Chapter 2

Linear Equations and Matrices

2.1 Linear equations: the beginning of algebra

The subject of algebra arose from the study of equations. The simplest kind of equations are linear equations, which are equations of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = c,$$

where $a_1, a_2, \ldots a_n$ are a set of numbers called the coefficients, $x_1, x_2, \ldots x_n$ are the variables and c is the constant term. In most familiar situations, the coefficients are real numbers, but in some of the other settings we will encounter later, such as coding theory, the coefficients might be elements of some a finite field. Such considerations will be taken up in later chapters.

The simplest linear equation one can imagine is an equation with only one variable, such as $ax = b$. For example, consider $3x = 4$. This equation is easy to solve since we can express the solution as $x = 3/4$. In general, if $a \neq 0$, then $x = \frac{b}{a}$, and this is the only solution. But if $a = 0$ and $b \neq 0$, there is no solution, since the equation is $0 = b$. And in the case where a and b are both 0, every real number x is a solution. This points outs a general property of linear equations. Either there is a unique solution (i.e. exactly one), no solution or infinitely many solutions.

Let's take another example. Suppose you are planning to make a cake using 10 ingredients and you want to limit the cake to 2000 calories. Let a_i be the number of calories per gram of the ith ingredient. Presumably, each a_i is nonnegative, although this problem may eventually be dealt with.

Next, let x_i be the number of grams of the ith ingredient. Then $a_1x_1+a_2x_2+\cdots+a_{10}x_{10}$ is the total number of calories in the recipe. Since you want the total number of calories in your cake to be at most 2000, you could consider the equation $a_1x_1 + a_2x_2 + \cdots + a_{10}x_{10} = 2000$. The totality of possible solutions $x_1, x_2, \ldots, x_{10}$ to this equation is the set of all possible recipes you can concoct with exactly 2000 calories. Decreasing the amount of any ingredient will then clearly decrease the total number of calories. Of course, any solution where some x_i is negative don't have a physical meaning.

A less simple example is the question of finding all common solutions of the equations $z = x^2 + xy^5$ and $z^2 = x + y^4$. Since the equations represent two surfaces in $\mathbb{R}^3$, we would expect the set of common solutions to be a curve. It's impossible to express the solutions in closed form, but we can study them locally. For example, both surfaces meet at $(1,1,1)^T$, so we can find the tangent line to the curve of intersection at $(1,1,1)^T$ by finding the intersection of the tangent planes of the surfaces at this point. This will at least give us a linear approximation to the curve.

General, nonlinear systems are usually very difficult to solve; their theory involves highly sophisticated mathematics. On the other hand, it turns out that systems of linear equations can be handled much more simply. There are elementary methods for solving them, and modern computers make it possible to handle gigantic linear systems with great speed. A general linear system having of m equations in n unknowns $x_1, \ldots, x_n$ can be expressed in the following form:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{23}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \tag{2.1}$$

Here, all the coefficients a_{ij} and all constants b_i are assumed for now to be real numbers. When all the constants $b_i = 0$, we will call the system *homogeneous*. The main problem, of course, is to find a procedure or algorithm for describing the *solution set* of a linear system as a subset of $\mathbb{R}^n$.

For those who skipped Chapter 1, let us insert a word about notation. A solution of (2.1) is an n-tuple of real numbers, i. e. an element of $\mathbb{R}^n$. By convention, n-tuples are always written as column vectors. To save space, we will use the notation

$$(u_1, u_2, \ldots, u_n)^T = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

The meaning of the superscript ^{T} will be clarified below. We would also like to point out that a brief summary of the highlights of this chapter may be found in the last section.

2.1.1 The Coefficient Matrix

To simplify notation, we will introduce the coefficient matrix.

Definition 2.1. The *coefficient matrix* of the above linear system is the $m \times n$ array

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}. \tag{2.2}$$

The *augmented coefficient matrix* is the $m \times (n+1)$ array

$$(A|\mathbf{b}) = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{23} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ldots & \vdots & \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{pmatrix}. \tag{2.3}$$

In general, an $m \times n$ *matrix* is simply a rectangular $m \times n$ array as in (2.2). When $m = n$, we will say that A is a square matrix of *degree n.*

Now let's look at the strategy for finding solving the system. First of all, we will call the set of solutions the *solution set.* The strategy for finding the solution set is to replace the original system with a sequence of new systems so that each new system has the same solution set as the previous one, hence as the original system.

Definition 2.2. Two linear systems are said to be *equivalent* if they have the same solution sets.

Two equivalent systems have the same number of variables, but don't need to have the same number of equations.

2.1.2 Gaussian reduction

The procedure for solving an arbitrary system is called *Gaussian reduction.* Gaussian reduction is an algorithm for solving an arbitrary system by performing a sequence of explicit operations, called *elementary row operations*, to bring the augmented coefficient matrix $(A|\mathbf{b})$ in (2.3) to a form called *reduced form*, or *reduced row echelon form.* First of all, we define reduced row echelon form.

Definition 2.3. A matrix A is said to be in *reduced row echelon form*, or simply, to be *reduced*, if it has three properties.

(i) The first non zero entry in each row of A is 1.

(ii) The first non zero entry in every row is to the right of the first non zero entry in all the rows above it.

(iii) Every entry above a first non zero entry is zero.

We will call a first non zero entry in a row its *corner entry.* A first non zero entry in a row which has not been made into 1 by a dilation is called the *pivot* of the row. Pivots aren't required to be 1.

For reasons that will be explained later, an $n \times n$ matrix in reduced row echelon form is called the *$n \times n$ identity matrix.* For example,

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Here are some more examples of reduced matrices:

$$\begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 5 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 & 0 & 9 \\ 0 & 0 & 0 & 1 & 4 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 3 & 0 & 9 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Notice that the last matrix in this example would be the coefficient matrix of a system in which the variable x_1 does not actually appear. The only variables that occur are $x_2, \dots, x_5$. Note also that the $n \times n$ identity matrix I_n and every matrix of zeros are also reduced.

2.1.3 Elementary row operations

The strategy in Gaussian reduction is to use a sequence of steps called *elementary row operations* on the rows of the coefficient matrix A to bring A into reduced form. There are three types of elementary row operations defined as follows:

- (Type I) Interchange two rows of A.
- (Type II) Multiply a row of A by a non zero scalar.
- (Type III) Replace a row of A by itself plus a multiple of a different row.

We will call Type I operations *row swaps* and Type II operations *row dilations.* Type III operations are called *transvections.* We will boycott this term. The main result is that an arbitrary matrix A can always be put into reduced form by a sequence of row operations. Before proving this, we will work an example.

Example 2.1. Consider the counting matrix

$$C = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

We can row reduce C as follows:

$$C \overset{R_2-4R_1}{\longrightarrow} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{pmatrix} \overset{R_3-7R_1}{\longrightarrow} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{pmatrix}$$

$$\overset{R_3-2R_2}{\longrightarrow} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{pmatrix} \overset{(-1/3)R_2}{\longrightarrow} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \overset{R_1-2R_2}{\longrightarrow} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Notice that we have indicated the row operations.

Proposition 2.1. *Every matrix A can be put into reduced form by some (not unique) sequence of elementary row operations.*

Proof. If $a_{11} \neq 0$, we can make it 1 by the dilation which divides the first row by a_{11}. We can then use row operations to make all other entries in the first column zero. If $a_{11} = 0$, but the first column has a non zero entry somewhere, suppose the first non zero entry is in the ith row. Then swapping the first and ith rows puts a non zero entry in the $(1,1)$ position. Now proceed as above, dividing the new first row by the inverse of the new $(1,1)$ entry, getting a corner entry in the $(1,1)$ position. Now that there we have a corner entry in $(1,1)$, we can use operations of type III to make all elements in the first column below the $(1,1)$ entry 0. If the first column consists entirely of zeros, proceed directly to the second column and repeat the procedure just described on the second column. The only additional step is that if the $(2,2)$ entry is a corner, then the $(1,2)$ entry may be made into 0 by another Type III operation. Continuing in this manner, we will eventually obtain a reduced matrix. □

REMARK: Of course, the steps leading to a reduced form are not unique. Nevertheless, the reduced form of A itself is unique. We now make an important definition.

Definition 2.4. The reduced form of an $m \times n$ matrix A is denoted by A_{red}. The *row rank*, or simply, the *rank* of an $m \times n$ matrix A is the number of non-zero rows in A_{red}.

2.2 Solving Linear Systems

Let A be an $m \times n$ matrix and consider the linear system whose augmented coefficient matrix is $(A|\mathbf{b})$. The first thing is to point out the role of row operations.

2.2.1 Equivalent Systems

Recall that two linear systems are said to be equivalent if they have the same solution sets. The point about Gaussian reduction is that performing a row operation on the augmented coefficient matrix of a linear system gives a new system which is equivalent to the original system.

For example, a row swap, which corresponds to interchanging two equations, clearly leaves the solution set unchanged. Similarly, multiplying the ith equation by a non-zero constant a does likewise, since the original system can be recaptured by multiplying the ith equation by a^{-1}. The only question is whether a row operation of Type III changes the solutions. Suppose the ith equation is replaced by itself plus a multiple k of the jth equation, where $i \neq j$. Then any solution of the original system is still a solution of the new system. But any solution of the new system is also a solution of the original system since subtracting k times the jth equation from the ith equation of the new system gives us back the original system. Therefore the systems are equivalent.

To summarize this, we state

Proposition 2.2. *Performing a sequence of row operations on the augmented coefficient matrix of a linear system gives a new system which is equivalent to the original system.*

2.2.2 The Homogeneous Case

We still have to find a way to write down the solution set. The first step will be to consider the homogeneous linear system with coefficient matrix $(A|\mathbf{0})$. This is the same as dealing with the coefficient matrix A all by itself.

Definition 2.5. The solution set of a homogeneous linear system with coefficient matrix A is denoted by $\mathcal{N}(A)$ and called the *null space* of A.

The method is illustrated by the following example.

