M_n(\mathcal{F})$.
    (b) What is the dimension of the space of all n x n symmetric matrices?
    (c) What is the dimension of the space of all n x n antisymmetric matrices?

12. Find a basis $\{A_i\}$ for the space $M_n(\mathcal{F})$ that consists only of matrices with the property that $A_i^2 = A_i$ (such matrices are called **idempotent** or **projections**). [*Hint*: The matrices

$$\begin{pmatrix}1 & 0\\0 & 0\end{pmatrix} \quad \begin{pmatrix}1 & 1\\0 & 0\end{pmatrix} \quad \begin{pmatrix}0 & 0\\1 & 0\end{pmatrix} \quad \begin{pmatrix}0 & 0\\1 & 1\end{pmatrix}$$

will work in the particular case of $M_2(\mathcal{F})$.]

13. Show that it is impossible to find a basis for $M_n(\mathcal{F})$ such that every pair of matrices in the basis commute with each other.

14. (a) Show that the set of all nonsingular n x n matrices forms a spanning set for $M_n(\mathcal{F})$. Exhibit a basis of such matrices.
    (b) Repeat part (a) with the set of all singular matrices.

15. Show that the set of all matrices of the form AB − BA do not span $M_n(\mathcal{F})$. [*Hint*: Use the trace.]

16. Is it possible to span $M_n(\mathcal{F})$ using powers of a single matrix A? In other words, can $\{I_n, A, A^2, \ldots, A^n, \ldots\}$ span $M_n(\mathcal{F})$? [*Hint*: Consider Exercise 4 above.]

3.7 INVERTIBLE MATRICES

We say that a matrix $A \in M_n(\mathcal{F})$ is **nonsingular** if $r(A) = n$, and **singular** if $r(A) < n$. Given a matrix $A \in M_n(\mathcal{F})$, if there exists a matrix $B \in M_n(\mathcal{F})$ such that $AB = BA = I_n$, then B is called an **inverse** of A, and A is said to be **invertible**.

Technically, a matrix B is called a **left inverse** of A if $BA = I$, and a matrix B' is a **right inverse** of A if $AB' = I$. Then, if $AB = BA = I$, we say that B is a **two-sided inverse** of A, and A is then said to be **invertible**.

Furthermore, if A has a left inverse B and a right inverse B′, then it is easy to see that B = B′ since B = BI = B(AB′) = (BA)B′ = IB′ = B′. We shall now show that if B is either a left or a right inverse of A, then A is invertible.

Theorem 3.21 A matrix $A \in M_n(\mathcal{F})$ has a right (left) inverse if and only if A is nonsingular. This right (left) inverse is also a left (right) inverse, and hence is an inverse of A.

Proof Suppose A has a right inverse B. Then $AB = I_n$ so that $r(AB) = r(I_n)$. Since $r(I_n)$ is clearly equal to n (Theorem 3.9), we see that r(AB) = n. But then from the corollary to Theorem 3.20 and the fact that both A and B are n x n matrices (so that $r(A) \le n$ and $r(B) \le n$), it follows that r(A) = r(B) = n, and hence A is nonsingular.

Now suppose that A is nonsingular so that r(A) = n. If we let E^j be the *jth* column of the identity matrix I_n, then for each j = 1, . . . , n the system of equations

$$\sum_{i=1}^{n} A^i x_i = AX = E^j$$

has a unique solution which we denote by $X = B^j$ (Theorem 3.16). Now let B be the matrix with columns B^j. Then the *jth* column of AB is given by

$$(AB)^j = AB^j = E^j$$

and hence $AB = I_n$. It remains to be shown that $BA = I_n$. To see this, note that $r(A^T) = r(A) = n$ (Theorem 3.19) so that A^T is nonsingular also. Hence applying the same argument shows there exists a unique n x n matrix C^T such that $A^T C^T = I_n$. Since $(CA)^T = A^T C^T$ and ${I_n}^T = I_n$, this is the same as $CA = I_n$. We now recall that it was shown prior to the theorem that if A has both a left and a right inverse, then they are the same. Therefore B = C so that BA = AB $= I_n$, and hence B is an inverse of A. Clearly, the proof remains valid if “right” is replaced by “left” throughout. ∎

Corollary 1 A matrix $A \in M_n(\mathcal{F})$ is nonsingular if and only if it has an inverse. Furthermore, this inverse is unique.

