Chapter 2

Introductory Examples

We begin by investigating the equation

$$ax = b \tag{2.1}$$

where a, b are given real numbers.

There are two cases to consider: (i) $a = 0$ and (ii) $a \neq 0$.

(i) Since $a = 0$, $ax = 0x = 0$ for every real number x. Thus *every* real number x is a solution if $b = 0$ whereas *no* real number x is a solution if $b \neq 0$.

(ii) Since $a \neq 0$, $x = \frac{b}{a}$ is the *unique* solution.

Thus, (2.1) has a unique solution if and only if $a \neq 0$. Otherwise, it has either no solutions or infinitely many solutions, according as $b \neq 0$ or $b = 0$.

We now consider the system of equations

$$ax + by = e \tag{2.2a}$$
$$cx + dy = f \tag{2.2b}$$

A solution of (2.2) consists of a pair of real numbers (x, y) such that both (2.2a) and (2.2b) are satisfied.

In order to try to find all solutions of (2.2), we try to replace (2.2a) and (2.2b) by equations of the form (2.1) which, taken together, have the same solutions as (2.2).

If we multiply (2.2a) by d and subtract b times (2.2b), as well as subtracting c times (2.2b) from a times (2.2a) we obtain

$$(ad - bc)x = (ed - bf) \tag{2.3a}$$
$$(ad - bc)y = (af - ec) \tag{2.3b}$$

Thus, if $ad - bc \neq 0$, we obtain the unique solution

$$\left(\frac{ed - bf}{ad - bc}, \frac{af - ec}{ad - bc}\right). \tag{2.4}$$

Direct substitution verifies that (2.4) solves our system of equations.

If, on the other hand, $ad - bc = 0$, then there is no solution whatsoever if either $ed - bf \neq 0$ or $af - ec \neq 0$, and every pair of real numbers (x, y) is a solution if both $ed - bf = 0$ and $af - ec = 0$.

The manner in which we derived (2.3) from (2.2) makes it clear that any solution of (2.2) is also a solution of (2.3). It follows that if $ad - bc = 0$, and either $ed - bf \neq 0$ or $af - ec \neq 0$, then (2.2) has no solution, for then (2.3) has none.

The situation is more delicate when $ad - bc = ed - bf = af - ec = 0$, for it is then possible that some solutions of (2.3) are not solutions of (2.2), as the next example shows.

Example 2.1. Take $a = 1, b = -1, c = d = e = f = 0$. Then (2.2) becomes

$$x - y = 0 \tag{2.5a}$$
$$0x + 0y = 0 \tag{2.5b}$$

and the complete set of solutions is plainly the set of all pairs of real numbers of the form (x, x). But (2.3) becomes

$$0x = 0 \tag{2.6a}$$
$$0y = 0 \tag{2.6b}$$

which is solved by any pair of real numbers (x, y). So, $(1, 0)$ solves (2.6) without solving (2.5).

Observation 2.2. Whether (2.2) has a unique solution is determined by $ad - bc$. For this reason $ad - bc$ is known as the *determinant* of the system of equations (2.2).

We continue our investigation of (2.2). At a number of places above, we claimed to have found all possible solutions. How can we be sure?

We address this question.

Suppose that (x_1, y_1) and (x_2, y_2) both solve (2.2). Then

$$a(x_1 - x_2) + b(y_1 - y_2) = (ax_1 + by_1) - (ax_2 + by_2) = e - e = 0 \tag{2.7a}$$
$$c(x_1 - x_2) + d(y_1 - y_2) = (cx_1 + dy_1) - (cx_2 + dy_2) = f - f = 0 \tag{2.7b}$$

Thus, any two solutions of (2.2) differ by a solution of

$$ax + by = 0 \tag{2.8a}$$
$$cx + dy = 0 \tag{2.8b}$$

Hence the general solution of (2.2) can be found by adding to (x_h, y_h), the general solution of (2.8), any one solution, (x_s, y_s), of (2.2). We therefore investigate (2.8).

Unlike (2.2), the system of equations (2.8) always has at least one solution, namely the *trivial* solution $x = 0, y = 0$.

Suppose that (x_1, y_1) and (x_2, y_2) are solutions of (2.8), and that λ, μ are real numbers. Then

$$a(\lambda x_1 + \mu x_2) + b(\lambda y_1 + \mu y_2) = \lambda(ax_1 + by_1) + \mu(ax_2 + by_2) = \lambda 0 + \mu 0 = 0$$
$$c(\lambda x_1 + \mu x_2) + d(\lambda y_1 + \mu y_2) = \lambda(cx_1 + dy_1) + \mu(cx_2 + dy_2) = \lambda 0 + \mu 0 = 0,$$

so that $(\lambda x_1 + \mu x_2, \lambda y_1 + \mu y_2)$ is again a solution of (2.8).

