Mathematics for Computer Science

Paul D. Mitchener

February 13, 2013

Contents

1 The Integers

1.1 Integers, Factors, and Primes

In this chapter, we study the set of *integers*, $\mathbb{Z}$, which is the set of numbers

$$\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

We use the rules of addition and multiplication without comment, as well as statements such as $a < b$, meaning an integer a is less than an integer b, and $a \leq b$, meaning an integer a is less than or equal to an integer b.

For an integer $a \in \mathbb{Z}$, we define the *absolute value*, $|a|$, to be the positive number we get by ignoring the sign. So, for example, $|4| = 4$ and $|-3| = 3$.

Our main goal in this chapter is to look in detail at how division works for integer numbers. This is useful in computer arithmetic and codes. We look at some of these applications in the next chapter.

Definition 1.1 Let $a, b \in \mathbb{Z}$. We say that a is a *factor* of b if we can write $a = kb$ for some other integer $k \in \mathbb{Z}$.

Example 1.2

- $6 = 3 \times 2$, so 3 and 2 are factors of 6.
- There is no integer n for which $6 = 4n$, so 4 is *not* a factor of 6.
- For any integer $n \in \mathbb{Z}$, $n = 1n$, so 1 and n are factors of n.

Algorithmically, a is a factor of b if the remainder when a is divided by b is 0. We return to this point of view in subsection 1.3.

Definition 1.3 A *prime number* if a positive integer $p \geq 2$ whose only factors are p and 1.

Note that we only consider positive numbers in this definition, and that we specifically exclude 1 from being a prime number.

Example 1.4

- The only factors of 5 are 5 and 1, so 5 is a prime number.
- $6 = 3 \times 2$, so 3 and 2 are factors of 6, and 6 is not a prime number.

To check whether a number, n, is prime, we can look at every positive integer $2 \leq k < n$ and check if it is a factor. If no number $2 \leq k < n$ is a factor of n, then n is a prime number.

This method is costly in computer time for large numbers. There are better ways of checking, but there is no known algorithm that is both efficient and always works. Such an algorithm would provide an efficient way of cracking codes, as we shall see later on.

Definition 1.5 Let $a, b \in \mathbb{Z}$. We call an integer $h \in \mathbb{Z}$ a *highest common factor* of a and b, and write $h = hcf(a, b)$ if:

- h is a factor of a, and h is a factor of b.
- If k is also a factor of both a and b, then k is a factor of h.

The highest common factor of integers a and b is also sometimes called the *greatest common divisor* of a and b. However, we only use the term highest common factor in these notes.

Example 1.6 The integer 4 is a highest common factor of both 12 and 8. So is -4.

Finding highest common factors of large numbers is both useful and hard. We will look at how to find highest common factors algorithmically later in this chapter.

1.2 Proofs by Induction

Recall the *principle of induction*:

Consider a statement $P(n)$ involving the integer n. *If* we know that:

- $P(1)$ is true;
- For every integer $k \geq 1$, $P(k)$ implies $P(k+1)$;

then $P(n)$ is true for every integer $n \geq 1$.

Here, the number 1 could be replaced by 0, or for that matter any other integer. Let us give an example.

Proposition 1.7 *The sum*

$$1 + 2 + 4 + \cdots + 2^n$$

is equal to $2^{n+1} - 1$.

Proof: Firstly, recall the mathematical notation for sums: we write

$$1 + 2 + 4 + \cdots + 2^n = \sum_{j=0}^{n} 2^j$$

(here we start with $k = 0$, as $2^0 = 1$).

To prove the statement in the proposition, we work by induction. When $n = 0$, the sum is

$$\sum_{j=0}^{0} 2^j = 1.$$

On the other hand, in this case,

$$2^{n+1} - 1 = 2 - 1 = 1$$

so the result is true when $n = 0$.

Suppose the result is true when $n = k$, where $k \geq 0$. Then

$$\sum_{j=0}^{k} 2^j = 2^{k+1} - 1.$$

But

$$\sum_{j=0}^{k+1} 2^j = 1 + 2 + 4 + \cdots + 2^k + 2^{k+1} = \sum_{j=0}^{k} 2^j + 2^{k+1}$$

so, using the result for $n = k$, we have

$$\sum_{j=0}^{k+1} 2^j = 2^{k+1} - 1 + 2^{k+1} = 2(2^{k+1}) - 1 = 2^{k+2} - 1$$

so the result is true when $n = k + 1$.

Hence the result is true for all $n \geq 0$ by induction. □

The following variation of the principle of induction is sometimes called *the principle of strong induction*:

Consider a statement $P(n)$ involving the integer n. *If* we know that:

- $P(1)$ is true;
- For every integer $k > 1$, $P(1), \ldots, P(k-1)$ all together imply $P(k)$;

then $P(n)$ is true for every integer $n \geq 1$.

Again, the number 1 could be replaced by 0, or for that matter any other integer.

Theorem 1.8 *Every integer $n \geq 2$ can be expressed as a product of one or more prime numbers.*

Proof: We work by strong induction.

The number 2 is itself a prime number, so the result is true for $n = 2$.

Let $k > 2$. Suppose that the result is true whenever $n < k$.

If k is a prime number, then the result is also true for k. If k is not a prime number, we can write

$$k = ab$$

where $2 \leq a < k$ and $2 \leq b < k$. Since a and b are both less than k, they can be written as a product of prime numbers. Hence $k = ab$ is also a product of prime numbers, and the result is true when $n = k$.

Therefore, by the principle of strong induction, the result is true for all $n \geq 2$.
□

It is in fact true that any integer $n \geq 2$ can be written *uniquely* as a product of prime numbers, although we have not proved uniqueness here. This result (with uniqueness) is known as the *fundamental theorem of arithmetic.*

1.3 Division

The following result is known as the *division algorithm.* It's not really an algorithm, but is a step in a number of different algorithms, as we shall see.

Theorem 1.9 *Let $a, b \in \mathbb{Z}$, with $b \neq 0$. Then we have a unique* quotient $q \in \mathbb{Z}$ *and* remainder $r \in \mathbb{Z}$ *where $0 \leq r < |b|$ such that*

$$a = qb + r.$$

□

We won't prove this result here. To understand it, note that the quotient q is the number we get when we divide a by b; the number r is the remainder. Observe that b is a factor of a if and only if the above remainder r is zero.

Note that both the quotient and remainder are integers.