Proof As we saw above, if B and C are both inverses of A, then B = BI = B(AC) = (BA)C = IC = C. ∎

In view of this corollary, the unique inverse to a matrix A will be denoted by A^{-1} from now on.

Corollary 2 If A is an n x n nonsingular matrix, then A^{-1} is nonsingular and $(A^{-1})^{-1} = A$.

Proof If A is nonsingular, then (by Theorem 3.21) A^{-1} exists so that $A^{-1}A = AA^{-1} = I$. But this means that $(A^{-1})^{-1}$ exists and is equal to A, and hence A^{-1} is also nonsingular. ∎

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

Proof The fact that A and B are nonsingular means that A^{-1} and B^{-1} exist. We therefore see that

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

and similarly $(AB)(B^{-1}A^{-1}) = I$. It then follows that $B^{-1}A^{-1} = (AB)^{-1}$. Since we have now shown that AB has an inverse, Theorem 3.21 tells us that AB must be nonsingular. ∎

Corollary 4 If A is nonsingular then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$.

Proof That A^T is nonsingular is a direct consequence of Theorem 3.19. Next we observe that

$$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$$

so that the uniqueness of the inverse tells us that $(A^T)^{-1} = (A^{-1})^T$. Note this also shows that A^T is nonsingular. ∎

Corollary 5 A system of n linear equations in n unknowns has a unique solution if and only if its matrix of coefficients is nonsingular.

Proof Consider the system AX = Y. If A is nonsingular, then a unique A^{-1} exists, and therefore we have $X = A^{-1}Y$ as the unique solution. (Note that this is essentially the content of Theorem 3.16.)

Conversely, if this system has a unique solution, then the solution space of the associated homogeneous system must have dimension 0 (Theorem 3.15). Then Theorem 3.13 shows that we must have r(A) = n, and hence A is nonsingular. ∎

A major problem that we have not yet discussed is how to actually find the inverse of a matrix. One method involves the use of determinants as we will see in the next chapter. However, let us show another approach based on the fact that a nonsingular matrix is row-equivalent to the identity matrix

(Theorem 3.10). This method has the advantage that it is algorithmic, and hence is easily implemented on a computer.

Since the j*th* column of a product AB is AB^j, we see that considering the particular case of $AA^{-1} = I$ leads to

$$(AA^{-1})^j = A(A^{-1})^j = E^j$$

where E^j is the j*th* column of I. What we now have is the nonhomogeneous system

$$AX = Y$$

(or $\sum_j a_{ij}x_j = y_i$) where $X = (A^{-1})^j$ and $Y = E^j$. As we saw in Section 3.2, we may solve for the vector X by reducing the augmented matrix to reduced row-echelon form. For the particular case of j = 1 we have

$$\text{aug } A = \begin{pmatrix} a_{11} & \cdots & a_{1n} & 1 \\ a_{21} & \cdots & a_{2n} & 0 \\ \vdots & & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & 0 \end{pmatrix}$$

and hence the reduced form will be

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & c_{11} \\ 0 & 1 & 0 & \cdots & 0 & c_{21} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & c_{n1} \end{pmatrix}$$

for some set of scalars c_{ij}. This means that the solution to the system is $x_1 = c_{11}$, $x_2 = c_{21}$, . . . , $x_n = c_{n1}$. But $X = (A^{-1})^1$ = the first column of A^{-1}, and therefore this last matrix may be written as

$$\begin{pmatrix} 1 & \cdots & 0 & a^{-1}{}_{11} \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & a^{-1}{}_{n1} \end{pmatrix} .$$