If we define an "addition" of solutions to (2.8) by

$$(x_1, y_1) \boxplus (x_2, y_2) := (x_1 + x_2, y_1 + y_2),$$

and if we define a "multiplication" by real numbers of solutions to (2.8) by

$$\lambda \boxdot (x, y) := (\lambda x, \lambda y)$$

then "adding" any two solutions of (2.8) yields a solution of (2.8), and "multiplying" a solution of (2.8) by a real number yields a solution of (2.8).

Systems of equations like (2.8) are called *homogeneous*. They are, of course, just the special case of (2.2) where $e = f = 0$. In particular, if in (2.2) either $e \neq 0$ or $f \neq 0$, then we call the corresponding system (2.8) the *associated homogeneous system*.

This example, and the features just highlighted, motivate these notes: Linear algebra may be thought of, in the first instance, as the systematic study of such systems of equations, their solutions and the transformations these admit.

The solutions of homogeneous systems of linear equations are elementary examples of *vector spaces*. Of course they are not the only examples.

The vectors of physics, such as force, also provide example, as the language suggests: the "sum" of two forces acting simultaneously is their resultant force, and multiplication by a real number corresponds to scaling the force.

Binary computer code is another example of a vector space, a point of view which finds application in theoretical computer science.

Solutions to specific systems of differential equations also form vector spaces.

Vector spaces also appear in number theory in several places, including the study of field extensions, and form the basis from which the important algebraic notion of *module* has been abstracted.

Finally, vector spaces, particularly inner product spaces, are central to the study of statistics and geometry.

Before launching into the formal study of linear algebra, we illustrate how linear algebra can be applied to solving *linear difference equations with constant coefficients*, by writing such difference equations in terms of *matrices*. This not only provides an application of linear algebra and its techniques, but also provides motivation for deeper investigation.

We begin by recalling the definition of a linear difference equation over $\mathbb{R}$.

Definition 2.3. Let $(x_n)_{x\in\mathbb{N}}$ be a sequence of real numbers. A *linear difference equation of degree k with constant coefficients* is an equation of the form

$$x_{n+k} + a_{k-1}x_{n+k-1} + \cdots + a_1 x_{n+1} + a_0 x_n = g(n), \tag{2.10}$$

where each $a_j \in \mathbb{R}$ and $g(n)$ is a function of $n \in \mathbb{N}$.

Example 2.4. Consider the difference equation

$$x_{n+2} - 4x_{n+1} + 3x_n = 0,$$

or $x_{n+2} = 4x_{n+1} - 3x_n$. We can represent this by the matrix equation

$$\begin{bmatrix} x_{n+2} \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix},$$

from which we deduce, by induction, that

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} x_1 \\ x_0 \end{bmatrix}.$$

Example 2.5. Consider the difference equation

$$x_{n+2} - 4x_{n+1} + 4x_n = 0,$$

or $x_{n+2} = 4x_{n+1} - 4x_n$. We can represent this by the matrix equation

$$\begin{bmatrix} x_{n+2} \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix},$$

from which we deduce, by induction, that

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} x_1 \\ x_0 \end{bmatrix}.$$

Example 2.6. Consider the difference equation

$$x_{n+2} - 4x_{n+1} + 5x_n = 0,$$

or $x_{n+2} = 4x_{n+1} - 5x_n$. We can represent this by the matrix equation

$$\begin{bmatrix} x_{n+2} \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix},$$

from which we deduce, by induction, that

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = \begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix}^n \begin{bmatrix} x_1 \\ x_0 \end{bmatrix}.$$

Thus, we have reduced the solving of these difference equations to "merely" computing, respectively

$$\begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix}^n, \quad \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix}^n \quad \text{and} \quad \begin{bmatrix} 4 & -5 \\ 1 & 0 \end{bmatrix}^n$$

Observation 2.7. These examples can, of course, be generalised to find the general solution of Equation (2.10) when $g(n) = 0$ for all $n \in \mathbb{N}$. We do not pursue this greater generality here, since our is merely to present some concrete examples. The general case is a simple application of our later theory, which is, in any event, needed for the analysis of the general case.

We have an explicit formula for computing the product of two matrices:

$$\text{If } \underline{\mathbf{A}} = [a_{ij}]_{m\times p} \text{ and } \underline{\mathbf{B}} = [b_{jh}]_{p\times q}, \text{ then } \underline{\mathbf{A}}\,\underline{\mathbf{B}} = [c_{ih}]_{m\times q}, \text{ where } c_{ih} := \sum_{j=1}^{p} a_{ij}b_{jh}.$$

This provides a recursive formula for $\underline{\mathbf{A}}^n$ $(n \in \mathbb{N})$: Writing $\underline{\mathbf{A}}^n := [a_{ij}^{(n)}]_{k\times k}$,

1. $a_{ij}^{(0)} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$ as $\underline{\mathbf{A}}^0 := \underline{\mathbf{1}}_k$
2. $a_{ij}^{(n+1)} = \sum_{h=1}^{k} a_{ih}a_{hj}^{(n)}$

While it is comforting to have a recursive formula and so be able to use a programmable calculator or computer for the actual calculation, it is easy to see that this is neither an efficient nor an insightful way to proceed.