Example 1.10 If we divide 12 by 5, we get 2, with remainder 2. In terms of the division algorithm, we have

$$12 = 2 \times 5 + 2.$$

If we divide 13 by -4, we get -3, with remainder 1. In terms of the division algorithm, we have

$$13 = (-3) \times (-4) + 1.$$

Our next result uses a technique called *proof by contradiction.* In a proof by contradiction, to prove a statement P, assume that P is not true. We use the fact that P is not true to arrive at something that is clearly nonsense- a contradiction. This contradiction means that the thing we *assumed* is false. Thus the statement that P is not true is false, ie: P is true.

Theorem 1.11 *There are infinitely many prime numbers.*

Proof: We prove this result by contradiction. Assume that there are only *finitely many* prime numbers. Write them in a list

$$p_1, p_2, \ldots, p_n.$$

Let

$$N = p_1 p_2 \cdots p_n + 1.$$

By the above theorem, N must have a factor which is a prime number (perhaps N itself). By our assumption, the only prime numbers are the ones in our list, $p_1, p_2, \ldots, p_n$. By the division algorithm, if we divide N by one of these numbers p_i, we have remainder 1, so no p_i is a factor of N.

This is a contradiction of our assumption that $p_1, p_2, \ldots, p_n$ are the only prime numbers. So there must be infinitely many prime numbers. □

1.4 Euclid's Algorithm

We now present an algorithm to find the highest common factor of two integers. It is called *Euclid's algorithm.*

Theorem 1.12 *Let $a, b \in \mathbb{Z}$. Write, with the division algorithm*

$$\begin{aligned}
a &= q_0 b + r_0 \\
b &= q_1 r_0 + r_1 \\
r_0 &= q_2 r_1 + r_2 \\
r_1 &= q_3 r_2 + r_3 \\
\vdots & \quad \vdots \\
r_{n-2} &= q_n r_{n-1} + r_n \\
r_{n-1} &= q_{n+1} r_n.
\end{aligned}$$

where $0 \leq r_0 < |b|$, and $0 \leq r_k < |r_{k-1}|$ for all k.

Then $r_n = hcf(a, b)$.

To clarify, we keep using the division algorithm, using the above pattern, until we get a zero remainder. The last non-zero remainder is a highest common factor of a and b.

If there is no remainder in the first step of the algorithm, we have $a = q_0 b$, meaning b is a factor of a, so the highest common factor of a and b is b itself.

We will come back to the proof. Before we do this, the most interesting thing is to see an example.

Example 1.13 Find the highest common factor of 630 and 132.

Solution: With repeated use of the division algorithm:

$$\begin{aligned}
630 &= 4 \times 132 + 102 \\
132 &= 1 \times 102 + 30 \\
102 &= 3 \times 30 + 12 \\
30 &= 2 \times 12 + 6 \\
12 &= 2 \times 6.
\end{aligned}$$

The last non-zero is remainder is 6. So $6 = hcf(630, 132)$.

To see why Euclid's algorithm works, we need the following lemma.

Lemma 1.14 *Let $a = qb + r$. Then $hcf(a, b) = hcf(b, r)$.*

Proof: Let $h = hcf(b, r)$. We need to show that h is a highest common factor of a and b.

- We have $b = mh$ and $q = nh$ for some $m, n \in \mathbb{Z}$. Hence

  $$a = qmh + nh = (qm + n)h$$

  meaning h is a factor of a. We already know h is a factor of b.

- Let c be a factor of both a and b. Write $a = kc$ and $b = lc$. Then the equation $a = qb + r$ gives us

  $$mc = qnc + r$$

  that is to say

  $$r = (m - qn)c.$$

  Thus c is also a factor of r. As c is a factor of b, c is a factor of h.

From this, it follows that $h = hcf(a, b)$.

□

Why Euclid's algorithm works:

Let $a, b \in \mathbb{Z}$. Write

$$\begin{aligned} a &= q_0 b + r_0 \\ b &= q_1 r_0 + r_1 \\ r_0 &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{n-2} &= q_n r_{n-1} + r_n \\ r_{n-1} &= q_{n+1} r_n. \end{aligned}$$

Then by the above,

$$hcf(a, b) = hcf(b, r_0) = hcf(r_0, r_1) = \cdots = hcf(r_{n-1}, r_n) = hcf(r_n, 0).$$

Now, notice that:

- $r_n = 1r_n$, and $0 = 0r_n$, so r_n is a factor of both r_n and 0.
- Let c be a factor of both r_n and 0. Then c is certainly a factor of r_n.

Hence $hcf(r_n, 0) = r_n$. □

1.5 Linear Diophantine Equations

In this section we look at solving equations of the form

$$ax + by = c$$

where:

- $a, b, c \in \mathbb{Z}$, and at least one of a and b is non-zero.
- We only look for solutions where $x, y \in \mathbb{Z}$.

Sometimes we can find solutions to such equations just by trying out different numbers. This is only recommended if a, b, c are all quite small.

Example 1.15 Find a solution to the equation

$$3x + 5y = 1.$$

Solution: We observe that $2 \times 5 = 10$, which is one more than 3×3. So

$$3 \times (-3) + 5 \times 2 = 1$$

and we can take $x = -3$ and $y = 2$.

For bigger numbers, we want a more systematic method. First we look at the special case where $c = hcf(a, b)$. In this case, we can find $x, y \in \mathbb{Z}$ such that $ax + by = c$. This is best illustrated using an example.

Example 1.16 Find a solution to the equation

$$630x + 132y = 6$$

where $x, y \in \mathbb{Z}$.

Solution: To proceed, one first step is to use Euclid's algorithm:

$$\begin{aligned} 630 &= 4 \times 132 + 102 \\ 132 &= 1 \times 102 + 30 \\ 102 &= 3 \times 30 + 12 \\ 30 &= 2 \times 12 + 6 \\ 12 &= 2 \times 6. \end{aligned}$$

We then look at what we have done and go backwards:

$$\begin{aligned} 6 &= 30 - 2 \times 12 \\ &= 30 - 2 \times (102 - 3 \times 30) \\ &= 7 \times 30 - 2 \times 102 \\ &= 7 \times (132 - 102) - 2 \times 102 \\ &= 7 \times 132 - 9 \times 102 \\ &= 7 \times 132 - 9 \times (630 - 4 \times 132) \\ &= -9 \times 630 + 43 \times 132 \end{aligned}$$

So we have a solution $x = -9$ and $y = 29$. It is worth quickly checking what we have found works using a calculator.

Note that in general we will have more than one possible solution. It is also possible to have no solutions.