Now, for each j = 1, . . . , n the system $AX = A(A^{-1})^j = E^j$ always has the same matrix of coefficients, and only the last column of the augmented matrix depends on j. Since finding the reduced row-echelon form of the matrix of coefficients is independent of this last column, it follows that we may solve all n systems simultaneously by reducing the single matrix

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} & | & 1 & \cdots & 0 \\ \vdots & & \vdots & | & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} & | & 0 & \cdots & 1 \end{pmatrix}.$$

In other words, the reduced form will be

$$\begin{pmatrix} 1 & \cdots & 0 & | & a^{-1}{}_{11} & \cdots & a^{-1}{}_{1n} \\ \vdots & & \vdots & | & \vdots & & \vdots \\ 0 & \cdots & 1 & | & a^{-1}{}_{n1} & \cdots & a^{-1}{}_{nn} \end{pmatrix}$$

where the matrix $A^{-1} = (a^{-1}{}_{ij})$ satisfies $AA^{-1} = I$ since $(AA^{-1})^j = A(A^{-1})^j = E^j$ is satisfied for each $j = 1, \ldots, n$.

Example 3.13 Let us find the inverse of the matrix A given by

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

We leave it as an exercise for the reader to show that the reduced row-echelon form of

$$\begin{pmatrix} -1 & 2 & 1 & | & 1 & 0 & 0 \\ 0 & 3 & -2 & | & 0 & 1 & 0 \\ 2 & -1 & 0 & | & 0 & 0 & 1 \end{pmatrix}$$

is

$$\begin{pmatrix} 1 & 0 & 0 & | & 1/6 & 1/12 & 7/12 \\ 0 & 1 & 0 & | & 1/3 & 1/6 & 1/6 \\ 0 & 0 & 1 & | & 1/2 & -1/4 & 1/4 \end{pmatrix}$$

and hence A^{-1} is given by

$$\begin{pmatrix} 1/6 & 1/12 & 7/12 \\ 1/3 & 1/6 & 1/6 \\ 1/2 & -1/4 & 1/4 \end{pmatrix}. \quad //$$

Exercises

1. Verify the reduced row-echelon form of the matrix given in Example 3.13.

2. Find the inverse of a general 2 x 2 matrix. What constraints are there on the entries of the matrix?

3. Show that a matrix is not invertible if it has any zero row or column.

4. Find the inverse of each of the following matrices:

$$(a)\ \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{pmatrix} \qquad (b)\ \begin{pmatrix} 1 & 3 & 4 \\ 3 & -1 & 6 \\ -1 & 5 & 1 \end{pmatrix} \qquad (c)\ \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 3 & 3 \end{pmatrix}$$

5. Use the inverse of the matrix in Exercise 4(c) above to find the solutions of each of the following systems:

$$(a)\ \begin{aligned} x + 2y + z &= 10 \\ 2x + 5y + 2z &= 14 \\ x + 3y + 3z &= 30 \end{aligned} \qquad (b)\ \begin{aligned} x + 2y + z &= 2 \\ 2x + 5y + 2z &= -1 \\ x + 3y + 3z &= 6 \end{aligned}$$

6. What is the inverse of a diagonal matrix?

7. (a) Prove that an upper-triangular matrix is invertible if and only if every entry on the main diagonal is nonzero (see Exercise 3.6.9 for the definition of an upper-triangular matrix).
(b) Prove that the inverse of a lower (upper) triangular matrix is lower (upper) triangular.

8. Find the inverse of the following matrix:

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

9. (a) Let A be any 2 x 1 matrix, and let B be any 1 x 2 matrix. Prove that AB is not invertible.

(b) Repeat part (a) where A is any m x n matrix and B is any n x m matrix with $n < m$.