Example 2.8. Try to compute the onethousandth power of $\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$ by this recursive procedure.

Plainly all the coefficients are positive integers and $a_{ij}^{(n+1)} > a_{ij}^{(n)}$. But little more can be said!

Even when you have used the recursive formula to complete such a calculation, you are unlikely to guess any formula for calculating the $a_{ij}^{(n)}$'s directly for $n > 2$. On the other hand, using the theory developed during this course, you will be able to see that for the matrices above, we have the explicit formulæ below.

Example 2.4 Since

$$\begin{bmatrix}4 & -3\\1 & 0\end{bmatrix}^n = \frac{1}{2}\begin{bmatrix}3 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}3 & 0\\0 & 1\end{bmatrix}^n\begin{bmatrix}1 & -1\\-1 & 3\end{bmatrix} = \frac{1}{2}\begin{bmatrix}3 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}3^n & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 3\end{bmatrix},$$

$$a_{11}^{(n)} = \frac{1}{2}(3^{n+1}-1) \qquad a_{12}^{(n)} = \frac{1}{2}(3-3^{n+1})$$

$$a_{21}^{(n)} = \frac{1}{2}(3^n-1) \qquad a_{22}^{(n)} = \frac{1}{2}(3-3^n).$$

Example 2.5 Since

$$\begin{bmatrix}4 & -4\\1 & 0\end{bmatrix}^n = \begin{bmatrix}2 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}2 & 1\\0 & 2\end{bmatrix}^n\begin{bmatrix}0 & 1\\1 & 2\end{bmatrix} = \begin{bmatrix}2 & 1\\1 & 0\end{bmatrix}\begin{bmatrix}2^n & n2^{n-1}\\0 & 2^n\end{bmatrix}\begin{bmatrix}0 & 1\\1 & 2\end{bmatrix},$$

$$a_{11}^{(n)} = (n+1)2^n \qquad a_{12}^{(n)} = -n2^{n+1}$$

$$a_{21}^{(n)} = n2^{n-1} \qquad a_{22}^{(n)} = (1-n)2^n = -(n-1)2^n.$$

Example 2.6 Since

$$\begin{aligned}\begin{bmatrix}4 & -5\\1 & 0\end{bmatrix}^n &= \frac{-i}{2}\begin{bmatrix}2+i & 2-i\\v1 & 1\end{bmatrix}\begin{bmatrix}2+i & 0\\0 & 2-i\end{bmatrix}^n\begin{bmatrix}1 & -2+i\\-1 & 2+i\end{bmatrix}\\ &= \frac{-i}{2}\begin{bmatrix}2+i & 2-i\\1 & 1\end{bmatrix}\begin{bmatrix}(2+i)^n & 0\\0 & (2-i)^n\end{bmatrix}\begin{bmatrix}1 & -2+i\\-1 & 2+i\end{bmatrix}\end{aligned}$$

where $i = \sqrt{-1}$,

$$a_{11}^{(n)} = \frac{-i}{2}\left((2+i)^{n+1}-(2-i)^{n+1}\right) \qquad a_{12}^{(n)} = \frac{-5i}{2}\left((2+i)^n-(2-i)^n\right)$$

$$a_{21}^{(n)} = \frac{-i}{2}\left((2+i)^n-(2-i)^n\right) \qquad a_{22}^{(n)} = \frac{-5i}{2}\left((2+i)^{n-1}-(2-i)^{n-1}\right).$$

Example 2.8 Since

$$\begin{bmatrix}1 & 2\\2 & 5\end{bmatrix}^n = \frac{\sqrt{2}-1}{2\sqrt{2}}\begin{bmatrix}\sqrt{2}+1 & 1\\-1 & \sqrt{2}+1\end{bmatrix}\begin{bmatrix}\left(3-2\sqrt{2}\right)^n & 0\\0 & \left(3+2\sqrt{2}\right)^n\end{bmatrix}\begin{bmatrix}\sqrt{2}+1 & -1\\1 & \sqrt{2}+1\end{bmatrix},$$

$$a_{11}^{(n)} = a\frac{1}{2\sqrt{2}}\left[\left(\sqrt{2}+1\right)\left(3-2\sqrt{2}\right)^n+\left(\sqrt{2}-1\right)\left(3+2\sqrt{2}\right)^n\right]$$

$$a_{12}^{(n)} = \frac{1}{2\sqrt{2}}\left[\left(3+2\sqrt{2}\right)^n-\left(3-2\sqrt{2}\right)^n\right]$$

$$a_{21}^{(n)} = \frac{1}{2\sqrt{2}}\left[\left(3+2\sqrt{2}\right)^n-\left(3-2\sqrt{2}\right)^n\right]$$

$$a_{22}^{(n)} = \frac{1}{2\sqrt{2}}\left[\left(\sqrt{2}-1\right)\left(3-2\sqrt{2}\right)^n+\left(\sqrt{2}+1\right)\left(3+2\sqrt{2}\right)^n\right].$$