Theorem 1.17 *Let $a, b, c \in \mathbb{Z}$, with at least one of a and b non-zero. Let $h = hcf(a, b)$. Then the equation*

$$ax + by = c$$

has an integer solution if and only if h is a factor of c.

Proof: By the above, there are integers $s, t \in \mathbb{Z}$ such that

$$as + bt = h$$

Further, we have $c = hk$, where $k \in \mathbb{Z}$. Let $x = ks$ and $y = kt$. Then

$$ax + by = aks + bkt = k(as + bt) = kh = c$$

so in this case we have an integer solution.

On the other hand, suppose we have $s, t \in \mathbb{Z}$ such that

$$as + bt = c.$$

Then $a = kh$ and $b = lh$, where $k, l \in \mathbb{Z}$, so

$$c = khs + lht = h(ks + lt)$$

and h is a factor of c. □

In particular, by the above, if h is *not* a factor of c, then the equation $ax + by = c$ has *no* integer solutions.

Example 1.18 Consider the equation

$$12x + 8y = 1.$$

Then the highest common factor of 12 and 8 is 4, and 4 is not a factor of of 1, so the equation has no integer solutions by the above.

We can also see this directly. The number on the left of the equation is always even (indeed, even a multiple of 4), whereas 1 is odd.

Now, if the equation $ax + by = c$ has any integer solution at all, it will have *lots* of integer solutions. We can find all solutions using the following result.

Theorem 1.19 *Let $a, b \in \mathbb{Z}$, with a and b not both zero. Let $h = hcf(a, b)$ Let $s, t \in \mathbb{Z}$ be such that $as + bt = h$.*

Then a pair of integers (x, y) is a solution to the equation $ax+by = h$ if and only if we have $k \in \mathbb{Z}$ such that

$$x = s + \frac{kb}{h} \qquad y = t - \frac{ka}{h}.$$

Proof: Since h is a factor of a and b, the expressions $\frac{kb}{h}$ and $\frac{ka}{h}$ are integers. We know that $as + bt = h$.

Let x and y be as above. Observe

$$ax + by = as + bt + \frac{kab}{h} - \frac{kab}{h} = as + bt = h.$$

Conversely, suppose $ax + by = h$. We also know that $as + bt = h$, so

$$a(x - s) + b(y - t) = 0.$$

Since h is the highest common factor of a and b, a/h and b/h are integers, and the highest common factor of $\frac{a}{h}$ and $\frac{b}{h}$ is 1.

Rearranging

$$\frac{a}{h}(x - s) = \frac{b}{h}(t - y)$$

so $\frac{a}{h}$ is a factor of $\frac{b}{h}(t - y)$.

Since the highest common factor of $\frac{a}{h}$ and $\frac{b}{h}$ is 1, this means that $\frac{a}{h}$ is a factor of $t - y$, so we have $k \in \mathbb{Z}$ where

$$t - y = \frac{ka}{h}$$

that is

$$y = t - \frac{ka}{h}.$$

But we know that $a(x - s) + b(y - t) = 0$. Rearranging, this means that

$$a(x - s) = \frac{kab}{h}$$

and so

$$x = s + \frac{kb}{h}.$$

This completes the proof. □

The set

$$S = \{(x, y) \mid ax + by = c,\ x \in \mathbb{Z},\ y \in \mathbb{Z}\}$$

is called the *general solution* to the equation $ax + by = c$. The work in this section gives us a procedure for finding the general solution to the equation $ax + by = c$:

- Use Euclid's algorithm to find the highest common factor, h.
- Is h a factor of c ? If not, the equation has no integer solutions. The general solution is the empty set, $\emptyset$.
- If h is a factor of c, run Euclid's algorithm backwards to find integers $s, t \in \mathbb{Z}$ such that $as + bt = h$.
- The equation $ax+by = c$ then has a particular solution $x = cs/h$, $y = ct/h$, where s and t are as above.
- The general solution to the equation $ax + by = c$ is given by

$$x = \frac{cs}{h} + \frac{kb}{h} \qquad y = \frac{ct}{h} - \frac{ka}{h} \qquad k \in \mathbb{Z}.$$

This seems like a lot to remember, but the keys are the use of Euclid's algorithm to find a particular solution, and proceeding to a particular solution. Most of the rest is common sense. Let us try with an example.

Example 1.20 Find the general solution to the equation

$$1071x + 462y = 42.$$

where $x, y \in \mathbb{Z}$.

Solution:

- We use Euclid's algorithm to find the highest common factor of 1071 and 462:

$$\begin{aligned} 1071 &= 2 \times 462 + 147 \\ 462 &= 3 \times 147 + 21 \\ 147 &= 7 \times 21 \end{aligned}$$

  So $hcf(1071, 462) = 21$.

- 21 is a factor of the right-hand side, 42, so we do have some solutions.
- Consider the equation

$$1071x + 462y = 21.$$

  Running Euclid's algorithm backwards, we see

$$\begin{aligned} 21 &= 462 - 3 \times 147 \\ &= 462 - 3 \times (1071 - 2 \times 462) \\ &= -3 \times 1071 + 7 \times 462 \end{aligned}$$

  So we have a particular solution $x = -3$, $y = 7$ to $1071x + 462y = 21$.

- Doubling, we have a particular solution $x = -6$, $y = 14$ to $1071x+462y = 42$. Hence we have the general solution

$$x = -6 + \frac{462k}{21} = -3 + 22k \qquad y = 14 - \frac{1071k}{21} = 14 - 51k$$

  where $k \in \mathbb{Z}$.

2 The Integers Modulo n

2.1 Modular Arithmetic

Definition 2.1 Let $m \geq 2$ be an integer. Then we say $a, b \in \mathbb{Z}$ are equivalent *modulo m* if a and b leave the same remainder when divided by m. In this case, we write

$$a \equiv b \mod m.$$

By the division algorithm, $a \equiv b \mod m$ if and only if there is an integer $k \in \mathbb{Z}$ such that

$$a = mk + b,$$

ie: m is a factor of $a - b$.

We can represent all numbers modulo m by numbers in the set

$$\{0, 1, 2, \ldots, m-1\},$$

ie: any integer is congruent to one (and only one) of these numbers module m.

Example 2.2

- 13 and 1 leave the same remainder when divided by 6. So $13 \equiv 1 \mod 6$.
- $5 \times 8 + 3 = 43$. So $43 \equiv 3 \mod 5$.
- $0 \equiv 20 \mod 10$.
- $13 \equiv 18 \mod 5$.
- $-2 \equiv 5 \mod 7$.
- All odd numbers are congruent to each-other modulo 2.