10. Summarize several of our results by proving the equivalence of the following statements for any n x n matrix A:
(a) A is invertible.
(b) The homogeneous system $AX = 0$ has only the zero solution.
(c) The system $AX = Y$ has a solution X for every n x 1 matrix Y.

11. Let A and B be square matrices of size n, and assume that A is nonsingular. Prove that $r(AB) = r(B) = r(BA)$.

12. A matrix A is called a **left zero divisor** if there exists a nonzero matrix B such that $AB = 0$, and A is called a **right zero divisor** if there exists a nonzero matrix C such that $CA = 0$. If A is an m x n matrix, prove that:
(a) If $m < n$, then A is a left zero divisor.
(b) If $m > n$, then A is a right zero divisor.
(c) If $m = n$, then A is both a left and a right zero divisor if and only if A is singular.

13. Let A and B be nonsingular symmetric matrices for which $AB - BA = 0$. Show that AB, $A^{-1}B$, AB^{-1} and $A^{-1}B^{-1}$ are all symmetric.

3.8 ELEMENTARY MATRICES

Recall the elementary row operations α, β, γ described in Section 3.2. We now let e denote any one of these three operations, and for any matrix A we define e(A) to be the result of applying the operation e to the matrix A. In particular, we define an **elementary matrix** to be any matrix of the form e(I). The great utility of elementary matrices arises from the following theorem.

Theorem 3.22 If A is any m x n matrix and e is any elementary row operation, then

$$e(A) = e(I_m)A \ .$$

Proof We must verify this equation for each of the three types of elementary row operations. First consider an operation of type α. In particular, let α be the interchange of rows i and j. Then

$$[e(A)]_k = A_k \quad \text{for } k \neq i, j$$

while

$$[e(A)]_i = A_j \quad \text{and} \quad [e(A)]_j = A_i \ .$$

On the other hand, using $(AB)_k = A_k B$ we also have

$$[e(I)A]_k = [e(I)]_k A \ .$$

If $k \neq i, j$ then $[e(I)]_k = I_k$ so that

$$[e(I)]_k A = I_k A = A_k \ .$$

If $k = i$, then $[e(I)]_i = I_j$ and

$$[e(I)]_i A = I_j A = A_j \ .$$

Similarly, we see that

$$[e(I)]_j A = I_i A = A_i \ .$$

This verifies the theorem for transformations of type α. (It may be helpful for the reader to write out e(I) and e(I)A to see exactly what is going on.)

There is essentially nothing to prove for type β transformations, so we go on to those of type γ. Hence, let e be the addition of c times row j to row i. Then

$$[e(I)]_k = I_k \quad \text{for } k \neq i$$

and

$$[e(I)]_i = I_i + cI_j \ .$$

Therefore

$$[e(I)]_i A = (I_i + cI_j)A = A_i + cA_j = [e(A)]_i$$

and for $k \neq i$ we have

$$[e(I)]_k A = I_k A = A_k = [e(A)]_k \ . \ \blacksquare$$

If e is of type α, then rows i and j are interchanged. But this is readily undone by interchanging the same rows again, and hence e^{-1} is defined and is another elementary row operation. For type β operations, some row is multiplied by a scalar c, so in this case e^{-1} is simply multiplication by 1/c. Finally, a type γ operation adds c times row j to row i, and hence e^{-1} adds $-c$ times row j to row i. Thus all three types of elementary row operations have inverses which are also elementary row operations.

By way of nomenclature, a square matrix $A = (a_{ij})$ is said to be **diagonal** if $a_{ij} = 0$ for $i \neq j$. The most common example of a diagonal matrix is the identity matrix.

Theorem 3.23 Every elementary matrix is nonsingular, and

$$[e(I)]^{-1} = e^{-1}(I) .$$

Furthermore, the transpose of an elementary matrix is an elementary matrix.