Example 2.2. Consider the homogeneous linear system

$$\begin{aligned} 0x_1 + x_2 + 2x_3 + 0x_4 + 3x + 5 - x_6 &= 0 \\ 0x_1 + 0x_2 + 0x_3 + x_4 + 2x_5 + 0x_6 &= 0. \end{aligned}$$

The coefficient matrix$A$ is already reduced. Indeed,

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

Our procedure will be to solve for the variables in the corners, which we will call the *corner variables.* We will express these variables in terms of the remaining variables, which we will call the *free variables.* In A above, the corner columns are the second and fourth, so x_2 and x_4 are the corner variables and the variables x_1, x_3, x_5 and x_6 are the free variables. Solving gives

$$\begin{aligned} x_2 &= -2x_3 - 3x_5 + x_6 \\ x_4 &= -2x_5 \end{aligned}$$

In this expression, the corner variables are dependent variables since they are functions of the free variables. Now let $(x_1, x_2, x_3, x_4, x_5, x_6)$ denoted an arbitrary vector in $\mathbb{R}^6$ which is a solution to the system, and let us call this 6-tuple *the general solution vector*. Replacing the corner variables by their expressions in terms of the free variables gives a new expression for the general solution vector involving just the free variables. Namely

$$\mathbf{x} = (x_1, -2x_3 - 3x_5 + x_6,\ x_3,\ -2x_5,\ x_5,\ x_6)^T.$$

The general solution vector now depends only on the free variables, and there is a solution for any choice of these variables.

Using a little algebra, we can compute the vector coefficients of each one of the free variables in $\mathbf{x}$. These vectors are called the *fundamental solutions*. In this example, the general solution vector $\mathbf{x}$ gives the following set of fundamental solutions:

$$\mathbf{f}_1 = (1,0,0,0,0,0)^T,\ \mathbf{f}_2 = (0,0,-2,\ 1,\ 0,\ 0)^T,\ \mathbf{f}_3 = (0,-3,\ 0,\ -2,\ 1,\ 0)^T,$$

and

$$\mathbf{f}_4 = (0,-1,\ 0,\ 0,\ 0,1)^T.$$

Hence the general solution vector has the form

$$\mathbf{x} = x_1\mathbf{f}_1 + x_3\mathbf{f}_2 + x_4\mathbf{f}_3 + x_5\mathbf{f}_4.$$

In other words, the fundamental solutions span the solution space, i.e. every solution is a linear combination of the fundamental solutions.

This example suggest the following

Proposition 2.3. *In an arbitrary homogeneous linear system with coefficient matrix A, any solution is a linear combination of the fundamental solutions, and the number of fundamental solutions is the number of free variables. Moreover,*

$$\#\text{corner variables} + \#\text{free variables} = \#\text{variables}. \tag{2.4}$$

Proof. The proof that every solution is a linear combination of the fundamental solutions goes exactly like the above example, so we will omit it. Equation (2.4) is an obvious consequence of the fact that every variable is either a free variable or a corner variable, but not both. □

There is something strange in Example 2.2. The variable x_1 never actually appears in the system, but it does give a free variable and a corresponding fundamental solution $(1,0,0,0,0,0)^T$. Suppose instead of A the coefficient matrix is

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

Now $(1, 0, 0, 0, 0, 0)^T$ is no longer a fundamental solution. In fact the solution set is now a subset of $\mathbb{R}^5$. The corner variables are x_1 and x_3, and there are now only three fundamental solutions corresponding to the free variables x_2, x_4, and x_5.

Even though (2.4) is completely obvious, it gives some very useful information. Here is a typical application.

Example 2.3. Consider a system involving 25 variables and assume there are 10 free variables. Then there are 15 corner variables, so the system has to have at least 15 equations, that is, there must be at least 15 linear constraints on the 25 variables.

We can also use (2.4) to say when a homogeneous system with coefficient matrix A has a unique solution (that is, exactly one solution). Now $\mathbf{0}$ is always a solution. This is the so called *trivial solution.* Hence if the solution is to unique, then the only possibility is that $\mathcal{N}(A) = \{\mathbf{0}\}$. But this happens exactly when there are no free variables, since if there is a free variable there will be non trivial solutions. Thus a homogeneous system has a unique solution if and only if every variable is a corner variable, which is the case exactly when the number of corner variables is the number of columns of A. It follows that if a homogeneous system has more variables than equations, there have to be non trivial solutions, since there has to be at least one free variable.

2.2.3 The Non-homogeneous Case

Next, consider the system with augmented coefficient matrix $(A|\mathbf{b})$. If $\mathbf{b} \neq \mathbf{0}$, the system is called *non-homogeneous.* To resolve the non-homogeneous case, we need to observe a result sometimes called the *Super-Position Principle.*

Proposition 2.4. *If a system with augmented coefficient matrix $(A|\mathbf{b})$ has a particular solution $\mathbf{p}$, then any other solution has the form $\mathbf{p} + \mathbf{x}$, where $\mathbf{x}$ varies over all solutions of the associated homogeneous equation. That is, $\mathbf{x}$ varies over $\mathcal{N}(A)$.*

Proof. Let us sketch the proof. (It is quite easy.) Suppose $\mathbf{p} = (p_1, \dots, p_n)$ and let $\mathbf{x} = (x_1, \dots, x_n)$ be an element of $\mathcal{N}(A)$. Then substituting $p_i + x_i$ into the system also gives a solution. Conversely, if $\mathbf{q}$ is another particular solution, then $\mathbf{p} - \mathbf{q}$ is a solution to the homogeneous system, i.e. an element of $\mathcal{N}(A)$. Therefore $\mathbf{q} = \mathbf{p} + \mathbf{x}$, where $\mathbf{x} = \mathbf{q} - \mathbf{p} \in \mathcal{N}(A)$. This completes the proof. □

Example 2.4. Consider the system involving the counting matrix C of Example 2.1:

$$\begin{aligned} 1x_1 + 2x_2 + 3x_3 &= a \\ 4x_1 + 5x_2 + 6x_3 &= b \\ 7x_1 + 8x_2 + 9x_3 &= c, \end{aligned}$$

where a, b and c are fixed arbitrary constants. This system has augmented coefficient matrix

$$(C|\mathbf{b}) = \begin{pmatrix} 1 & 2 & 3 & a \\ 4 & 5 & 6 & b \\ 7 & 8 & 9 & c \end{pmatrix}.$$

We can use the same sequence of row operations as in Example 2.1 to put $(C|\mathbf{b})$ into reduced form $(C_{red}|\mathbf{c})$ but to minimize the arithmetic with denominators, we will actually use a different sequence.

$$(C|\mathbf{b}) = \overset{R_2-R_1}{\rightarrow} \begin{pmatrix} 1 & 2 & 3 & a \\ 3 & 3 & 3 & b-a \\ 7 & 8 & 9 & c \end{pmatrix} \overset{R_3-2R_2}{\rightarrow} \begin{pmatrix} 1 & 2 & 3 & a \\ 3 & 3 & 3 & b-a \\ 1 & 2 & 3 & c-2b+2a \end{pmatrix} \overset{R_3-R_1}{\rightarrow}$$

$$\begin{pmatrix} 1 & 2 & 3 & a \\ 3 & 3 & 3 & b-a \\ 0 & 0 & 0 & c-2b+a \end{pmatrix} \overset{(-1/3)R_3}{\rightarrow} \begin{pmatrix} 1 & 2 & 3 & a \\ -1 & -1 & -1 & (1/3)a-(1/3)b \\ 0 & 0 & 0 & c-2b+a \end{pmatrix} \overset{R_2+R_1}{\rightarrow}$$

$$\begin{pmatrix} 1 & 2 & 3 & a \\ 0 & 1 & 2 & (4/3)a-(1/3)b \\ 0 & 0 & 0 & c-2b+a \end{pmatrix} \overset{R_1-2R_2}{\rightarrow} \begin{pmatrix} 1 & 0 & -1 & (-5/3)a+(2/3)b \\ 0 & 1 & 2 & (4/3)a-(1/3)b \\ 0 & 0 & 0 & c-2b+a \end{pmatrix}.$$

The reduced system turns out to be the same one we obtained by using the sequence in Example 11.2. We get

$$\begin{aligned} 1x_1 + 0x_2 - 1x_3 &= (-5/3)a + (2/3)b \\ 0x_1 + 1x_2 + 2x_3 &= (4/3)a - (1/3)b \\ 0x_1 + 0x_2 + 0x_3 &= a - 2b + c \end{aligned}$$

Clearly the above system may in fact have no solutions. In fact, from the last equation, we see that whenever $a - 2b + c \neq 0$, there cannot be a solution. Such a system is called *inconsistent.* For a simpler, example, think of three lines in $\mathbb{R}^2$ which don't pass through a common point. This is an example where the system has three equations but only two variables.

Example 2.5. Let's solve the system of Example 2.4 for $a = 1$, $b = 1$ and $c = 1$. In that case, the original system is equivalent to

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

It follows that $x_1 = -1 + x_3$ and $x_2 = 1 - 2x_3$. This represents a line in $\mathbb{R}^3$.

The line if the previous example is parallel to the line of intersection of the three planes

$$\begin{aligned} 1x_1 + 2x_2 + 3x_3 &= 0 \\ 4x_1 + 5x_2 + 6x_3 &= 0 \\ 7x_1 + 8x_2 + 9x_3 &= 0, \end{aligned}$$

These planes meet in a line since their normals are contained in a plane through the origin. On the other hand, when $a - 2b + c \neq 0$, what happens is that the line of intersection of any two of the planes is parallel to the third plane (and not contained in it).

that slightly perturbing the lines will

2.2.4 Criteria for Consistency and Uniqueness

To finish our treatment of systems (for now), we derive two criteria, one for consistency and the other for uniqueness. Consider the $m \times n$ linear system (2.1) with coefficient matrix A and augmented coefficient matrix $(A|\mathbf{b})$.