Addition and multiplication can be performed modulo m.

Proposition 2.3 *Let $a_1 \equiv a_2 \mod m$, and $b_1 \equiv b_2 \mod m$. Then $a_1 + b_1 \equiv a_2 + b_2 \mod m$, and $a_1 a_2 \equiv b_1 b_2 \mod m$.*

Proof: We know that m is a factor of $a_1 - a_2$, and of $b_1 - b_2$. Hence, m is a factor of

$$(a_1 - a_2) + (b_1 - b_2) = (a_1 + b_1) - (a_2 - b_2).$$

But this means that $a_1 + b_1 \equiv a_2 + b_2 \mod m$.

Now, observe

$$a_1 b_1 - a_2 b_2 = a_1 b_1 - a_1 b_2 + a_1 b_2 - a_2 b_2 = a_1(b_1 - b_2) + +b_2(a_1 - a_2).$$

Since m is a factor of $b_1 - b_2$, and $a_1 - a_2$, we have that m is a factor of $a_1 b_1 - a_2 b_2$. This means that $a_1 a_2 \equiv b_1 b_2 \mod m$. □

Usually when performing arithmetic modulo m, we arrange things so that we get an answer in the range $\{0, 1, \ldots, m-1\}$.

Example 2.4

Working modulo 12:

$$11 + 3 \equiv 14 \equiv 2 \mod 2.$$

This should be familiar from telling the time. The time 3 hours after 11 o'clock is of course 2 o'clock. Yes- telling the time is arithmetic modulo 12.

Working modulo 7:

$$6 \times 5 \equiv 30 \equiv 2 \mod 7.$$

Division, however, is not something we can assume. Nor is cancelling factors to solve equations. This is something we explore next.

2.2 Solving Equations

Suppose we want to divide b by a modulo m. This means finding x such that

$$ax \equiv b \mod m.$$

The modular equation means that

$$ax + my = b \qquad y \in \mathbb{Z}.$$

But this is a linear Diophantine equation, with x and y as unknowns (with x the one we're really interested in). We know how to solve such equations from our work in the previous chapter.

Example 2.5 Solve

$$35x \equiv 21 \mod 91.$$

Solution:

- Write this equation

$$35x + 91y = 21$$

  where $x, y \in \mathbb{Z}$.

- Perform Euclid's algorithm to find $hcf(91, 35)$:

$$\begin{array}{rcl} 91 & = & 2 \times 35 + 21 \\ 35 & = & 21 + 14 \\ 21 & = & 14 + 7 \\ 14 & = & 2 \times 7 \end{array}$$

  So $hcf(91, 35) = 7$.

- 7 is a factor of the right-hand side, 21, so we have some solutions.

- Consider the equation

  $$35x + 91y = 7$$

  Running Euclid's algorithm backwards, we see that

  $$\begin{aligned} 7 &= 21 - 14 \\ &= 21 - (35 - 21) \\ &= 2 \times 21 - 35 \\ &= 2 \times (91 - 2 \times 35) - 35 \\ &= 2 \times 91 - 5 \times 35 \end{aligned}$$

  So the equation

  $$35x + 91y = 7$$

  has a particular solution

  $$x = -5, y = 2.$$

  Hence the equation $35x + 91y = 21$ has the particular solution $x = -15$, $y = 6$, and so general solution

  $$x = -15 + \frac{91k}{7} = -15 + 13k \qquad y = 6 - \frac{35k}{7} = 6 - 5k$$

  where $k \in \mathbb{Z}$.

- Going back to the modular equation, our original equation

  $$35x \equiv 21 \mod 91$$

  has general solution

  $$x = -15 + 13k \qquad k \in \mathbb{Z}.$$

  In other words, $x \equiv -15 \mod 13$, ie: $x \equiv 2 \mod 13$.

  Looking for solutions in the range

  $$\{0, 1, 2, \ldots, 90\}$$

  we see that we have possible solutions

  $$x \equiv 2, 15, 28 \mod 39.$$

- Finally, we should use a calculator to check our answers. We want, in the fact case, to check that 2 × 35 differs from 21 by a multiple of 9, ie: that $((2 \times 35) - 21)/9$ is an integer.

  The check in the other cases are similar. Doing this here, fortunately everything works, so we haven't made a mistake- phew!

Notice that we got *three* answers when solving the equation $35x \equiv 21 \mod 91$. It is possible that we have one, more than one or no answer when trying to solve this problem.

Example 2.6 Solve

$$4x \equiv 1 \mod 6.$$

Solution:

We might see immediately that no even number can possibly be equivalent to 1 modulo 4, so that this equation has no solution. If we see that, we've saved time.

But if we didn't see that immediately, we could follow the formal procedure.

- Write our equation
$$4x + 6y = 1 \qquad x, y \in \mathbb{Z}$$
- By Euclid's algorithm (or just spotting it), $hcf(4, 6) = 2$.
- 2 is not a factor of the right-hand side, 1. So the equation has no solutions.

The following result gives a condition under which we have a *unique* solution to our equation.

Theorem 2.7 *Let p be a prime number. Let $a \in \{1, \ldots, p-1\}$, $b \in \mathbb{Z}$. Then the equation*

$$ax \equiv b \mod p$$

has a unique solution modulo p.

Proof: Following our usual procedure, first note that we want to solve the equation

$$ax + py = b \qquad x, y \in \mathbb{Z}.$$

Since p is prime, and $a \in \{1, \ldots, p-1\}$, we have $1 = hcf(a, p)$. The number 1 is certainly a factor of b, so the equation has solutions.

Running Euclid's algorithm backwards, we can pick out $s, t \in \mathbb{Z}$ such that $as + pt = 1$. The general solution to the equation $ax + py = b$ is then

$$x = bs + kp \qquad y = bt - ka \qquad k \in \mathbb{Z}.$$

We see immediately that $x \equiv bs \mod p$. This is (modulo p) the only solution to our original equation, as required. □

2.3 The Chinese Remainder Theorem

The result we are here calling the *Chinese remainder theorem* was originally stated in the 3rd century AD by the Chinese mathematician Sun Zi, and generalised in the 13th century by Qin Jiushao.

In modern language, the result is as follows.