Proof By definition, e(I) is row equivalent to I and hence has the same rank as I (Theorem 3.4). Thus e(I) is nonsingular since $r(I_n) = n$, and hence $e(I)^{-1}$ exists. Since it was shown above that e^{-1} is an elementary row operation, we apply Theorem 3.22 to the matrix e(I) to obtain

$$e^{-1}(I)e(I) = e^{-1}(e(I)) = I .$$

Similarly, applying Theorem 3.22 to $e^{-1}(I)$ yields

$$e(I)e^{-1}(I) = e(e^{-1}(I)) = I .$$

This shows that $e^{-1}(I) = [e(I)]^{-1}$.

Now let e be a type α transformation that interchanges rows i and j (with $i < j$). Then the *ith* row of e(I) has a 1 in the *jth* column, and the *jth* row has a 1 in the *ith* column. In other words,

$$[e(I)]_{ij} = 1 = [e(I)]_{ji}$$

while for $r, s \neq i, j$ we have

$$[e(I)]_{rs} = 0 \quad \text{if } r \neq s$$

and

$$[e(I)]_{rr} = 1 .$$

Taking the transpose shows that

$$[e(I)]^T{}_{ij} = [e(I)]_{ji} = 1 = [e(I)]_{ij}$$

and

$$[e(I)]^T{}_{rs} = [e(I)]_{sr} = 0 = [e(I)]_{rs} .$$

Thus $[e(I)]^T = e(I)$ for type α operations.

Since I is a diagonal matrix, it is clear that for a type β operation which simply multiplies one row by a nonzero scalar, we have $[e(I)]^T = e(I)$.

Finally, let e be a type γ operation that adds c times row j to row i. Then e(I) is just I with the additional entry $[e(I)]_{ij} = c$, and hence $[e(I)]^T$ is just I with the additional entry $[e(I)]_{ji} = c$. But this is the same as c times row i added to row j in the matrix I. In other words, $[e(I)]^T$ is just another elementary matrix. ∎

We now come to the main result dealing with elementary matrices. For ease of notation, we denote an elementary matrix by E rather than by e(I). In other words, the result of applying the elementary row operation e_i to I will be denoted by the matrix $E_i = e_i(I)$.

Theorem 3.24 Every nonsingular n x n matrix may be written as a product of elementary n x n matrices.

Proof It follows from Theorem 3.10 that any nonsingular n x n matrix A is row equivalent to I_n. This means that I_n may be obtained by applying r successive elementary row operations to A. Hence applying Theorem 3.22 r times yields

$$E_r \cdots E_1 A = I_n$$

so that

$$A = E_1^{-1} \cdots E_r^{-1} I_n = E_1^{-1} \cdots E_r^{-1} \,.$$

The theorem now follows if we note that each E_i^{-1} is an elementary matrix according to Theorem 3.23 (since $E_i^{-1} = [e(I)]^{-1} = e^{-1}(I)$ and e^{-1} is an elementary row operation). ∎

Corollary If A is an invertible n x n matrix, and if some sequence of elementary row operations reduces A to the identity matrix, then the same sequence of row operations reduces the identity matrix to A^{-1}.

Proof By hypothesis we may write $E_r \cdots E_1 A = I$. But then multiplying from the right by A^{-1} shows that $A^{-1} = E_r \cdots E_1 I$. ∎

Note this corollary provides another proof that the method given in the previous section for finding A^{-1} is valid.

There is one final important property of elementary matrices that we will need in a later chapter. Let E be an n x n elementary matrix representing any of the three types of elementary row operations, and let A be an n x n matrix. As we have seen, multiplying A from the left by E results in a new matrix with the same rows that would result from applying the elementary row operation to A directly. We claim that multiplying A from the right by E^T results in a new matrix whose columns have the same relationship as the rows of EA. We will prove this for a type γ operation, leaving the easier type α and β operations to the reader (see Exercise 3.8.1).