Proposition 2.5. *Suppose the coefficient matrix A has rank k, that is A_{red} has k corners. Then $\mathcal{N}(A) = \{\mathbf{0}\}$ if and only if $k = n$. The (possibly non-homogeneous) linear system $(A|\mathbf{b})$ is consistent if and only if the rank of A and of $(A|\mathbf{b})$ coincide. If $(A|\mathbf{b})$ is consistent and $k = n$, then the solution is unique. Finally, if A is $n \times n$, the system* (2.1) *is consistent for all* $\mathbf{b}$ *if and only if the rank of A equals n.*

Proof. The first statement is a repetition of a result we already proved. The second follows as in the previous example, because if the rank of $(A|\mathbf{b})$ is greater than k, then the last equation amounts to saying $0 = 1$. If A is $n \times n$ of rank n, then it is clear that $(A|\mathbf{b})$ and A have the same rank, namely n. It remains to show that if A and $(A|\mathbf{b})$ have the same rank for all $\mathbf{b}$, then A has rank n. But if the rank of A is less than n, one can (exactly as in

Example 2.4) produce a $\mathbf{b}$ for which $(A|\mathbf{b})$ has rank greater than the rank of A. □

Systems where $m = n$ are an important special case as they are neither under determined (fewer equations than unknowns) nor over determined (more equations than unknowns). When A is $n \times n$ of rank n, the system (2.1) is called *nonsingular.* Thus the nonsingular systems are the square systems which are always consistent and always have unique solutions. We also say that an $n \times n$ matrix *nonsingular* if it has maximal rank n. If the rank of A is less than n, we say that A is *singular.*

Example 2.6. An amusing geometric criterion for a 3×3 matrix

$$A = \begin{pmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \mathbf{a_3} \end{pmatrix}$$

to be nonsingular is that

$$\mathbf{a_1} \cdot (\mathbf{a_2} \times \mathbf{a_3}) \neq 0.$$

Indeed, we know that $\mathbf{a_1}$, $\mathbf{a_2}$, and $\mathbf{a_3}$ are not in a plane through the origin if and only if $\mathbf{a_1} \cdot (\mathbf{a_2} \times \mathbf{a_3}) \neq 0$. But the above Proposition also says that the rank of A is three precisely when there is no non-zero vector orthogonal to each of $\mathbf{a_1}$, $\mathbf{a_2}$, and $\mathbf{a_3}$.

The expression $\mathbf{a_1} \cdot (\mathbf{a_2} \times \mathbf{a_3})$ is called the determinant of A and abbreviated $\det(A)$. In algebraic terms, we have

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \tag{2.5}$$

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}.$$

Determinants will be taken up in a later chapter.

Exercises

Exercise 2.1. Consider the linear system

$$\begin{aligned} x_1 + 2x_2 + 4x_3 + 4x_4 &= 7 \\ x_2 + x_3 + 2x_4 &= 3 \\ x_1 + 0x_2 + 2x_3 + 0x_4 &= 1 \end{aligned}$$

(a) Let A be the coefficient matrix of the associated homogeneous system. Find the reduced form of A.

(b) Determine whether the system is consistent and, if so, find the general solution.

(c) Find the fundamental solutions of $A\mathbf{x} = \mathbf{0}$ and show that the fundamental solutions span $\mathcal{N}(A)$.

(d) Is the system $A\mathbf{x} = \mathbf{b}$ consistent for all $\mathbf{b} \in \mathbb{R}^3$? If not, find an equation which the components of $\mathbf{b}$ must satisfy.

Exercise 2.2. If A is 9×27, explain why the system $A\mathbf{x} = \mathbf{0}$ must have at least 18 fundamental solutions.

Exercise 2.3. Consider the system $A\mathbf{x} = \mathbf{0}$ where $A = \left(\begin{smallmatrix} 1 & -1 & 2 & -1 & 1 \\ -2 & 2 & 1 & -2 & 0 \end{smallmatrix}\right)$. Find the fundamental solutions and show they span $\mathcal{N}(A)$.

Exercise 2.4. Let A be the 2×5 matrix of Problem 2.3. Solve the compounded linear system

$$\left(A | \begin{smallmatrix} 1 & -1 \\ -2 & 0 \end{smallmatrix}\right).$$

Exercise 2.5. Set up a linear system to determine whether $(1, 0, -1, 1)$ is a linear combination of $(-1, 1, 2, 0)$, $(2, 1, 0, 1)$ and $(0, 1, 0, -1)$ with real coefficients. What about when the coefficients are in $\mathbb{Z}_3$? Note that in $\mathbb{Z}_3$, $-1 = 2$.

Exercise 2.6. A baseball team has won 3 mores games at home than on the road, and lost 5 more at home than on the road. If the team has played a total of 42 games, and if the number of home wins plus the number of road losses is 20, determine the number of home wins, road wins, home losses and road losses.

Exercise 2.7. For what real values of a and b does the system

$$\begin{aligned} x + ay + a^2z &= 1 \\ x + ay + abz &= a \\ bx + a^2y + a^2bz &= a^2b \end{aligned}$$

have a unique solution?

Exercise 2.8. True or False: If the normals of three planes in $\mathbb{R}^3$ through the origin lie in a plane through the origin, then the planes meet in a line.

Exercise 2.9. Suppose A is a 12×15 matrix of rank 12. How many fundamental solutions are there in $\mathcal{N}(A)$?

Exercise 2.10. . How many 2×2 matrices of rank 2 are there if we impose the condition that the entries are either 0 or 1? What about 3×3 matrices of rank 3 with the same condition?

Exercise 2.11. Find the ranks of each of the following matrices:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 8 & 27 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 2 \\ 1 & 4 & 4 \\ 1 & 8 & 8 \end{pmatrix}.$$

Can you formulate a general result from your results?

Exercise 2.12. If a chicken and a half lay and egg and a half in a day and a half, how many eggs does a single chicken lay in one day? Bonus marks for relating this to linear equations.

2.3 Matrix Algebra

Our goal in this section is to introduce matrix algebra and to show how it it is closely related to the theory of linear systems.

2.3.1 Matrix Addition and Multiplication

Let $\mathbb{R}^{m \times n}$ denote the set of all $m \times n$ matrices with real entries. There are three basic algebraic operations on matrices. These are addition, scalar multiplication and matrix multiplication. There are conditions which govern when two matrices can be added and when they can be multiplied. In particular, one cannot add or multiply any pair of matrices. First of all, suppose $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ and r is a scalar. Then the scalar multiple rA of

A is the matrix $rA = (ra_{ij}) \in \mathbb{R}^{m\times n}$ in which every element of A has been multiplied by r. For example, if A is 2×3, then

$$3A = \begin{pmatrix} 3a_{11} & 3a_{12} & 3a_{13} \\ 3a_{21} & 3a_{22} & 3a_{23} \end{pmatrix}.$$

Matrix addition can only be carried out on matrices of the same dimension. When A and B have the same dimension, say $m \times n$, we take their sum in the obvious manner. If $A = (a_{ij})$ and $B = (b_{ij})$, then $A + B$ is defined to be the $m \times n$ matrix $A + B := (a_{ij} + b_{ij})$. In other words, the (i, j) entry of $A + B$ is $a_{ij} + b_{ij}$. For example,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}.$$

Addition and scalar multiplication can be combined in the usual way to give linear combinations of matrices (of the same dimension). Here is an example.

Example 2.7. Let

$$A = \begin{pmatrix} 1 & 1 & 0 & 2 \\ 2 & -4 & 0 & 1 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 2 & 1 & 2 \end{pmatrix}.$$

Then

$$3A = \begin{pmatrix} 3 & 3 & 0 & 6 \\ 6 & -12 & 0 & 3 \end{pmatrix} \qquad \text{and} \qquad A + B = \begin{pmatrix} 2 & 2 & 1 & 2 \\ 2 & -2 & 1 & 3 \end{pmatrix}.$$

The $m \times n$ matrix such that every entry is 0 is called the $m \times n$ *zero matrix*. Clearly, the $m \times n$ zero matrix is an additive identity for addition of $m \times n$ matrices. Now that the additive identity is defined, we can also note that any $m \times n$ matrix A has as an additive inverse $-A$, since $A + (-A) = O$.

2.3.2 Matrices Over $\mathbb{F}_2$: Lorenz Codes and Scanners

So far we have only considered matrices over the real numbers. After we define fields, in the next Chapter, we will be able to compute with matrices over other fields, such as the complex numbers $\mathbb{C}$. Briefly, a field is a set with addition and multiplication which satisfies the basic algebraic properties of the integers, but where we can also divide.

The smallest field is $\mathbb{F}_2$, the integers mod 2, which consists of 0 and 1 with the usual rules of addition and multiplication, except that $1 + 1$ is defined to be 0: $1+1=0$. The integers mod 2 are most used in computer science since. (Just look at the on-off switch on a PC.) Adding 1 represents a change of state while adding 0 represents status quo.

Matrices over $\mathbb{F}^2$ are themselves quite interesting. For example, since $\mathbb{F}_2$ has only two elements, there are precisely 2^{mn} such matrices. Addition on such matrices also has some interesting properties, as the following example shows.

Example 2.8. For example,

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

and

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

In the first sum, the parity of every element in the first matrix is reversed. In the second, we see every matrix over $\mathbb{F}_2$ is its own additive inverse.

Example 2.9. Random Key Crypts. Suppose Rocky wants to send a message to Bullwinkle, and he wants to make sure that Boris and Natasha won't be able to learn what it says. Here is what the ever resourceful flying squirrel can do. First he encodes the message as a sequence of zeros and ones. For example, he can use the binary expansions of 1 through 26, thinking of 1 as a, 2 as b etc. Note that $1 = 1, 2 = 10, 3 = 11, 4 = 100\ldots, 26 = 11010$. Now he represents each letter as a five digit string: $a = 00001$, $b = 00010$, $c = 00011$, and so on, and encodes the message. Rocky now has a long string zeros and ones, which is usually called the *plaintext*. Finally, to make things more compact, he arranges the plaintext into a 01 matrix by adding line breaks at appropriate places. Let's denote this matrix by P, and suppose P is $m \times n$. Now the fun starts. Rocky and Bullwinkle have a list of $m \times n$ matrices of zeros and ones that only they know. The flying squirrel selects one of these matrices, say number 47, and tells Bullwinkle. Let E be matrix number 47. Cryptographers call E the *key*. Now he sends the *ciphertext* $enc_E(P) = P + E$ to Bullwinkle. If only Rocky and Bullwinkle know E, then the matrix P containing the plaintext is secure. Even if Boris and Natasha succeed in overhearing the ciphertext $P + E$, they will still have to know E to find out what P is. The trick is that the key E has to be sufficiently random so that neither Boris nor Natasha can guess it. For example, if E is the all ones matrix, then P isn't very secure since Boris and Natasha will surely try it. Notice that once Bullwinkle receives the ciphertext, all he has to do is add E and he gets P. For

$$enc_E(P) + E = (P + E) + E = P + (E + E) = P + O = P.$$

This is something even a mathematically challenged moose can do. Our hero's encryption scheme is extremely secure if the key E is sufficiently random and it is only used once. (Such a crypt is called a *one time pad.*) However, if he uses E to encrypt another plaintext message Q, and Boris and Natasha pick up both $enc_E(P) = P + E$ and $enc_E(Q) = Q + E$, then they can likely find out what both P and Q say. The reason for this is that

$$(P + E) + (Q + E) = (P + Q) + (E + E) = P + Q + O = P + Q.$$

The point is that knowing $P + Q$ may be enough for a cryptographer to deduce both P and Q. However, as a one time pad, the random key is very secure (in fact, apparently secure enough for communications on the hot line between Washington and Moscow).

