4 Linear Systems

Revision of Gaussian elimination

In this section we will be looking at how to solve, systematically, systems of linear equations. We begin by looking at some easy examples.

Example 4.1 (a) (Two equations in two unknowns) Find two real numbers x and y which satisfy, simultaneously, the equations

$$x + 3y = 11$$

and

$$2x - y = 1.$$

(b) (Two equations in three unknowns) Find all solutions (if any) to the two simultaneous equations

$$x - 2y + 2z = 5$$

and

$$2x - 4y + 4z = 9.$$

(c) (Three equations in three unknowns) Find all solutions (if any) to the three simultaneous equations

$$x + y + z = -1,$$

$$2x - y + 3z = -10,$$

$$4x + y + 5z = -12.$$

(d) Interpret your answers to (a), (b) and (c) geometrically.

We have now seen that systems of linear equations may have no solutions at all: in this case the system of equations is said to be **inconsistent**. Otherwise, as we will see, there must either be a unique solution or else infinitely many solutions. This fits in with our geometric interpretation (for up to three unknowns) of the sets of solutions in terms of points, lines and planes.

General systems of linear equations

There are usually several different ways to think about a system of linear equations. One way is to think in terms of multiplication of a vector by a matrix.

Example 4.2 Consider the following system of linear equations (two equations in two unknowns).

$$2x + y = -1$$

$$-3x + y = 3.$$

Set $\mathbf{x} = (x, y)$. Find a 2×2 matrix A and a vector $\mathbf{b}$ so that the system of equations above takes the form

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

Find A^{-1} and hence find the solution vector $\mathbf{x}$. Check your answer!

Of course there is no limit to the number of equations or the number of unknowns. When there are n unknowns we will often call them x_1, x_2, $\ldots$, x_n. However when the number of unknowns is small we may also use x, y, z, t etc..

The general system of linear equations has m equations in n unknowns. We call the unknowns x_1, x_2, $\ldots$, x_n. The system of equations then has the following form.

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

Here the unknowns may be regarded as a **vector of unknowns** $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$ (or $\mathbb{C}^n$ if we are working with complex numbers) and the other numbers are constants, which may be real or complex. The constants a_{ij} are the **coefficients** and form an $m \times n$ **coefficient matrix** (or **matrix of coefficients**),

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

The constants $b_1, b_2, \ldots, b_m$ form a **constant vector** (or **vector of constants**)in $\mathbb{R}^m$ (or $\mathbb{C}^m$) and the system of equations may then be rewritten as a matrix equation

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

or more efficiently as

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

We then seek **solution vectors** $\mathbf{x}$ in $\mathbb{R}^n$ (or $\mathbb{C}^n$). The set of all such solution vectors is the **solution set** for the system of equations.

This is quite a useful theoretical approach and we will return to it later. In practise, however, we usually solve systems of equations by Gaussian elimination. To make our calculations more efficient, we may record all the information about the system of equations above in an **augmented matrix**, which has m rows and $n+1$ columns,

$$[A|\mathbf{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}\right] \quad \text{or simply} \quad \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{bmatrix}.$$

The vertical line here is optional but helps us to remember that we are dealing with the augmented matrix corresponding to our system of equations. In a way, all that we are really doing here is saving ourselves some time by leaving out the names of the variables and the + signs.

Example 4.3 For the following system of linear equations (three equations in three unknowns) write down the matrix of coefficients A, the constant vector $\mathbf{b}$ and the augmented matrix $[A|\mathbf{b}]$. (You need not solve the system of equations.)

$$2x_1 - 4x_2 + x_3 = 9,$$
$$-2x_1 + x_2 + 2x_3 = 10,$$
$$x_1 - 3x_2 - 7x_3 = -2.$$

What effects do the following have on the augmented matrix?

(a) Swap the first equation with the third equation.

(b) Add the first equation to the second equation.

(c) Multiply the third equation by -2?

This example shows that we can perform our usual Gaussian elimination by applying certain **elementary row operations** to our augmented matrix to make it simpler. We are allowed to do any one of the following.

- Swap any two rows.
- Multiply any row by a non-zero constant.
- Add any multiple of one of the rows to a different row.

This is rather similar to the way we simplified determinants earlier. Note that when you perform one of these elementary row operations the set of solutions does not change.

Having done one of these elementary row operations we have a new matrix and we can then do further row operations. The idea is, of course, to arrive at a matrix representing a system of equations which is easier to handle.

WARNING! It is very easy to lose information if you perform an incorrect row operation.

Common error 1 If you multiply a row by 0 you end up replacing a sensible equation with the equation $0 = 0$. (This is true but not very useful.)