Let γ be the addition of c times row j to row i. Then the rows of E are given by $E_k = I_k$ for $k \neq i$, and $E_i = I_i + cI_j$. Therefore the columns of E^T are given by

$$(E^T)^k = I^k \quad \text{for } k \neq i$$

and

$$(E^T)^i = I^i + cI^j \ .$$

Now recall that the k*th* column of AB is given by $(AB)^k = AB^k$. We then have

$$(AE^T)^k = A(E^T)^k = AI^k = A^k \quad \text{for } k \neq i$$

and

$$(AE^T)^i = A(E^T)^i = A(I^i + cI^j) = AI^i + cAI^j = A^i + cA^j \ .$$

This is the same relationship as that found between the rows of EA where $(EA)_k = A_k$ and $(EA)_i = A_i + cA_j$ (see the proof of Theorem 3.22).

Exercises

1. Let A be an n x n matrix, and let E be an n x n elementary matrix representing a type α or β operation. Show that the columns of AE^T have the same relationship as the rows of EA.

2. Write down 4 x 4 elementary matrices that will induce the following elementary operations in a 4 x 4 matrix when used as left multipliers. Verify that your answers are correct.
   (a) Interchange the 2*nd* and 4*th* rows of A.
   (b) Interchange the 2*nd* and 3*rd* rows of A.
   (c) Multiply the 4*th* row of A by 5.
   (d) Add k times the 4*th* row of A to the 1*st* row of A.
   (e) Add k times the 1*st* row of A to the 4*th* row of A.

3. Show that any $e_\alpha(A)$ may be written as a product of $e_\beta(A)$'s and $e_\gamma(A)$'s. (The notation should be obvious.)

4. Pick any 4 x 4 matrix A and multiply it from the *right* by each of the elementary matrices found in the previous problem. What is the effect on A?

5. Prove that a matrix A is row equivalent to a matrix B if and only if there exists a nonsingular matrix P such that B = PA.

6. Reduce the matrix

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

to the reduced row-echelon form R, and write the elementary matrix corresponding to each of the elementary row operations required. Find a nonsingular matrix P such that PA = R by taking the product of these elementary matrices.

7. Let A be an n x n matrix. Summarize several of our results by proving that the following are equivalent:
(a) A is invertible.
(b) A is row equivalent to I_n .
(c) A is a product of elementary matrices.

8. Using the results of the previous problem, prove that if $A = A_1 A_2 \cdots A_k$ where each A_i is a square matrix, then A is invertible if and only if each of the A_i is invertible.

The remaining problems are all connected, and should be worked in the given order.

9. Suppose that we define elementary column operations exactly as we did for rows. Prove that every elementary column operation on A can be achieved by multiplying A on the *right* by an elementary matrix. [*Hint*: You can either do this directly as we did for rows, or by taking transposes and using Theorem 3.23.]

10. Show that an m x n reduced row-echelon matrix R of rank k can be reduced by elementary column operations to an m x n matrix C of the form

$$C = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \end{pmatrix}$$

where the first k entries on the main diagonal are 1's, and the rest are 0's.

11. From the previous problem and Theorem 3.3, show that every m x n matrix A of rank k can be reduced by elementary row and column operations to the form C. We call the matrix C the **canonical form** of A.

12. We say that a matrix A is **row-column-equivalent** (abbreviated by r.c.e.) to a matrix B if A can be transformed into B by a finite number of elementary row and column operations. Prove:
(a) If A is a matrix, e is an elementary row operation, and e′ is an elementary column operation, then (eA)e′ = e(Ae′).
(b) r.c.e. is an equivalence relation.
(c) Two m x n matrices A and B are r.c.e. if and only if they have the same canonical form, and hence if and only if they have the same rank.

13. If A is any m x n matrix of rank k, prove that there exists a nonsingular m x m matrix P and a nonsingular n x n matrix Q such that PAQ = C (the canonical form of A).

14. Prove that two m x n matrices A and B are r.c.e. if and only if there exists a nonsingular m x m matrix P and a nonsingular n x n matrix Q such that PAQ = B.