Example 2.10. (**Scanners**) We can also interpret matrices over $\mathbb{F}_2$ in another natural way. Consider a black and white photograph as being a rectangular array consisting of many black and white dots. By giving the white dots the value 0 and the black dots the value 1, our black and white photo is therefore transformed into a matrix over $\mathbb{F}_2$. Now suppose we want to compare two black and white photographs whose matrices A and B have the same dimensions, that is, both are $m \times n$. It turns out to be inefficient for a computer to scan the two matrices to see in how many positions they agree. However, suppose we consider the sum $A + B$. When $A + B$ has a 1 in the (i, j)-component, then $a_{ij} \neq b_{ij}$, and when it has 0, then $a_{ij} = b_{ij}$. Hence the sum two identical photographs will be the zero matrix, and the sum of two complementary photographs will sum to the all ones matrix. An obvious and convenient measure of how similar the two matrices A and B are is the number of non zero entries of $A+B$. This number. which is easily tabulated, is known as the *Hamming distance* between A and B.

2.3.3 Matrix Multiplication

The third algebraic operation, *matrix multiplication*, is the most important and the least obvious to define. For one thing, the product of two matrices of the same dimension is only defined if the matrices are square. The product AB of two matrices A and B exists only when the number of columns of A equals the number of rows of B.

Definition 2.6. Let A be $m \times n$ and B $n \times p$. Then the *product* AB of A and B is the $m \times p$ matrix C whose entry in the ith row and kth column is

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

Thus

$$AB = \big(\sum_{j=1}^{n} a_{ij} b_{jk}\big).$$

Put another way, if the columns of A are $\mathbf{a}_1, \dots, \mathbf{a}_n$, then the rth column of AB is

$$b_{1r}\mathbf{a}_1 + b_{2r}\mathbf{a}_2 + \dots b_{nr}\mathbf{a}_n.$$

Hence the rth column of AB is the linear combination of the columns of A using the entries of the rth column of B as the scalars. One can also express AB as a linear comination of the rows of B. This turns out to be connected with row operations. The reader is invited to work this out explicitly.

Example 2.11. Here are two examples.

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 6 & 0 \\ -2 & 7 \end{pmatrix} = \begin{pmatrix} 1 \cdot 6 + 3 \cdot (-2) & 1 \cdot 0 + 3 \cdot 7 \\ 2 \cdot 6 + 4 \cdot (-2) & 2 \cdot 0 + 4 \cdot 7 \end{pmatrix} = \begin{pmatrix} 0 & 21 \\ 4 & 28 \end{pmatrix}.$$

Note how the columns of the product are linear combinations. Computing the product in the opposite order gives a different result:

$$\begin{pmatrix} 6 & 0 \\ -2 & 7 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 6 \cdot 1 + 0 \cdot 2 & 6 \cdot 3 + 0 \cdot 4 \\ -2 \cdot 1 + 7 \cdot 2 & -2 \cdot 3 + 7 \cdot 4 \end{pmatrix} = \begin{pmatrix} 6 & 18 \\ 12 & 22 \end{pmatrix}.$$

From this example, we have a pair of 2×2 matrices A and B such that $AB \neq BA$. More generally, multiplication of $n \times n$ matrices is not commutative, although there is a notable exception: if A and B are 1×1, then $AB = BA$.

2.3.4 The Transpose of a Matrix

Another operation on matrices is *transposition*, or taking the transpose. If A is $m \times n$, the *transpose* A^T of A is the $n \times m$ matrix $A^T := (c_{rs})$, where $c_{rs} = a_{sr}$. This is easy to remember: the ith row of A^T is just the ith column of A. Here are two obvious facts. First,

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

Second, a matrix and its transpose have the same diagonal. A matrix A which is equal to its transpose (that is, $A = A^T$) is called *symmetric.* Clearly, every symmetric matrix is square. The symmetric matrices over $\mathbb{R}$ turn out to be especially important, as we will see later.

Example 2.12. If

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

then

$$A^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}.$$

An example of a 2×2 symmetric matrix is

$$\begin{pmatrix} 1 & 3 \\ 3 & 5 \end{pmatrix}.$$

Notice that the dot product $\mathbf{v} \cdot \mathbf{w}$ of any two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ can be expressed as a matrix product, provided we use the transpose. In fact,

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = \sum v_i w_i.$$

The transpose of a product has an amusing property:

$$\big(AB\big)^T = B^T A^T.$$

This transpose identity can be seen as follows. The (i, j) entry of $B^T A^T$ is the dot product of the ith row of B^T and the jth column of A^T. Since this is the same thing as the dot product of the jth row of A and the ith column of B, which is the (j, i) entry of AB, and hence the (i, j) entry of $(AB)^T$, we see that $(AB)^T = B^T A^T$. Suggestion: try this out on an example.

2.3.5 The Algebraic Laws

Except for the commutativity of multiplication, the expected algebraic properties of addition and multiplication all hold for matrices. Assuming all the sums and products below are defined, matrix algebra obeys following laws:

(1) **Associative Law**: Matrix addition and multiplication are associative:

$$\big(A + B\big) + C = A + \big(B + C\big) \quad \text{and} \quad \big(AB\big)C = A\big(BC\big).$$

(2) **Distributive Law**: Matrix addition and multiplication are distributive:

$$A\big(B+C\big) = AB + AC \qquad \text{and} \qquad \big(A+B\big)C = AC + BC.$$

(3) **Scalar Multiplication Law**: For any scalar r,

$$\big(rA\big)B = A\big(rB\big) = r\big(AB\big).$$

(4) **Commutative Law for Addition**: Matrix addition is commutative: $A + B = B + A$.

Verifying these properties is a routine exercise. I suggest working a couple of examples to convince yourself, if necessary. Though seemingly uninteresting, the associative law for multiplication will often turn to be a very important property.

Recall that the $n \times n$ *identity matrix* I_n is the matrix having one in each diagonal entry and zero in each entry off the diagonal. For example,

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that it makes sense to refer to the identity matrix over $\mathbb{F}_2$, since ! is the multiplicative identity of $\mathbb{F}^2$.

The identity matrix I_n is a multiplicative identity for matrix multiplication. More precisely, we have

Proposition 2.6. *If A is an $m \times n$ matrix, then $AI_n = A$ and $I_mA = A$.*

Proof. This is an exercise. □

Exercises

Exercise 2.13. Make up three matrices A, B, C so that AB and BC are defined. Then compute AB and $(AB)C$. Next compute BC and $A(BC)$. Compare your results.

Exercise 2.14. Suppose A and B are symmetric $n \times n$ matrices. (You can even assume $n = 2$.)

(a) Decide whether or not AB is always symmetric. That is, whether $(AB)^T = AB$ for all symmetric A and B?

(b) If the answer to (a) is no, what condition ensures AB is symmetric?

Exercise 2.15. Suppose B has a column of zeros. How does this affect any product of the form AB? What if A has a row or a column of zeros?

Exercise 2.16. Let A be the 2×2 matrix over $\mathbb{F}_2$ such that $a_{ij} = 1$ for each i, j. Compute A^m for any integer $m > 0$. Does this question make sense if $m < 0$? (Note A^j is the product $AA \cdots A$ of A with itself j times.)

Exercise 2.17. Generalize this question to 2×2 matrices over $\mathbb{F}_2p$.

Exercise 2.18. Let A be the $n \times n$ matrix over $\mathbb{R}$ such that $a_{ij} = 2$ for all i, j. Find a formula for A^j for any positive integer j.

Exercise 2.19. Verify Proposition 2.6 for all $m \times n$ matrices A over $\mathbb{R}$.

Exercise 2.20. Give an example of a 2×2 matrix A such that every entry of A is either 0 or 1 and $A^2 = I_2$ as a matrix over $\mathbb{F}_2$, but $A^2 \neq I_2$ as a matrix over the reals.

2.4 Elementary Matrices and Row Operations

The purpose of this section is make an unexpected connection between matrix multiplication and row operations. We will see that in fact row operations can be done by matrix multiplication. For example, in the $2 \times n$ case, we use the following three types of 2×2 matrices:

$$E_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0 \\ 0 & r \end{pmatrix}, \qquad E_3 = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}.$$

These matrices enable us to do row operations of types I, II and III respectively via left or pre-multiplication, so they are called *elementary matrices.* For example,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix},$$

$$\begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ra & rb \\ c & d \end{pmatrix},$$

and

$$\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + sc & b + sd \\ c & d \end{pmatrix}.$$

Suitably modified, the same procedure works in general. An $m \times m$ matrix obtained from I_m by performing a single row operation is called an *elementary $m \times m$ matrix*. Here is the main point.

Proposition 2.7. *Let A be an $m \times m$ matrix, and assume E is an elementary $m \times m$ matrix. Then EA is the matrix obtained by performing the row operation corresponding to E on A.*

Proof. This follows from the fact that

$$EA = (EI_m)A,$$

and EI_m is the result of applying the row operation corresponding to E to I_m. Thus left multiplication by E will do the same thing to A that it does to I_m. □

Since any matrix can be put into reduced form by a sequence of row operations, and since row operations can be performed by left multiplication by elementary matrices, we have

Proposition 2.8. *An arbitrary $m \times n$ matrix A can be put into reduced form by a performing sequence of left multiplications on A using only $m \times m$ elementary matrices.*

Proof. This follows from the above comments. □

This procedure can be expressed as follows: starting with A and replacing it by $A_1 = E_1A$, $A_2 = E_2(E_1A))$ and so forth, we get the sequence

$$A \to A_1 = E_1A \to A_2 = E_2(E_1A) \to \cdots \to E_k(E_{k-1}(\cdots(E_1A)\cdots)),$$

the last matrix being A_{red}. What we obtain by this process is a matrix

$$B = (E_k(E_{k-1}\cdots(E_1A)\cdots))$$

with the property that $BA = A_{red}$. We want to emphasize that although B is expressed as a certain product of elementary matrices, the way we have chosen these matrices is never unique. However, it will turn out that B is unique in certain cases, one of which is the case where A is a nonsingular $n \times n$ matrix.