Theorem 2.8 *Let $m, n \geq 2$, with $hcf(m,n) = 1$. Let $a, b \in \mathbb{Z}$. Then we have $x \in \mathbb{Z}$ such that the equations*

$$x \equiv a \mod m \qquad x \equiv \mod n$$

both hold. Further, this x is unique modulo mn. □

In other words, the two equations

$$x \equiv a \mod m \qquad x \equiv \mod n$$

have a solution, and this solution is unique modulo mn.

Definition 2.9 We call two integers a and b *coprime* if $hcf(a,b) = 1$, ie: they have no common factors other then 1.

To see how to compute this solution, we (as usual) convert to a linear Diophantine equation. To see how to do this, consider the following example.

Example 2.10 Find all solutions $x \in \mathbb{Z}$ to the simultaneous equations

$$x \equiv 1 \mod 3 \qquad x \equiv 2 \mod 5.$$

Solution: We solve

$$x = 1 + 3k \qquad x = 2 + 5l$$

where $x, k, l \in \mathbb{Z}$.

Combining these equations

$$1 + 3k = 2 + 5l$$

so

$$3k - 5l = 1$$

where $k, l \in \mathbb{Z}$. Observe that $k = 2$ and $l = 1$ is a particular solution- if we don't spot this, we could find it using our usual Euclid's algorithm technieque.

Hence, as usual, the general solution is

$$k = 2 - 5n \qquad l = 1 - 3n \qquad n \in \mathbb{Z}.$$

So

$$x = 1 + 3(2 - 5n) = 7 - 15n \qquad n \in \mathbb{N}$$

We conclude that

$$x \equiv 7 \mod 15.$$

2.4 Fermat's Little Theorem

The result known as *Fermat's little theorem* involves prime numbers and congruences. It was discovered in the 17th century by the lawyer and amateur mathematician Pierre de Fermat. We call it Fermat's little theorem to distinguish it from the more famous (and *much* harder to prove) result that is Fermat's Last Theorem.

Theorem 2.11 *Let p be a prime number. Let $a \in \mathbb{Z}$. Then*

$$a^p \equiv a \mod p.$$

Proof: In this proof, we assume $a \geq 1$. We can then deduce the other cases.

We work by induction on a. First of all, note that the result is clear when $a = 1$.

Suppose that the result is true for a, so that

$$a^p \equiv a \mod p.$$

We need the corresponding formula for $a + 1$. Observe, by the binomial theorem

$$(a+1)^p = a^p + pa^{p-1} + \frac{1}{2}p(p-1)a^{p-2} + \cdots + pa + 1.$$

Now, any of the above terms with a factor of p are congruent to 0 modulo p. Hence

$$(a+1)^p \equiv a^p + 1 \equiv a + 1 \mod p$$

and we are done. □

Corollary 2.12 *Let p be a prime number. Let $a \in \mathbb{Z}$ be such that p is not a factor of a. Then*

$$a^{p-1} \equiv 1 \mod p.$$

Proof: Let $x = a^{p-1}$. By Fermat's little theorem, we know that

$$ax \equiv a^p \equiv a \mod p.$$

By theorem 2.7, we know that there is a unique value of x modulo p which solves this equation. Certainly, $x = 1$ is a solution. So by uniqueness, $x \equiv 1 \mod p$. By we know that $x = a^{p-1}$, so

$$a^{p-1} \equiv 1 \mod p$$

and we are done. □

We also refer to the above corollary as Fermat's Little Theorem.

Example 2.13 Find $3^{72} \mod 55$.
Solution:

- First we factor 55 into a product of prime numbers, each of which we we can work on with Fermat's little theorem. This is a case of straightforward checking of factors; here, we have $55 = 11 \times 5$, and 11 and 5 are prime numbers.

- We now work out $3^{72} \mod 11$. To do this, we apply Fermat's little theorem with $p = 11$. Doing this, we know

$$3^{10} \equiv 1 \mod 11.$$

  Now

$$72 = 7 \times 10 + 2$$

  so

$$3^{72} = (3^{10})^7 \times 3^2.$$

  By the above, we have

$$3^{72} \equiv 3^2 \equiv 9 \mod 11.$$

- Now we work out $3^{72} \mod 5$. Apply Fermat's little theorem with $p = 5$. Then

$$3^4 \equiv 1 \mod 5.$$

  Now

$$72 = 18 \times 4$$

  so

$$3^{72} \equiv (3^4)^{18} \equiv 1 \mod 5$$

- Let $x = 3^{72}$. We know

$$x \equiv 9 \mod 11 \qquad x \equiv 1 \mod 5.$$

  By the Chinese remainder theorem, we know there is a unique solution of x modulo 55. If we find it, that is our answer.

  As usual, we can write these equations

$$x = 9 + 11k \qquad x = 1 + 5l$$

  $k, l \in \mathbb{Z}$. Hence $9 + 11k = 1 + 5l$, and

$$11k - 5l = 1 - 9 = -8.$$

  By inspection, we spot a solution $k = 2$, $l = 6$. If we can't spot a solution in this way (because the numbers are too large), we could have used Euclid's algorithm.

  Plugging this in, we see that our solution is

$$x \equiv 9 + (11 \times 2) \equiv 31 \mod 55.$$

2.5 RSA Cryptography

RSA is an algorithm, named after its inventors, Rivest, Shamir and Adleman. It is used to send secret messages that are encoded using a widely known *public key*, but can only be decoded using a secret *private key*.

There are three steps to RSA cyptography- *key generation*, *encryption*, and *decryption*.

Key Generation

- Choose two prime numbers p and q and compute $n = pq$.
- Pick $e \in \mathbb{Z}$ such that $1 < e < (p-1)(q-1)$, and e and $(p-1)(q-1)$ are coprime.
- Find $d > 0$ such that

$$de \equiv 1 \mod (p-1)(q-1).$$

The pair (n, e) is the *public key* for encryption. The number d is the *private key*, and is used, along with n for *decryption*.

The security of the RSA algorithm rests on the fact that knowing just n and e, it is very hard to find d; there is no efficient computer algorithm to do this when the prime numbers p and q are large.

Part of finding d from n and e is finding the prime numbers p and q such that $n = pq$. If n is a very large number, this is a hard problem to solve with currently known algorithms, potentially taking hours or even weeks of computer time.

Encryption

To send a secret message:

- Represent the message (or perhaps an individual letter of the message) by an integer, M, such that $0 \leq M < n$.
- Find $c < n$ such that $c \equiv M^e \mod n$.
- Transmit the number c.

Decryption

Decoding an encrypted message c means finding M if we know c.