Common error 2 For safety, row operations should be done one at a time. Let us see what happens if you perform two incompatible row operations simultaneously. Consider the system of two equations in two unknowns

$$x + y = 2,$$
$$2x + 3y = 4.$$

Our augmented matrix for this system is

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & 3 & 4 \end{array}\right].$$

Clearly these two equations give us enough information to determine x and y uniquely.

Suppose we decide to subtract the first row from the second row and simultaneously to subtract the second row from the first row. We would obtain the new augmented matrix

$$\left[\begin{array}{cc|c} -1 & -2 & -2 \\ 1 & 2 & 2 \end{array}\right].$$

However this corresponds to the system of equations

$$-x - 2y = -2,$$
$$x + 2y = 2.$$

Something has gone wrong: these two equations now say the same thing (one is just minus the other) and we end up with a whole line's worth of solutions.

We must have cheated somehow: what did we do wrong? The answer is that we should really perform row operations one at a time. If we start by subtracting the first row from the second row we obtain the matrix

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & 2 & 2 \end{array}\right].$$

If we then subtract the **new** second row from the first, we do not get the matrix $\left[\begin{array}{cc|c} -1 & -2 & -2 \\ 1 & 2 & 2 \end{array}\right]$. Instead we obtain the matrix

$$\left[\begin{array}{cc|c} 0 & -1 & 0 \\ 1 & 2 & 2 \end{array}\right].$$

This is better, though not very systematic! If we convert this back to a system of equations, we obtain

$$-y = 0 \quad \text{and} \quad x + 2y = 2.$$

The first equation gives $y = 0$ and substituting into the second equation we get $x = 2$. Finally we can check, directly, that this really does solve the original system of equations.

Common error 3. Unlike with the calculation of determinants, we are not allowed to perform column operations during Gaussian elimination. (As an exercise you can look at what happens in the example above if you perform one or more column operations.)

As long as we do it without cheating, we can perform several elementary row operations one after the other with the aim of making our matrix simpler to work with.

Definition 4.4 We say that two matrices B and C are **row equivalent to each other** and use the symbol $B \sim C$ if you can get from B to C by performing a sequence of elementary row operations.

In fact row operations are reversible, so whenever $B \sim C$ we also have $C \sim B$.

The following rule summarises what we have said above.

RULE 4.5 Let B be the augmented matrix corresponding to a system of linear equations. Let C be any matrix which is row equivalent to B (i.e. $B \sim C$: you can get from B to C by performing a sequence of elementary row operations). Then C is the augmented matrix corresponding to another system of linear equations, and the two systems of linear equations have exactly the same sets of solutions.

We now look at a systematic way to do Gaussian elimination (in terms of row operations). This may not always give the quickest solution to the problem, but will have the advantage that it will always lead us to the answer. Our plan is to eliminate early variables first, and then to eliminate the later ones. This corresponds to attempting to reduce our matrix to what is called **echelon form**, at which point the system of equations is easy to deal with.

Here are three examples of matrices which are in echelon form: the black line is unnecessary, but is there to help reveal the pattern.

$$\begin{bmatrix} 2 & 1 & 0 & 4 \\ 0 & -3 & 4 & -2 \\ 0 & 0 & 2 & -4 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1-2i & 2 & 2+2i \\ 0 & 0 & 2i & -6 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & -6 \\ 0 & 0 & 0 \end{bmatrix}$$

We will give a formal definition of echelon form matrices in a moment. First let us look at see which systems of linear equations correspond to these matrices.

Example 4.6 Consider each of the three matrices above. Regarding them as augmented matrices (so the last column is the column of constants) write down the corresponding systems of linear equations and find all the solutions (if any).

Now that we have seen that matrices in this form are easy to deal with, here is a definition of echelon form. (Some authors call this form **row echelon form** in order to distinguish it from **column echelon form**.)

Definition 4.7 For a matrix to be in echelon form both of the following must be satisfied:

(a) if there are any rows which are full of zeros, they must all be grouped together at the bottom of the matrix;

(b) for any two of the other rows (i.e. rows which are not full of zeros) the first non-zero entry in the lower row must be (strictly) further to the right than the first non-zero entry in the upper row. (Note that it is not enough for these entries to be in the same column.)

In this case, we call these non-zero entries the **leading** entries (or **distinguished**, or **pivot**) entries of the echelon form matrix, and we call the columns in which these distinguished entries occur the **pivot columns**. We call the corresponding variables (i.e. x_j where $1 \leq j \leq n$ and the jth column is a pivot column) the **leading variables** (or **pivot variables**) and the remaining variables the **free variables**.

Some texts insist that leading entries must be equal to 1 **for the matrix to be in echelon form.** We do not insist on this here, though we do require this for **reduced echelon form** (see below). Of course you can arrange to make the leading entries 1 by simply multiplying the relevant rows by appropriate non-zero constants.

Example 4.8 For each of the following matrices, determine whether or not it is in echelon form. For those which are in echelon form, find the leading entries, the pivot columns, the leading variables and the free variables.