Note that we could have expressed B without parentheses writing it simply as $B = E_kE_{k-1}\cdots E_1$, due to the fact that, by the associative law, the parens can be rearranged at will.

When we are reducing a matrix A with entries in $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$ or even $\mathbb{F}_2$, then the elementary matrices we need to use also have entries in $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$ or $\mathbb{F}_2$, hence the matrix B which brings A into reduced form also has entries in the corresponding place. Hence we may state

Proposition 2.9. *For any $m \times n$ matrix A (with entries in $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$ or $\mathbb{F}_2$), there is an $m \times m$ matrix B (with entries in $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{C}$ or $\mathbb{F}_2$), which is a product of elementary matrices, such that $BA = A_{red}$.*

Example 2.13. Let's compute the matrix B produced by the sequence of row operations in Example 2.1 which puts the counting matrix C in reduced form. Examining the sequence of row operations, we see that B is the product

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

Thus

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

Be careful to express the product in the correct order. The first row operation is the made by the matrix on the right and the last by the matrix on the left. Thus

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

That is, $BC = C_{red}$.

In the above computation, you should not be explicitly multiplying the elementary matrices out. Start at the right and apply the sequence of row operations working to the left. A convenient way of doing this is to begin with the 3×6 matrix $(A|I_3)$ and carry out the sequence of row operations. The final result will be $(A_{red}|B)$. Thus if we start with

$$(A|I_3) = \begin{pmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{pmatrix},$$

we end with

$$(A_{red}|B) = \begin{pmatrix} 1 & 0 & -1 & -5/3 & 2/3 & 0 \\ 0 & 1 & 2 & 4/3 & -1/3 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{pmatrix}.$$

2.4.1 Application to Linear Systems

How does this method apply to solving linear systems? Note that the linear system (2.1 can be expressed in the compact matrix form

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

where A is the coefficient matrix, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ is the column of variables, and $\mathbf{b} = (b_1, b_2, \ldots, b_m)^T$ is the column of constants.

Starting with a system $A\mathbf{x} = \mathbf{b}$, where A is $m \times n$, multiplying this equation by any elementary matrix E gives a new linear system $EA\mathbf{x} = E\mathbf{b}$, which we know is equivalent to the original system. Therefore, applying Proposition 2.9, we obtain

Proposition 2.10. *Given the linear system $A\mathbf{x} = \mathbf{b}$, there exists a square matrix B which is a product of elementary matrices, such that the original system is equivalent to $A_{red}\mathbf{x} = B\mathbf{b}$.*

What's useful is that given E, there exists an elementary matrix F such that $FE = I_m$. It follows (after a little thought) that there exists a square matrix C such that $CB = I_m$. We will expand on this in the following section.

The advantage of knowing the matrix B which brings A into reduced form is that at least symbolically one can handle an arbitrary number of systems as easily as one. In other words, one can just as easily solve a matrix linear equation $A\mathbf{X} = \mathbf{D}$, where $\mathbf{X} = (x_{ij})$ is a matrix of variables and $\mathbf{D} = (D_{jk})$ is a matrix of constants. If A is $m \times n$ and D has p columns, then X is $n \times p$ and D is $m \times p$. This matrix equation is equivalent to $A_{red}X = BD$.

Exercises

Exercise 2.21. Find the reduced row echelon form for each of the following matrices, which are assumed to be over $\mathbb{R}$:

$$A_1 = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 3 & 1 \\ 1 & 2 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 2 & -1 & 1 \\ 2 & 3 & 1 & 0 \\ 0 & 1 & 2 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Exercise 2.22. Repeat Exercise 2.21 for the following matrices, except assume that each matrix is defined over $\mathbb{Z}_2$:

$$C_1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad C_3 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Exercise 2.23. Find matrices B_1, B_2 and B_3 which are products of elementary matrices such that B_iA_i is reduced, where A_1, A_2, A_3 are the matrices of Exercise 1.

Exercise 2.24. Find matrices D_1, D_2 and D_3 defined over $\mathbb{Z}_2$ which are products of elementary matrices such that D_iC_i is reduced, where C_1, C_2, C_3 are the matrices of Exercise 2.

Exercise 2.25. Prove carefully the if E is an elementary matrix and F is the elementary matrix that performs the inverse operation, then $FE = EF = I_n$.

Exercise 2.26. Write down all the 3×3 elementary matrices E over $\mathbb{Z}_2$. For each E, find the matrix F defined in the previous exercise such that $FE = EF = I_3$.

Exercise 2.27. Repeat Exercise 2.26 for the elementary matrices over $\mathbb{Z}_3$.

Exercise 2.28. List all the row reduced 2×3 matrices over $\mathbb{Z}_2$.

2.5 Matrix Inverses

Given an elementary matrix E, we noted in the last section that there exists another elementary matrix F such that $FE = I_m$. A little thought will convince you that not only is $FE = I_m$, but $EF = I_m$ as well. Doing a row operation then undoing it produces the same result as first undoing it and then doing it. Either way you are back to where you started. The essential property is pointed out in the next

Definition 2.7. Suppose two $m \times m$ matrices A and B have the property that $AB = BA = I_m$. Then we say A is an *inverse* of B (and B is an inverse of A).

We will use A^{-1} to denote an inverse of A. In the 2×2 examples above,

$$E_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \Rightarrow E_1^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$E_2 = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} \Rightarrow E_2^{-1} = \begin{pmatrix} r^{-1} & 0 \\ 0 & 1 \end{pmatrix},$$

and

$$E_3 = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \Rightarrow E_3^{-1} = \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix}.$$

2.5.1 A Necessary and Sufficient for Existence

Recall that an $n \times n$ matrix A is called invertible if A has rank n. We therefore have the

Proposition 2.11. *If an $n \times n$ matrix A is invertible, there exists an $n \times n$ matrix B such that $BA = I_n$.*

Proof. This follows from Proposition 2.9 since $A_{red} = I_n$. □

This relates the notion of invertibilty say for real numbers with that of invertibilty for matrices. Thus we would like to know when A has a two sided inverse, not just an inverse on the left.

Theorem 2.12. *An $n \times n$ matrix A has an inverse B if and only if A has rank n. Moreover, there can only be one inverse. Finally, if an $n \times n$ matrix A has some left inverse, then A is invertible and the left inverse is the unique inverse A^{-1}.*

Proof. Suppose first that A has two inverses B and C. Then

$$B = BI_n = B(AC) = (BA)C = I_nC = C.$$

Thus t$B = C$, so the inverse is unique. Next, suppose A has a left inverse B. We will show that the rank of A is n. For this, we have to show that if $A\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = \mathbf{0}$. But if $A\mathbf{x} = \mathbf{0}$, then

$$B(A\mathbf{x}) = (BA)\mathbf{x} = I_n\mathbf{x} = \mathbf{0}. \tag{2.6}$$

Thus indeed, A does have rank n. Now suppose A has rank n. Then we know the system $A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$. It follows that there exists an $n \times n$ matrix X so that $AX = I_n$. This follows from knowing the system $A\mathbf{x} = \mathbf{e}_i$ has a solution for each i, where $\mathbf{e}_i$ is the ith column of I_n. Thus there exist $n \times n$ matrices B and X so that $BA = AX = I_n$. We now show that $B = X$. Repeating the above argument, we have

$$B = BI_n = B(AX) = (BA)X = I_nX = X.$$

Thus A has an inverse if and only if it has rank n. To finish the proof, suppose A has a left inverse B: that is B is $n \times n$ and $BA = I_n$. But we just showed in (2.6) that A has rank n, so (as we concluded above), A^{-1} exists and equals B. □

This theorem explains why we call square matrices of maximal rank invertible.

Corollary 2.13. *If A invertible, then the system*

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

has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof. We leave this as an exercise. □

The product of any two invertible $n \times n$ matrices A and B is also invertible. Indeed, $(AB)^{-1} = B^{-1}A^{-1}$. For

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

This is used in the proof of the following useful Proposition.

Proposition 2.14. *Every invertible matrix is a product of elementary matrices.*

The proof is left as an exercise. Of course, by the previous Proposition, the converse is also true: any product of elementary matrices is invertible.

2.5.2 Methods for Finding Inverses

We have two ways of finding the matrix B so that $BA = A_{red}$. The first is simply to multiply out the sequence of elementary matrices which reduces A. This is not as bad as it sounds since multiplying elementary matrices is very easy. The second method is to form the augmented matrix $(A|I_n)$ and row reduce. The final result will be in the form $(I_n|A^{-1})$. This is the method used in most textbooks. Let's begin with an example.

Example 2.14. Suppose we want to find an inverse for

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

Since we only need to solve the matrix equation $XA = I_3$, we can use our previous strategy of row reducing $(A|I_3)$.

$$(A|I_3) = \begin{pmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{pmatrix} \to$$

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

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

Hence

$$A^{-1} = B = \begin{pmatrix} 5 & -4 & 2 \\ -2 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix},$$

since, by construction, $BA = I_3$.

Example 2.15. To take a slightly more interesting example, let

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

The catch is that we will assume that the entries of A are elements of $\mathbb{F}_2 = \{0, 1\}$. Imitating the above procedure, we obtain that

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

Note that the correctness of this result should be checked by computing directly that

$$I_4 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

There is somewhat less obvious third technique which is sometimes also useful. If we form the *augmented coefficient matrix* $(A \mid \mathbf{b})$, where $\mathbf{b}$ represents the column vector with components $b_1, b_2, \dots b_m$ and perform the row reduction of this augmented matrix, the result will be in the form $(I_n | \ \mathbf{c})$, where the components of $\mathbf{c}$ are certain linear combinations of the components of $\mathbf{b}$. The coefficients in these linear combinations give us the entries of A^{-1}. Here is an example.