If we know the private key, d, we can do this using the following result.

Theorem 2.14 *$c^d \equiv M \mod n$.*

Proof: Recall $c \equiv M^e \mod n$. Observe

$$c^d \equiv M^{de} \mod n.$$

Recall that $de \equiv 1 \mod (p-1)(q-1)$, so

$$de = 1 + k(p-1)(q-1)$$

for some $k \in \mathbb{Z}$, meaning

$$M^{de} = M \times (M^{(p-1)(q-1)})^k.$$

By Fermat's little theorem

$$M^{p-1} \equiv 1 \mod p \qquad M^{q-1} \equiv 1 \mod q$$

so, by the above

$$M^{de} \equiv M \mod p \qquad M^{de} \equiv M \mod q$$

By the Chinese remainder theorem, the pair of equations

$$x \equiv M \mod p \qquad x \equiv M \mod q$$

has a unique solution modulo $n = pq$. We conclude

$$M^{de} \equiv M \mod n$$

so $c^d \equiv M \mod n$ and we are done. □

Example 2.15
Solution:

- Choose prime numbers $p = 61$ and $q = 53$. We compute $n = pq = 61 \times 53 = 3233$.
- We have $(p-1)(q-1) = (61-1)(53-1) = 3120$. Choose $1 < e < 3120$ such that e and 3120 are coprime. Let us pick $e = 17$.
- Find $d > 0$ such that $de \equiv 1 \mod 3120$.

  Using our usual methods to solve such equations, we find $d = 2753$.

So our public key is $(n, e) = (3233, 17)$. Our private key is $d = 2753$.
So we encrypt using the function $E\colon \{0, \dots, n-1\} \to \{0, \dots, n-1\}$ given by

$$c = E(M) \equiv M^{17} \mod 3223.$$

We decrypt using the function $D\colon \{0, \dots, n-1\} \to \{0, \dots, n-1\}$ given by

$$M = D(c) \equiv c^{2753} \mod 3223.$$

For instance, in order to encrypt $M = 65$, we calculate

$$c \equiv 65^{17} \mod 3223$$

which is, using the techniques of the previous section, 2790 modulo 3223.

To decrypt $c = 2790$, we calculate

$$M \equiv 2790^{2753} \mod 3223$$

which is, using the techniques of the previous section, 65.

The code is secure, as knowing just $n = 3223$ and $e = 17$, it is very hard to find $d = 2753$ which we need to decrypt.

3 Matrices, Vectors and Systems of Linear Equations

3.1 Elementary Row Operations

In contrast to the previous two sections, we are now going to work over the set of real numbers, $\mathbb{R}$, rather than the set of integers. The real numbers includes all integers, fractions, and irrational numbers such as $\sqrt{2}$ and π. It does *not* include imaginary numbers such as $i = \sqrt{-1}$.

Consider a system of m linear equations in n variables,

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

where $a_{ij},\ b_j \in \mathbb{R}$ are fixed known real numbers, and the $x_i \in \mathbb{R}$ are the unknown variables we want to find.

To *solve* the system means finding all possible values of the variables x_i. A system of linear equations could have one solution, more than one solution, or no solutions; there are examples of each type.

Example 3.1 Solve the simultaneous equations

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

Solution:

Working formally, we:

- Multiply the first equation by $\frac{1}{2}$:

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

- Add three times the first equation to the second equation:

$$\begin{aligned} x - 2y &= 2 \\ -2y &= 8 \end{aligned}$$

- Multiply the second equation by $-\frac{1}{2}$:

$$\begin{aligned} x - 2y &= 2 \\ y &= -4 \end{aligned}$$

- Add two times the second equation to the first equation:

$$\begin{array}{rcl} x & = & -6 \\ y & = & -4 \end{array}$$

In the above example, everything is straightforward. For examples with more equations and variables, things get more complex- it is useful to formalise the process of solving these equations and have an algorithm.

Definition 3.2 An *$m \times n$ matrix* of real numbers is an array

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

An *augmented matrix* is an array of the form

$$(A|b) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right)$$

The above augmented matrix is called the matrix of the system of linear equations

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

Example 3.3 The system of linear equations

$$\begin{array}{rcl} 2x - 4y & = & 4 \\ -3x + 4y & = & 2 \end{array}$$

is represented by the augmented matrix

$$\left(\begin{array}{cc|c} 2 & -4 & 4 \\ -3 & 4 & 2 \end{array}\right)$$

Now, notice that when solving a system of linear equations, there are some operations we can perform to try to simplify it and find a solution. Specifically, we can:

- Swapping two equations

- Multiply one equation by a non-zero real number.
- Add (or subtract) a multiple of one equation from another (different) equation.

None of these operations changes the solution of a system of linear equations, and the idea is to perform them until our system is in such a state that we can just 'read off' a solution; we saw this in our above example.

Definition 3.4 An *elementary row operation* (ERO) on a matrix is one of the following three things:

- Swapping two rows of the matrix.
- Multiplication of one row of the matrix by a non-zero number.
- The addition or subtraction of a multiple of one row of the matrix to another (different) row.

We write $C \sim D$ if we obtain the matrix D by performing a finite number of EROs on C.

Remark 3.5 The elementary row operations are all reversible. So if $C \sim D$, then $D \sim C$. In fact, $\sim$ is an equivalence relation.

The following result is clear from our remarks, and is the key to solving systems of linear equations using EROs.

Theorem 3.6 *Consider a system of linear equations with augmented matrix* $(A|b)$*. Suppose* $(A|b) \sim (A'|b')$*. Then the system of linear equations with augmented matrix* $(A'|b')'$ *has the same solutions.* □

In other words, we write down our system in matrix form, and perform elementary row operations until we can read off a solution.