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

Our usual procedure of Gaussian elimination, interpreted in terms of elementary row operations, allows us to **row reduce** any matrix to echelon form and this is how we solve systems of linear equations. Our examples above showed us how we deal with the system of equations corresponding to such a matrix. We may formalise this as follows.

RULE 4.9 Every matrix B is row equivalent to another matrix C which is in echelon form. If the original matrix B is the augmented matrix of a system of linear equations then the solutions of this system, if any, may be easily determined by looking at this echelon form matrix C. (Remember that the solution vectors corresponding to C are the same as those for B.)

(a) If there are any rows of C which have a non-zero entry in the last column but zeros everywhere else then there are no solutions: the system of equations is **inconsistent**.

(b) Otherwise there is at least one solution: the system of equations is **consistent**.

  (i) If all of the variables are leading variables, the solution is unique.

  (ii) If there is at least one free variable then there are infinitely many solutions. The free variables may take any values, and this then determines the values of the leading variables.

Example 4.10 Find all solutions, if any, of the following systems of equations.

(a) (Three equations in three unknowns.)

$$x + 2y - 2z = 3,$$
$$2x + y + 3z = 4,$$
$$4x + 5y - z = 1.$$

(b) (Two equations in four unknowns.)

$$x_1 - 2x_2 + x_3 - x_4 = 3,$$
$$-2x_1 + 3x_2 - x_3 + 2x_4 = 5.$$

In fact the echelon form of the matrix you reach is not unique: we can (though we often do not need to) carry on doing further row operations to simplify the matrix further.

Here are some examples of matrices in what is called **reduced echelon form**.

$$\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 1+3i \\ 0 & 0 & 1 & 2-5i \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

The systems of linear equations corresponding to these augmented matrices are even easier to solve than for augmented matrices in echelon form, but the extra work required to reach this form usually means it is not worth reducing this far. You can probably guess the definition of reduced echelon form from these examples.

Definition 4.11 A matrix A is said to be in **reduced echelon form** if it is in echelon form, all of the leading (pivot) entries are 1 and the only entry in each pivot column which is non-zero is the leading (pivot) entry itself.

Example 4.12 What are the possible echelon forms for a 2×2 matrix? What reduced echelon forms are possible?

We conclude this sub-section by noting some special cases.

RULE 4.13 Suppose you have a system of m linear equations in n unknowns, with coefficient matrix A and constant vector $\mathbf{b}$ (so, as usual, the system of equations has the form $A\mathbf{x} = \mathbf{b}$).

(a) (Homogeneous systems of equations). If $\mathbf{b}$ is the zero vector then the system of equations is said to be homogeneous. In this case the last column of the augmented matrix remains full of zeros throughout row reduction. (You can omit it as long as you do not forget it.) There is always at least one solution to such a system of equations, namely that obtained by setting $\mathbf{x}$ to be the appropriate zero vector.

(b) (The **More Unknowns Theorem**) If there are more unknowns than equations then the augmented matrix has at least two more columns than it has rows. Reducing to echelon form, we see that there must be at least one free variable. As a result, it is impossible for there to be a unique solution: either the system of equations is inconsistent or else there are infinitely many solutions.

(c) Combining (a) and (b), if $\mathbf{b}$ is the zero vector and there are more unknowns than equations then there must be infinitely many solutions to the system of equations.

This last fact is very important. We will see it again when we discuss linear dependence and independence later in the module.

Example 4.14 How many solutions does the following system of linear equations have? (You need not find these solutions.)

$$2x + 3y - 5z + t = 0,$$
$$4x - 5y + z - 3t = 0,$$
$$7x - y + 2z - t = 0.$$

Square matrices: n equations in n unknowns

In the special case where there are the same number as equations as unknowns, the matrix of coefficients A is a square matrix. In this case, A may or may not be invertible.

RULE 4.15 Consider a system of n equations in n unknowns

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$
$$\vdots$$
$$a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n.$$

Let A be the $n \times n$ matrix of coefficients, $\mathbf{b}$ the vector of constants and $\mathbf{x}$ the vector of unknowns, so that the system of equations has the form $A\mathbf{x} = \mathbf{b}$. If the matrix A is invertible, then the system of equations has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

We will consider this situation in more detail later. Note that this rule does not tell you what happens if A is singular: in fact we will see that in that case there may be no solutions or there may be infinitely many solutions.

We saw earlier that Rule 4.15 is not too bad a way to solve 2 equations in 2 unknowns. Unfortunately larger matrices can take a long time to invert, and it is usually better to solve the equations by Gaussian elimination. There is, however, one slight shortcut available which makes the application of Rule 4.15 a little more efficient.