Example 2.16. Again let

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

Now form

$$\begin{pmatrix} 1 & 2 & 0 & a \\ 1 & 3 & 1 & b \\ 0 & 1 & 2 & c \end{pmatrix}$$

and row reduce. The result is

$$\begin{pmatrix} 1 & 0 & 0 & 5a - 4b + 2c \\ 0 & 1 & 0 & -2a + 2b - c \\ 0 & 0 & 1 & a - b + c \end{pmatrix}.$$

Thus we see the inverse is

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

2.5.3 Matrix Groups

In this section, we will give some examples of what are called matrix groups or linear groups. The basic example of a matrix group is the set $GL(n,\mathbb{R})$ of all invertible elements of $\mathbb{R}^{n\times n}$. Thus,

$$GL(n,\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} \mid A^{-1} \text{ exists}\}. \tag{2.7}$$

Notice that, by definition, every element in $GL(n,\mathbb{R})$ has an inverse. Moreover, I_n is an element of $GL(n,\mathbb{R})$, and if A and B are elements of $GL(n,\mathbb{R})$, then so is their product AB. These three properties define what we mean by a matrix group.

Definition 2.8. A subset G of $\mathbb{R}^{n\times n}$ is called a *matrix group* if the following three conditions hold:

(i) if $A, B \in G$, then $AB \in G$,

(ii) $I_n \in G$, and

(iii) if $A \in G$, then $A^{-1} \in G$.

It turns out that these three axioms are broad enough to give each matrix group an extremely rich structure. Of course, as already noted above, $GL(n,\mathbb{R})$ is a matrix group. In fact, if $G \subset \mathbb{R}^{n\times n}$ is any matrix group, then $G \subset GL(n,\mathbb{R})$ (why?). A subset of $GL(n,\mathbb{R})$ which is also a matrix group is called a *subgroup* of $GL(n,\mathbb{R})$. Thus we want to consider subgroups of $GL(n,\mathbb{R})$. The simplest example of a subgroup of $GL(n,\mathbb{R})$ is $\{I_n\}$: this is the so called *trivial subgroup.*

To get some simple yet interesting examples, let us consider permutation matrices.

Example 2.17 (Permutation Matrices). A matrix P obtained from I_n by a finite sequence of row swaps is called a *permutation matrix.* In other words, a permutation matrix is a matrix $P \in \mathbb{R}^{n\times n}$ such that there are row swap matrices $S_1, \dots, S_k \in \mathbb{R}^{n\times n}$ for which $P = S_1 \cdots S_k$. (Recall that a row swap matrix is by definition an elementary matrix obtained by interchanging two rows of I_n.) Clearly, I_n is a permutation matrix (why?), and any product of permutation matrices is also a permutation matrix. Thus we only need to see that the inverse of a permutation matrix is also a permutation matrix. Let $P = S_1 \cdots S_k$ be a permutation matrix. Then $S^{-1} = S_k^{-1} \cdots S_1^{-1}$, so P^{-1} is indeed a permutation matrix since $S_i^{-1} = S_i$ for each index i.

Let $P(n)$ denote the set of $n \times n$ permutation matrices. One can also describe $P(n)$ as the set of all matrices obtained from I_n by permuting the rows of I_n. Thus $P(n)$ is the set of all $n \times n$ matrices whose only entries are 0 or 1 such that every row and every column has exactly one non-zero entry. It follows from elementary combinatorics that $P(n)$ has exactly n elements.

The inverse of a permutation matrix has a beautiful expression.

Proposition 2.15. *If P is a permutation matrix, then $P^{-1} = P^T$.*

Proof. This follows from the above discussion. We leave the details for the exercises. □

To give an explicit example, let us compute $P(3)$.

Example 2.18. $P(3)$ consists of the following six 3×3 permutation matrices; namely I_3 and

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

Definition 2.9 (The orthogonal group). Let $Q \in \mathbb{R}^{n \times n}$. Then we say that Q is *orthogonal* if and only if $QQ^T = I_n$. The set of all $n \times n$ orthogonal matrices is denoted by $O(n, \mathbb{R})$. We call $O(n, \mathbb{R})$ the *orthogonal group.*

Proposition 2.16. *$O(n, \mathbb{R})$ is a subgroup of $GL(n, \mathbb{R})$.*

Proof. It follows immediately from the definition and Theorem 2.12 that if Q is orthogonal, then $Q^T = Q^{-1}$. Consequently, since $QQ^T = I_n$ implies $Q^TQ = I_n$, whenever Q is orthogonal, so is Q^{-1}. The identity I_n is clearly orthogonal, so it remains to show that the product of two orthogonal matrices is orthogonal. Let Q and R be orthogonal. Then

$$QR(QR)^T = QR(R^TQ^T) = Q(RR^T)Q^T = QQ^T = I_n.$$

□

By Proposition 2.15, we have $P(n) \subset O(n, \mathbb{R})$. That is, every permutation matrix is orthogonal. Hence $P(n)$ is a subgroup of $O(n, \mathbb{R})$.

The condition $Q^TQ = I_n$ which defines an orthogonal matrix Q is equivalent to the property that as a transformation of $\mathbb{R}^n$ to itself, Q preserves inner products. That is, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}. \tag{2.8}$$

Indeed, since $Q^TQ = I_n$,

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y} = \mathbf{x}^TQ^TQ\mathbf{y} = (Q\mathbf{x})^TQ\mathbf{y} = Q\mathbf{x} \cdot Q\mathbf{y}. \tag{2.9}$$

Conversely, if $Q\mathbf{x}\cdot Q\mathbf{y} = \mathbf{x}\cdot\mathbf{y}$, then $Q^TQ = I_n$. (Just let $\mathbf{x} = \mathbf{e}_i$ and $\mathbf{y} = \mathbf{e}_j$.) In particular, we can now conclude

Proposition 2.17. *Every element of the orthogonal group $O(n,\mathbb{R})$ preserves lengths of vectors and also distances and angles between vectors.*

Proof. This follows from the identity $\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}||\mathbf{y}|\cos\theta$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. □

The preceding Proposition tells us that the orthogonal group $O(n,\mathbb{R})$ is intimately related to the geometry of $\mathbb{R}^n$. If Q is orthogonal, then the columns of Q are mutually orthogonal unit vectors, which is a fact we will frequently use.

The orthogonal group for $O(2,\mathbb{R})$ is especially interesting. It has an important subgroup $SO(2)$ called the *rotation group* which consists of the rotation matrices

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Note that $R_0 = I_2$. The fact that $SO(2)$ is a subgroup of $O(2,\mathbb{R})$ follows from trigonometry. For example, the sum formulas for $\cos(\theta+\mu)$ and $\sin(\theta+\mu)$, are equivalent to the geometrically obvious formulas

$$R_\theta R_\mu = R_\mu R_\theta = R_{\theta+\mu} \tag{2.10}$$

for all θ and μ. We will discuss this later that in more detail.

Example 2.19. There are three subgroups of $GL(n,\mathbb{R})$ which we encountered in Chapter 4: the group $\mathcal{D}_n$ of all invertible diagonal matrices (those diagonal matrices with no zeros on the diagonal), the group $\mathcal{L}_n$ of all lower triangular matrices with only ones on the diagonal, and the group $\mathcal{U}_n$ of all upper triangular matrices with ones on the diagonal. In the proof of Theorem 2.19, we actually used the fact that $\mathcal{L}_n$ and $\mathcal{U}_n$ are matrix groups.

Exercises

Exercise 2.29. Find the inverse of each of the following real matrices or show that the inverse does not exist.

$$\text{(a)} \begin{pmatrix} 1 & 2 \\ 4 & 1 \end{pmatrix} \text{(b)} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{pmatrix} \text{(c)} \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \text{(d)} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

Exercise 2.30. If the field is $\mathbb{Z}_2$, which of the matrices in Exercise 1 are invertible?

Exercise 2.31. Suppose $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, and assume that $\Delta = ad - bc \neq 0$. Show that $A^{-1} = \frac{1}{\Delta}\left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$. What does the condition $\Delta \neq 0$ mean in terms of the rows of A?

Exercise 2.32. Suppose A has an inverse. Find a formula for the inverse of A^T?

Exercise 2.33. Prove Proposition 2.14.

Exercise 2.34. Suppose A is $n \times n$ and there exists a right inverse B, i.e. $AB = I_n$. Show A invertible.

Exercise 2.35. Let $C = \left(\begin{smallmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{smallmatrix}\right)$. Find a general formula for C^{-1}.

Exercise 2.36. Show that if A and B are $n \times n$ and have inverses, then $(AB)^{-1} = B^{-1}A^{-1}$. What is $(ABCD)^{-1}$ if all four matrices are invertible?

Exercise 2.37. Suppose A is invertible $m \times m$ and B is $m \times n$. Solve the equation $AX = B$.

Exercise 2.38. Suppose A and B are both $n \times n$ and AB is invertible. Show that both A and B are invertible. (See what happens if $B\mathbf{x} = \mathbf{0}$.)

Exercise 2.39. Let A and B be two $n \times n$ matrices over $\mathbb{R}$. Suppose $A^3 = B^3$, and $A^2B = B^2A$. Show that if $A^2 + B^2$ is invertible, then $A = B$. (Hint: Consider $(A^2 + B^2)A$.)

Exercise 2.40. Without computing, try to guess the inverse of the matrix

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

(Hint: are the columns orthogonal?)

Exercise 2.41. Is it TRUE or FALSE that if an $n \times n$ matrix with integer entries has an inverse, then the inverse also has integer entries?

Exercise 2.42. Consider the matrix

$$B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix}.$$

Find B^{-1}, if it exists, when B is considered to be a real matrix. Is B invertible when it is considered as a matrix over $\mathbb{Z}_2$?

Exercise 2.43. A real $n \times n$ matrix Q such that $Q^TQ = I_n$ is called *orthogonal*. Find a formula for the inverse of an arbitrary orthogonal matrix Q. Also show that the inverse of an orthogonal matrix is also orthogonal.

Exercise 2.44. Show that the product of two orthogonal matrices is orthogonal.

Exercise 2.45. A matrix which can be expressed as a product of row swap matrices is called a *permutation matrix*. These are the matrices obtained by rearranging the rows of I_n. Show that every permutation matrix is orthogonal. Deduce that if P is a permutation matrix, then $P^{-1} = P^T$.