Example 3.7 Solve the system of linear equations

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

Solution:

This system has matrix:

$$\left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 2 & 1 & -1 & 2 \\ 1 & -2 & 1 & 1 \end{array}\right)$$

Now we perform EROs:

- Subtract twice row 1 from row 2, and subtract row 1 from row 3:

$$\left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 2 & 1 & -1 & 2 \\ 1 & -2 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 0 & 3 & -5 & -1 \\ 0 & -1 & -1 & 0 \end{array}\right)$$

- Swap row 2 and row 3, then multiply row 2 by -1:

$$\sim \left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 3 & -5 & -1 \end{array}\right)$$

- Subtract 3 times row 2 from row 3:

$$\sim \left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & -8 & -1 \end{array}\right)$$

- Multiply row 3 by $-1/8$:

$$\sim \left(\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & \frac{1}{8} \end{array}\right)$$

- Subtract row 3 from row 2, and twice row 3 from row 1:

$$\sim \left(\begin{array}{ccc|c} 1 & -1 & 0 & \frac{3}{4} \\ 0 & 1 & 0 & -\frac{1}{8} \\ 0 & 0 & 1 & \frac{1}{8} \end{array}\right)$$

- Add row 2 to row 1:

$$\sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & \frac{5}{8} \\ 0 & 1 & 0 & -\frac{1}{8} \\ 0 & 0 & 1 & \frac{1}{8} \end{array}\right)$$

We can now convert back into a system of linear equations and read off the solutions:

$$\begin{aligned} x &= \tfrac{5}{8} \\ y &= -\tfrac{1}{8} \\ z &= \tfrac{1}{8} \end{aligned}$$

3.2 General Solutions

In the examples we have seen so far, our systems of linear equations have had a unique solution. It is also possible for a system of linear equations to have many solutions, or no solution.

Given a system of linear equations,

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

the set of all possible ordered sets, $(x_1, \dots, x_n)$, for which the above equations are all true is called the *general solution.*

Example 3.8 Consider the system of linear equations

$$\begin{aligned}
v + 2w - 3x + 3y + 2z &= 0 \\
-v - w + 3x + z &= 3 \\
v + 2w + y &= 1 \\
-v - w + 4y + 5z &= 4 \\
v + 2w + x + 7y + z &= 8
\end{aligned}$$

Solution:

The system has augmented matrix

$$(A|b) = \left(\begin{array}{ccccc|c}
1 & 2 & -3 & 3 & 2 & 0 \\
-1 & -1 & 3 & 0 & 1 & 3 \\
1 & 2 & 0 & 1 & 0 & 1 \\
-1 & -1 & 0 & 4 & 5 & 4 \\
1 & 2 & 1 & 7 & 6 & 8
\end{array}\right).$$

$$\begin{aligned}
(A|b) &\sim \left(\begin{array}{ccccc|c}
1 & 2 & -3 & 3 & 2 & 0 \\
0 & 1 & 0 & 3 & 3 & 3 \\
0 & 0 & 3 & -2 & -2 & 1 \\
0 & 1 & -3 & 7 & 7 & 4 \\
0 & 0 & 4 & 4 & 4 & 8
\end{array}\right) \sim \left(\begin{array}{ccccc|c}
1 & 2 & -3 & 3 & 2 & 0 \\
0 & 1 & 0 & 3 & 3 & 3 \\
0 & 0 & 3 & -2 & -2 & 1 \\
0 & 0 & -3 & 4 & 4 & 1 \\
0 & 0 & 1 & 1 & 1 & 2
\end{array}\right) \\
&\sim \left(\begin{array}{ccccc|c}
1 & 2 & -3 & 3 & 2 & 0 \\
0 & 1 & 0 & 3 & 3 & 3 \\
0 & 0 & 1 & 1 & 1 & 2 \\
0 & 0 & -3 & 4 & 4 & 1 \\
0 & 0 & 3 & -2 & -2 & 1
\end{array}\right) \sim \left(\begin{array}{ccccc|c}
1 & 2 & -3 & 3 & 2 & 0 \\
0 & 1 & 0 & 3 & 3 & 3 \\
0 & 0 & 1 & 1 & 1 & 2 \\
0 & 0 & 0 & 7 & 7 & 7 \\
0 & 0 & 0 & -5 & -5 & -5
\end{array}\right) \\
&\sim \left(\begin{array}{ccccc|c}
1 & 2 & -3 & 3 & 2 & 0 \\
0 & 1 & 0 & 3 & 3 & 3 \\
0 & 0 & 1 & 1 & 1 & 2 \\
0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0
\end{array}\right) \sim \left(\begin{array}{ccccc|c}
1 & 0 & -3 & -3 & -4 & -6 \\
0 & 1 & 0 & 3 & 3 & 3 \\
0 & 0 & 1 & 1 & 1 & 2 \\
0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0
\end{array}\right) \\
&\sim \left(\begin{array}{ccccc|c}
1 & 0 & 0 & 0 & -1 & 0 \\
0 & 1 & 0 & 3 & 3 & 3 \\
0 & 0 & 1 & 1 & 1 & 2 \\
0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0
\end{array}\right) \sim \left(\begin{array}{ccccc|c}
1 & 0 & 0 & 0 & -1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0
\end{array}\right),
\end{aligned}$$

Converting back to equations:

$$\begin{array}{rcl} v - z &=& 0 \\ w &=& 0 \\ x &=& 1 \\ y + z &=& 1 \\ 0 &=& 0 \end{array}$$

The final equation '$0 = 0$' doesn't tell us anything, so we ignore it. Now, to work out the general solution, we work backwards through the variables, starting with z and the bottom equation. If we don't know what a variable is, we set it to be an arbitrary real number. We denote such arbitrary numbers by letters in the Greek alphabet, $\alpha, \beta, \gamma, \ldots$.

So, above, we can't determine z. Set $z = \alpha$ where $\alpha \in \mathbb{R}$. Substitute this into the other equations. Then we get $y = 1 - \alpha$, $x = 1$, $w = 0$, and $v = z = \alpha$. So we have general solution

$$\begin{array}{rcl} v &=& \alpha \\ w &=& 0 \\ x &=& 1 \\ y &=& 1 - \alpha \\ z &=& \alpha \end{array} \qquad \alpha \in \mathbb{R}.$$

Example 3.9 Solve the system of linear equations

$$\begin{array}{rcrcrcr} 2x &+& y &+& z &=& 1 \\ 4x &+& 2y &+& 3z &=& -1 \\ 6x &+& 3y &-& z &=& 11 \end{array}.$$

Solution:

The augmented matrix of this system is

$$\left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 4 & 2 & 3 & -1 \\ 6 & 3 & -1 & 11 \end{array}\right) \sim \left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & -4 & 8 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & -4 \end{array}\right) \sim \left(\begin{array}{ccc|c} 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

Going back to equations, the last matrix gives us:

$$\begin{array}{rcl} x + y &=& 0 \\ z &=& 0 \\ 0 &=& 1 \end{array}$$

The final equation '$0 = 1$' can *never* be satisfied- it is nonsense. This means that our system of linear equations does not have any solutions. Getting the equation '$0 = 1$' or something similar is what always happens with this method when a system of equations can't be solved.