It is difficult to envisage how anyone could have guessed these formulæ!

But the theory we develop will make them obvious.

Observation 2.9. Whereas all the coefficients of each of the matrices $\underline{\mathbf{A}}^n$ are plainly integers, and obviously all positive in the last example, the formulæ for $a_{ij}^{(n)}$ as function of n may be replete with negative numbers, fractions, irrational numbers and even complex numbers, making it difficult to divine from the explicit formulæ the fact that the coefficients are always integers, possibly even positive ones!

This illustrates a recurring theme in mathematics: In order to solve problems which are simple to express, it is frequently necessary to go beyond the terms in which the problem is expressed, to a deeper or more abstract level, in order to find a solution. We seem to have made the problems more complicated. But this has made them easier to solve!

Perhaps the most striking recent example of this is Andrew Wiles' proof in 1995 of Fermat's Last Theorem:

> The integer equation $x^n + y^n = z^n$, with $x, y, z \neq 0$, has no solution if $n > 2$.

This was enunciated in 1657, but no proof was known until Andrew Wiles' work in the 1990s! While the problem is simple to express and understand — a student in the first year of high school can begin to work on it — its proof by Andrew Wiles depends upon results drawn from algebraic topology, algebraic geometry and other fields of mathematics.

2.1 Exercises

Exercise 2.1. Solve the following system of equations, where the solutions are to be real numbers.

$$\begin{array}{rcrcrcr} x & + & 7y & + & 4z & = & 21 \\ 3x & - & 6y & + & 5z & = & 2 \\ 5x & + & y & - & 3z & = & 14 \end{array}$$

Exercise 2.2. Let $f, g : \mathbb{R} \to \mathbb{R}$ be real valued functions of the real variable x. Show that if $y = f(x)$ and $y = g(x)$ both satisfy the differential equation

$$\frac{d^2y}{dx^2} - 2\frac{dy}{dx} - 3y = 0$$

and if λ, μ are any real numbers, then the function $h : \mathbb{R} \to \mathbb{R}$ defined by

$$h(x) := \lambda f(x) + \mu g(x)$$

for all $x \in \mathbb{R}$, must also satisfy the given differential equation.

Exercise 2.3. Solve the system of differential equations

$$\begin{aligned} x'(t) - 2y'(t) &= x(t) \\ x'(t) + y'(t) &= y(t) + x(t) \end{aligned}$$

where $x(t)$ and $y(t)$ denote real valued functions of the real variable t, and $'$ stands for the derivative.

Exercise 2.4. Find all integral matrices $\underline{\mathbf{A}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ satisfying $\underline{\mathbf{A}}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Exercise 2.5. Let $\underline{\mathbf{A}} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$. Prove that for $n \in \mathbb{N}$, $n \geq 1$, $\underline{\mathbf{A}}^n = \begin{bmatrix} a^n & 0 \\ 0 & b^n \end{bmatrix}$.

Exercise 2.6. (i) Let $\underline{\mathbf{A}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. Prove that for $n \in \mathbb{N}$, $\underline{\mathbf{A}}^n = \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix}$.

(ii) For $\underline{\mathbf{A}} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$, find $\underline{\mathbf{A}}^n$.

Exercise 2.7. For each of the following matrices $\underline{\mathbf{A}}$, find $\underline{\mathbf{A}}^n$.

(i) $\underline{\mathbf{A}} = \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}$.

(ii) $\underline{\mathbf{A}} = \begin{bmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}$.

(iii) $\underline{\mathbf{A}} = \begin{bmatrix} a & 1 & 0 & 0 \\ 0 & a & 1 & 0 \\ 0 & 0 & a & 1 \\ 0 & 0 & 0 & a \end{bmatrix}$.

(iv) $\underline{\mathbf{A}} = a\underline{\mathbf{1}}_k + \underline{\mathbf{N}}_k$, where $\underline{\mathbf{N}}_k := [x_{ij}]_{k\times k}$ is given by $x_{ij} = \begin{cases} 1 & \text{if } j = i+1 \\ 0 & \text{otherwise.} \end{cases}$