Exercise 2.46. Show that the following two matrices are permutation matrices and find their inverses:

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

Exercise 2.47. You are a code-breaker (more accurately, a cryptographer) assigned to crack a secret cipher constructed as follows. The sequence 01 represents A, 02 represents B and so forth up to 26, which represents Z. A space between words is indicated by inserting 00. A text can thus be encoded as a sequence. For example, 1908040002090700041507 stands for "the big dog". We can think of this as a vector in $\mathbb{R}^{22}$. Suppose a certain text has been encoded as a sequence of length 14,212=44×323, and the sequence has been broken into 323 consecutive intervals of length 44. Next, suppose each sub-interval is multiplied by a single 44×44 matrix C. The new sequence obtained by laying the products end to end is called the cipher text, because it has now been enciphered, and it is your job to decipher it. Discuss the following questions.

(i) How does one produce an invertible 44×44 matrix in an efficient way, and how does one find its inverse?

(ii) How many of the sub-intervals will you need to decipher to break the whole cipher by deducing the matrix C?

Exercise 2.48. Prove Proposition2.15.

Exercise 2.49. Show that if Q is orthogonal, then the columns of Q are mutually orthogonal unit vectors. Prove that this is also true for the rows of Q.

Exercise 2.50. * Show that every element H of $O(2, \mathbb{R})$ that isn't a rotation satisfies $H^T = H$ and $H^2 = I_2$. (Note I_2 is the only rotation that satisfies these conditions.)

2.6 The $LPDU$ Decomposition

In this section, we will show that every $n \times n$ invertible matrix A has an interesting factorization $A = LPDU$, where each of the matrices L, P, D, U has a particular form. This factorization is frequently used in solving large systems of linear equations by a process known as back substitution. We will not go into back substitution here, but the reader can consult a text on applied linear algebra, e.g. *Linear Algebra and its Applications* by G. Strang. In addition to its usefulness in applied linear algebra, the $LPDU$ decomposition is also of theoretical interest, since each P gives a class of matrices called a Schubert cell, which has many interesting properties.

In order to describe the necessary ingredients L, D, P, and U, we need column operations, pivots and permutation matrices. Of these, the permutation matrices are especially interesting, as we will encounter them in a number of other contexts later.

2.6.1 The Basic Ingredients: L, P, D and U

In the $LPDU$ decomposition, L is lower triangular and has only 1's on its diagonal, P is a permutation matrix, D is a diagonal matrix without any zeros on its diagonal, and U is upper triangular and has only 1's on its diagonal. The notable feature of these types of matrices is that each one can be constructed from just one kind of elementary matrix.

Let's introduce the cast of characters starting with lower triangular matrices. An $n \times n$ matrix $L = (l_{ij})$ is called *lower triangular* if all entries of L strictly above its diagonal are zero. In other words, $l_{ij} = 0$ if $i < j$. Similarly, a matrix is called *upper triangular* if all entries strictly below its diagonal are zero. Clearly, the transpose of a lower triangular matrix is upper triangular and vice versa. We will only be dealing with the upper or lower triangular matrices all of whose diagonal entries are 1's. These matrices are called, respectively, *upper triangular unipotent* and *lower triangular unipotent* matrices.

Example 2.20. Every lower triangular 3×3 unipotent matrix has the form

$$L = \begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix}.$$

The transpose $U = L^T$ is

$$U = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}.$$

We've already used lower triangular unipotent matrices for row reduction, namely, the elementary matrices of Type III which are also lower triangular. For these matrices, exactly one of a, b, c is different from zero. Left multiplication on A by such a matrix replaces a row of A by itself plus a certain multiple of one the rows above it. You can check without any difficulty the basic property that the product of two or more lower triangular elementary matrices of Type III is again lower triangular. Moreover, all the diagonal entries of the product will be ones (why?). More generally, the product of two or more lower triangular matrices is again lower triangular, and the product of two or more lower triangular unipotent matrices is again lower triangular unipotent.

Notice also that if L is lower triangular unipotent, then we can find lower triangular elementary matrices of Type III, say $E_1, E_2, \dots, E_k$ so that $E_k \cdots E_2 E_1 L = I_n$. Since the inverse of each E_i is another lower triangular elementary matrix of type III, we therefore see that $L = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$. Thus both L and L^{-1} can be expressed as a product of lower triangular elementary matrices of Type III. In particular, the inverse of a lower triangular unipotent matrix is also lower triangular unipotent.

We summarize this discussion in the following proposition:

Proposition 2.18. *The product of two lower triangular unipotent matrices is also lower triangular unipotent, and the inverse of a lower triangular unipotent matrix is a lower triangular unipotent matrix. The corresponding statements in the upper triangular unipotent case also hold.*

What this Proposition says is that the lower (resp. upper) triangular unipotent matrices is closed under the operations of taking products and inverses. Hence they form a pair of matrix groups.

Recall that an $n \times n$ matrix which can be expressed as a product of elementary matrices of Type II (row swaps for short) is called a *permutation matrix*. The $n \times n$ permutation matrices are exactly those matrices which can be obtained by rearranging the rows of I_n. We have already seen that they form a subgroup of $O(n, \mathbb{R})$. In particular, inverse of a permutation matrix P is P^T.

2.6.2 The Main Result

We now come to the main theorem.

Theorem 2.19. *Let $A = (a_{ij})$ be an invertible matrix over $\mathbb{R}$. Then A can be expressed in the form $A = LPDU$, where L is lower triangular unipotent, U is upper triangular unipotent, P is a permutation matrix, and D is a diagonal matrix with all its diagonal entries non zero. Furthermore, the matrices P and D are unique.*

This result gives the invertible matrices an interesting structure. Each of L, P, D, U is constructed from just one kind of elementary matrix. Note that we need to add that every invertible diagonal matrix is a product of Type I elementary matrices. Because of our assertion that the diagonal matrix D is unique, its diagonal entries have quite a bit of significance. The ith diagonal entry d_{ii} of D is usually called the ith *pivot* of A.

Example 2.21. Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be invertible. If $a \neq 0$, then the $LPDU$ decomposition of A is

$$A = \begin{pmatrix} 1 & 0 \\ -c/a & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & (ad-bc)/a \end{pmatrix} \begin{pmatrix} 1 & -b/a \\ 0 & 1 \end{pmatrix}.$$

However, if $a = 0$, then $bc \neq 0$ and A can be expressed either

as

$$LPD = \begin{pmatrix} 1 & 0 \\ d/b & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & b \end{pmatrix}$$

or

$$PDU = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 1 & d/c \\ 0 & 1 \end{pmatrix}.$$

This example tells us that L and U are not necessarily unique.

Proof of Theorem 2.19. This proof has the desirable feature that it is algorithmic. That is, it gives a clear procedure for finding the $LPDU$ factorization. The first step in the proof is to scan down the first column of A until we find the first non zero entry. Such an entry exists since A is invertible. Let $\sigma(1)$ denote the row this entry is in, and let $d_{\sigma(1)} = a_{\sigma(1),1}$ denote the entry itself. Now perform a sequence of row operations to make the entries below $a_{\sigma(1),1}$ equal to zero. This transforms the first column of A into

$$(0, \ldots, 0, d_{\sigma(1)}, 0, \ldots, 0)^T. \tag{2.11}$$

This reduction is performed by pre-multiplying A by a sequence of lower triangular elementary matrices of the third kind. We therefore obtain a lower triangular unipotent matrix L_1 so that the first column of L_1A has the form (2.11). The next step is to use the non zero entry $d_{\sigma(1)}$ in the first column to annihilate all the entries in the $\sigma(1)$-st row to the right of $d_{\sigma(1)}$. Since post multiplying by elementary matrices performs column operations, we can multiply L_1A on the right by a sequence of upper triangular elementary matrices of the third kind to produce a matrix of the form $(L_1A)U_1$ whose first column has the form (2.11) and whose $\sigma(1)$-st row is

$$(d_{\sigma(1)}, 0, \dots, 0). \tag{2.12}$$

Moreover, Proposition 2.18 guarantees that U_1 will be upper triangular unipotent. We now have the first column and $\sigma(1)$-st row of A in the desired form and from now on, they will be unchanged.

To continue, we repeat this procedure on the second column of L_1AU_1. Suppose the first non zero entry of the second column sits in the k-th row. Since we already cleared out all non zero entries of the $\sigma(1)$st row, $k \neq \sigma(1)$. Now set $\sigma(2) = k$ and repeat the same procedure we carried out for the first column and $\sigma(1)$st row. Continuing, we eventually obtain lower triangular unipotent matrices L_i and upper triangular unipotent matrices U_i and a rearrangement $\sigma(1), \sigma(2), \dots, \sigma(n)$ of $1, 2, \dots, n$ so that

$$(L_nL_{n-1}\cdots L_1)A(U_1U_2\cdots U_n)^{-1} = Q,$$

where Q is the matrix whose ith column has $d_{\sigma(i)}$ as its $\sigma(i)$th entry and zeros in every other entry, where $d_{\sigma(i)}$ is the first non zero entry in the ith column of $(L_{i-1}\cdots L_1)A$. We can clearly factor Q as PD where D is the diagonal matrix whose ith diagonal entry is $d_{\sigma(i)}$ and P is the permutation matrix with ones exactly where Q had non zero entries. This gives us the expression

$$A = (L_nL_{n-1}\cdots L_1)^{-1}PD(U_1U_2\cdots U_n)^{-1}.$$

But this is the desired factorization $A = LPDU$. Indeed, $L' = L_nL_{n-1}\cdots L_1$ is lower triangular unipotent since it is a product of lower triangular elementary matrices of type III, and hence its inverse L is also lower triangular unipotent. The same remarks hold when we put $U = (U_1U_2\cdots U_n)^{-1}$, so we have established the existence of the $LPDU$ factorization. We will leave the proof of the uniqueness of P and D as an exercise. □

Example 2.22. To illustrate the proof, let

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

Since the first non zero entry in the first column of A is $a_{13} = -1$, the first two steps are to subtract the first column twice from the second and to subtract it once from the third. The result is

$$AU_1U_2 = \begin{pmatrix} 0 & 2 & -2 \\ 0 & 4 & -5 \\ -1 & 0 & 0 \end{pmatrix}.$$

Next we subtract twice the first row from the second, which gives

$$L_1AU_1U_2 = \begin{pmatrix} 0 & 2 & -2 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix}.$$

Finally, we add the second column to the third and then factor, getting

$$Q = L_1AU_1U_2U_3 = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix}.$$

Now write

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

Thus we obtain the $LPDU$ factorization

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

2.6.3 The Case $P = I_n$

In the case where $P = I_n$, A can be row reduced without using row interchanges. In fact, in a sense that can be made precise, general invertible matrices do not require a row interchange. This is because a row interchange is only necessary when a zero shows up on a diagonal position during row reduction.