First let us look at 2×2 matrices. Suppose we want to solve $A\mathbf{x} = \mathbf{y}$ where

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

If A is invertible, we know that

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

and so the unique solution is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{\det(A)} \begin{bmatrix} dy_1 - by_2 \\ -cy_1 + ay_2 \end{bmatrix}$$

or in other words,

$$x_1 = \frac{\begin{vmatrix} y_1 & b \\ y_2 & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} \quad \text{and} \quad x_2 = \frac{\begin{vmatrix} a & y_1 \\ c & y_2 \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}.$$

Notice that you find x_j by replacing the jth column of A by the constant vector, and then dividing the determinant of the new matrix by the determinant of A.

As usual, this is no coincidence: by combining Rule 4.15 with Rule 3.60, we obtain the following rule, called **Cramer's rule**.

RULE 4.16 (Cramer's rule) Let A be an invertible $n \times n$ matrix and let $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ be a constant vector. Let C_j be the matrix obtained from A by replacing the jth column of A with the vector $\mathbf{b}$. Then the unique solution vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ to the system of linear equations $A\mathbf{x} = \mathbf{b}$ is given by

$$x_j = \frac{\det(C_j)}{\det(A)}.$$

Remember that this rule only helps if the matrix is invertible, and even then for large matrices the method is rather slow.

Example 4.17 Use Cramer's rule to solve each of the following systems of linear equations. Check your answers!

(a) (Two equations in two unknowns)

$$2x_1 - x_2 = -1,$$
$$x_1 + x_2 = 10.$$

(b) (Three equations in three unknowns)

$$2x_1 + x_2 + x_3 = 8,$$
$$x_1 + x_2 - x_3 = -1,$$
$$x_1 + 2x_2 + x_3 = 1.$$

Gauss–Jordan inversion of a matrix

In the previous chapter we used determinants to determine whether or not a matrix was singular, and to find the inverses of invertible matrices. We now look at the Gauss–Jordan method, which is based on row operations.

Example 4.18 Consider the matrix $A = \begin{bmatrix} 1 & 1 \\ 4 & 3 \end{bmatrix}$, and let $I = I_2$ be the 2×2 identity matrix. Form the augmented matrix

$$B = [A \,|\, I] = \left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 4 & 3 & 0 & 1 \end{array}\right].$$

Using row operations, find the reduced echelon form of the augmented matrix B. This reduced echelon form matrix has the form

$$[I \,|\, C]$$

for some 2×2 matrix C. What is the connection between C and A?

This example illustrates the Gauss–Jordan method. More generally we may determine whether an $n \times n$ matrix is invertible and, if so, find its inverse by the same technique.

RULE 4.19 Let A be an $n \times n$ matrix and let $I = I_n$ (the $n \times n$ identity matrix). Form the augmented matrix $B = [A \,|\, I]$. Using row operations, find the reduced echelon form of the augmented matrix B. If the reduced echelon form matrix has the form $[I \,|\, C]$ for some $n \times n$ matrix C, then A is invertible and $A^{-1} = C$. Otherwise, A is singular.

Example 4.20 Apply the Gauss–Jordan method to the following matrices to determine whether or not they are invertible, and to find inverses for those which are invertible:

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

We conclude this chapter with some rules connecting together all the information we have about systems of n equations in n unknowns and the corresponding $n \times n$ coefficient matrices.

RULE 4.21 Let A be an $n \times n$ matrix. The following statements are equivalent. (This means that if any of the following statements are true, then so are all the others.)

(a) The matrix A is invertible.

(b) The determinant of A is not equal to zero.

(c) The matrix A is row equivalent to the $n \times n$ identity matrix I_n.

(d) The equation

$$A\mathbf{x} = \mathbf{0}$$

(which corresponds to a homogeneous system of n equations in n unknowns) has the **unique** solution $\mathbf{x} = \mathbf{0}$.

(e) For every $n \times 1$ column vector of constants $\mathbf{b}$, the equation

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

has exactly one solution.

(f) For every $n \times 1$ column vector of constants $\mathbf{b}$, the equation

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

has **at least one** solution.

If we negate all the statements above, we obtain our final rule of this chapter.

RULE 4.22 Let A be an $n \times n$ matrix. The following statements are equivalent.

(a) The matrix A is singular.

(b) The determinant of A is 0.

(c) The matrix A is not row equivalent to the $n \times n$ identity matrix I_n.

(d) The equation

$$A\mathbf{x} = \mathbf{0}$$

has at least one non-zero solution $\mathbf{x}$. (It then follows immediately that this equation has infinitely many solutions.)

(e) There is at least one $n \times 1$ column vector of constants $\mathbf{b}$, so that the equation

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

has either no solutions at all or else has more than one solution. (From earlier results, in the latter case there must be infinitely many solutions.)

(f) There is at least one $n \times 1$ column vector of constants $\mathbf{b}$, so that the equation

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

has no solutions at all.