Proposition 2.20. *If an invertible matrix A admits an LDU decomposition, then the matrices L, D and U are all unique.*

Proof. If A has two LDU decompositions, say $A = L_1D_1U_1 = L_2D_2U_2$, then we can write $L_1^{-1}L_2D_2U_2 = D_1U_1$. Hence $L_1^{-1}L_2D_2 = D_1U_1U_2^{-1}$. But in this equation, the matrix on the left hand side is lower triangular and the matrix on the right hand side is upper triangular. This tells us that immediately that $D_1 = D_2$, and also that $L_1^{-1}L_2 = U_1U_2^{-1} = I_n$ (why?). Hence $L_1 = L_2$ and $U_1 = U_2$, so the proof is completed. □

Going back to the 2×2 case considered in Example 2.21, the LDU decomposition for A is therefore unique when $a \neq 0$. We also pointed out in the same example that if $a = 0$, then L and U are not unique.

If one were only interested in solving a square system $A\mathbf{x} = \mathbf{b}$, then finding the $LPDU$ factorization of A isn't necessary. In fact, it turns out that one can post multiply A by a permutation matrix Q which is concocted to move zero pivots out of the way. That is, if Q is chosen carefully, there exists a factorization $AQ = LDU$. The only affect on the system is to renumber the variables, replacing $\mathbf{x}$ by $Q^{-1}\mathbf{x}$. The L, D, U and Q are no longer unique.

2.6.4 The symmetric LDU decomposition

Suppose A is an invertible symmetric matrix which has an LDU decomposition. Then it turns out that L and U are not only unique, but they are related. In fact, $U = L^T$. This makes finding the LDU decomposition very simple. The reasoning for this goes as follows. If $A = A^T$ and $A = LDU$, then

$$LDU = (LDU)^T = U^TD^TL^T = U^TDL^T$$

since $D = D^T$. Therefore the uniqueness of L, D and U implies that $U = L^T$. The upshot is that to factor $A = LDU$, all one needs is to do row operations of Type III on A such that higher rows act on lower rows to bring A into upper triangular form B. This means all we need to do is to find a lower triangular unipotent matrix L' so that $L'A$ is upper triangular, i.e. $L'A = B$. Then the matrices D and U can be found by inspection. In fact, D is diagonal so $d_{ii} = b_{ii}$ for all i, and since all the b_{ii} are all nonzero, we can write $B = DU$, where U is an upper triangular unipotent matrix. This means $L'A = DU$, so we have found both U and D. Hence L is also known since, by the uniqueness result just proved,

$L = U^T$. Of course, it's also the case that $L = (L')^{-1}$, but this is the hard way to go.

Example 2.23. Consider the symmetric matrix

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

The strategy is to apply Type III row operations, only allowing higher rows to operate on lower rows, to bring A into upper triangular form, which is our DU. Doing so, we find that A reduces to

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

Hence

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and

$$U = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus $A = LDU$ where U is as above, $L = U^T$ and $D = \mathrm{diag}(1, 2, 1)$.

Summarizing, we state

Proposition 2.21. *If A is an (invertible) $n \times n$ symmetric matrix without zero pivots, then the $LPDU$ decomposition of A has the form $A = LDL^T$.*

When A is invertible and symmetric but has zero pivots, then a permutation matrix $P \neq I_n$ is needed. Expressing $A = LPDU$, it turns out that we may construct L and U so that $U = L^T$ still holds. This says that PD is also symmetric (why?). Since $P^T = P^{-1}$, we see that

$$PD = (PD)^T = D^T P^T = DP^{-1},$$

so $PDP = D$. I claim two conditions must be fulfilled in order to have this. The first is that since P is a permutation matrix and hence PD is D with its rows permuted, PDP cannot be a diagonal matrix unless $P = P^{-1}$. Since P is a permutation matrix, $P = P^{-1}$ if and only if $P = P^T$. We can therefore conclude that P is a symmetric permutation matrix. Moreover, this tells us that $PD = DP$, so P and D commute. Now look at PDP^{-1}. It can be shown that PDP^{-1} is always a diagonal matrix with the same diagonal entries as D, except in a different order. In fact, let the ith diagonal entry of PDP^{-1} be $d_{\sigma(i)}$. Then σ is the permutation which is determined by P. This gives the second condition; since $PDP^{-1} = D$, the diagonal of D must be left unchanged by the permutation σ. Thus D and P cannot be arbitrary when A is symmetric. We therefore have

Proposition 2.22. *Let A be a symmetric invertible matrix. Then there exists an expression $A = LPDU$ with L, P, D, U as usual and furthermore :*

(i) $U = L^T$,

(ii) $P = P^T = P^{-1}$, *and*

(iii) $PD = DP$.

Conversely, if L, P, D, U satisfy the above three conditions, then $LPDU$ is symmetric.

The next example shows how the disussion preceeding the last Propositon works.

Example 2.24. Let

$$A = \begin{pmatrix} 0 & 2 & 4 & -4 \\ 2 & 4 & 2 & -2 \\ 4 & 2 & -8 & 7 \\ -4 & -2 & 7 & -8 \end{pmatrix}.$$

The first step is to make the (3,1) an (4,1) entries of A zero while keeping A symmetric. This is done by using symmetric row and column operations. That is, we replace A by $A_1 = E_1AE_1^T$, $A_2 = E_2^TA_1E_2^T$ etc. Begin with E_1 which differs from I_4 only in the $(3, 2)$-entry, which is -2 instead of 0. Computing $E_1AE_1^T$ gives

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 2 & 4 & -4 \\ 2 & 4 & 2 & -2 \\ 4 & 2 & -8 & 7 \\ -4 & -2 & 7 & -8 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} =$$

$$\begin{pmatrix} 0 & 2 & 0 & -4 \\ 2 & 4 & -6 & -2 \\ 0 & -6 & 0 & 11 \\ -4 & -2 & 7 & -8 \end{pmatrix}.$$

Notice A_1 is symmetric (why?). Next let $A_2 = E_2A_1E_2^T$, where E_2 is obtained from I_4 by adding twice the second row to the fourth row. The result is

$$A_2 = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 2 & 4 & -6 & 6 \\ 0 & -6 & 0 & -1 \\ 0 & 6 & -1 & 0 \end{pmatrix}.$$

The next step is to remove the 4 in the $(2,2)$-position of A_2. This is done by symmetric elimination. We subtract the first row from the second row and the first column from the second column. This gives

$$A_3 = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 2 & 0 & -6 & 6 \\ 0 & -6 & 0 & -1 \\ 0 & 6 & -1 & 0 \end{pmatrix}.$$

It is easy to see how to finish. After two more eliminations, we end up with

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

Hence we get

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

and

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

We leave it as an exercise to find L and hence U.

Exercises

Exercise 2.51. Find the $LPDU$ decompositions of the following matrices:

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

Exercise 2.52. Find the inverse of

$$\begin{pmatrix}1&a&b&c\\0&1&d&e\\0&0&1&f\\0&0&0&1\end{pmatrix}.$$

Exercise 2.53. Show directly that an invertible upper triangular matrix B can be expressed $B = DU$, where D is a diagonal matrix with non zero diagonal entries and U is upper an triangular matrix all of whose diagonal entries are ones. Is this still true if B is singular?

Exercise 2.54. Show that the product of any two lower triangular matrices is lower triangular. Also show that the inverse of a lower triangular invertible matrix is lower triangular. What are the diagonal entries of the inverse?

Exercise 2.55. Let A be a 3×3 lower triangular unipotent matrix. Find a formula expressing A as a product of lower triangular elementary matrices of type III.

Exercise 2.56. Find the $LPDU$ decomposition of

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

Exercise 2.57. Find the LDU decomposition of

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

Exercise 2.58. Prove that in any $LPDU$ decomposition, P and D are unique. (Let $LPDU = L'P'D'U'$ and consider $L^{-1}L'P'D' = PDUU'^{-1}$.)

Exercise 2.59. Find a 3×3 matrix A such that the matrix L in the $A = LPDU$ decomposition isn't unique.

Exercise 2.60. Let A be the matrix of Example 2.24. Find the matrices L and U. Also, show that $PD = DP$.

2.7 Summary

This chapter was an introduction to linear systems and matrices. We began by introducing the general linear system of m equations in n unknowns with real coefficients. There are two types of systems called homogeneous and non-homogeneous according to whether the constants on the right hand sides of the equations are all zeros or not. The solutions make up the solution set. If the system is homogeneous, the solution set is a subspace of $\mathbb{R}^n$. In order to write the system in a compact form, we introduced the coefficient matrix for homogeneous systems and the augmented coefficient matrix for non-homogeneous systems. We then wrote down the three row operations of Gaussian reduction. The row operations give a specific set of rules for bringing the coefficient matrix and augmented coefficient matrix into a normal form known as reducued form. The point is that performing a row operation on the coefficient matrix (or augmented coefficient matrix) gives a new coefficient matrix (or augmented coefficient matrix) whose associated linear system has exactly the same solution space (or set in the non-homogeneous case).

After a matrix is put into reducued form, we can read off its rank (the number of non-zero rows). We then obtained criteria which are necessary and sufficient for the existence and uniqueness of solutions. A non-homogeneous system has a solution if and only if its augmented coefficient matrix and coefficient matrix have the same rank. A unique solution exists if and only if the augmented coefficient matrix and coefficient matrix have the same rank and the rank is the number of unknowns.

We next introduced matrix algebra, addition and multiplication. Matrices of the same size can always be added but to form AB, the number of rows of B must be the same as the number of columns of A. We saw how elementary matrices perform row opertions, so that matrices can be row reduced by multiplication. This lead to the notion of the inverse of an $n \times n$ matrix A, a matrix B such that $AB = BA = I_n$. We saw $BA = I_n$ is enough to guarantee $AB = I_n$, and also, the invertible $n \times n$ matrices are exactly those of rank n. A key fact is that a square linear system $A\mathbf{x} = \mathbf{b}$ with A invertible has unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

We then introduced matrix groups, or as they are also known, linear groups, and gave several examples. After that, we discussed a way of factoring an invertible matrix as $LPDU$. This is an often used method both in applied mathematics in solving large systems and in pure mathematics in the study of matrix groups.