In other words, there is *no* x, y and z satisfying the original three equations. Our general solution is the empty set.

3.3 Reduced Echelon Form

Our aim in this section is to express our method of solving linear equations as a formal algorithm.

Definition 3.10 A matrix is said to be in *reduced echelon form* if:

- Any rows consisting entirely of zeros are below the other rows.
- The leading entry in any non-zero row is a 1, and the leading 1 is in a column to the right of the leading 1 of the row above it.
- Each leading 1 is the only non-zero entry in its column.

Example 3.11 The following matrices are in reduced echelon form:

$$\begin{pmatrix} 1 & 0 & 0 & 43 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 4 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 2 & 0 & 0 & 0 & -1 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 & 0 & 0 & 2 & 3 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 0 & -2 & -3 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We can easily read off the solution of a system of linear equations whose matrix is in reduced echelon form.

Example 3.12 Consider a system of linear equations in variables x, y, z with augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 43 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 4 \end{array}\right)$$

Then we have solution

$$x = 43, \ y = -2, \ z = 4.$$

Example 3.13 Consider a system of linear equations in variables $x_1, x_2, \ldots, x_9$ with augmented matrix

$$\left(\begin{array}{ccccccccc|c} 0 & 1 & 2 & 0 & 0 & 0 & -1 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 & 0 & 0 & 2 & 3 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 0 & -2 & -3 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Then we have corresponding equations

$$\begin{array}{rcl} x_2 + 2x_3 - x_7 + x_8 & = & 5 \\ x_4 + 2x_7 + 3x_8 & = & 4 \\ x_5 - 2x_7 - 3x_8 & = & 3 \\ x_6 - x_7 - x_8 & = & 2 \\ x_9 & = & 1 \end{array}$$

Then we have general solution (reading from the last variable upwards):

$$\begin{array}{rcl} x_9 &=& 1 \\ x_8 &=& \alpha \\ x_7 &=& \beta \\ x_6 &=& 2+\alpha+\beta \\ x_5 &=& 3+3\alpha+2\beta \\ x_4 &=& 4-3\alpha-2\beta \\ x_3 &=& \gamma \\ x_2 &=& 5-\alpha+\beta-2\gamma \\ x_1 &=& \delta \end{array}$$

where $\alpha, \beta, \gamma, \delta \in \mathbb{R}$.

Note that the variables corresponding to the leading 1's are determined by the other variables. Those not corresponding to the leading 1's, we set to be arbitrary real numbers in the general solution.

So, if we can transform a matrix into reduced echelon form, we can just read off the solution of the corresponding system of linear equations.

The proof of our next result is just as important as the statement.

Theorem 3.14 *Any matrix can be put into reduced echelon form by a finite sequence of elementary row operations.*

Proof: Let

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

be an $m \times n$ *matrix* of real numbers. Note that a_{ij} is the entry of A in row i and column j . We present an algorithm to put A into reduced echelon form by elementary row operations.

Start with $i = 1$, $j = 1$.

- If $a_{ij} = 0$, swap row i with some other row below it to ensure that $a_{ij} \neq 0$. If all entries in column j are zero, increase j by 1.
- Multiply row i by $1/a_{ij}$ to make the leading entry 1.
- Eliminate all other entries in column j by subtracting multiples of row i from the other rows.
- Increase i and j by 1, and return to the first step.

The algorithm stops when we are done with the last row or column of the matrix. □

3.4 Vectors and Linear Transformations

Definition 3.15 An *n-dimensional vector* is an $n \times 1$ matrix of real numbers. We write $\mathbb{R}^n$ to denote the set of n-dimensional vectors.

Example 3.16

$$\begin{pmatrix} 1 \\ -\frac{1}{2} \\ 2 \\ 0 \end{pmatrix} \in \mathbb{R}^4$$

is an example of a 4-dimensional vector.

We think of a vector as representing a point in n-dimensional space. We say that two vectors are equal precisely when their corresponding entries are equal.

Definition 3.17 Let

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \ v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n.$$

Then we define the *sum* $u + v$ to be the vector

$$u + v = \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix}.$$

Given a real number $\alpha \in \mathbb{R}$, we define the *scalar multiple* αu to be the vector

$$\alpha u = \begin{pmatrix} \alpha u_1 \\ \alpha u_2 \\ \vdots \\ \alpha u_n \end{pmatrix}.$$

We also define

$$u - v = u + (-1)v = \begin{pmatrix} u_1 - v_1 \\ u_2 - v_2 \\ \vdots \\ u_n - v_n \end{pmatrix}.$$

Example 3.18 Consider the vectors

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \in \mathbb{R}^2.$$

Then

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 3 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 5 \end{pmatrix}.$$

Recall that for sets A and B, a function $f\colon A \to B$ is a rule which sends an element $a \in A$ to an element $f(a) \in B$.

Definition 3.19 A *linear transformation* is a function $T\colon \mathbb{R}^m \to \mathbb{R}^n$ such that

$$T(\alpha u + \beta v) = \alpha T(u) + \beta T(v)$$

for all $\alpha, \beta \in \mathbb{R}$ and $u, v \in \mathbb{R}^m$.

Example 3.20 We can define a linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ by writing

$$T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ x + 3y \\ 2x - y \end{pmatrix}.$$

Definition 3.21 Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \qquad v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots v_n \end{pmatrix}$$

be an $m \times n$ matrix of real numbers and an n-dimensional vector respectively. Then we define the *product* $Av \in R^m$ by:

$$Av = \begin{pmatrix} a_{11}v_1 + a_{12}v_2 + \dots + 1_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \dots + 1_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \dots + 1_{mn}v_n \end{pmatrix}.$$

Example 3.22 Let

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

Then

$$Av = \begin{pmatrix} 1 \times 1 + 2 \times 2 + 0 \times (-1) \\ -1 \times 11 \times 2 + 3 \times (-1) \end{pmatrix} = \begin{pmatrix} 5 \\ -2 \end{pmatrix}.$$

Using the above, a system of linear equations

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

can be written as a single equation involving vectors. To be precise, let

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \qquad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots b_m \end{pmatrix}.$$

Let us write the variables $x_1, x_2, \dots, x_n$ as a vector

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots x_n \end{pmatrix}.$$

Then the system of linear equations is satisfied if and only if the vector equation

$$AX = b$$

holds. We will sometimes use this compact notation.

The following result is obtained by carefully checking the relevant definitions. The proof is not very interesting, so we leave it out.

Theorem 3.23 *Let A be an $m \times n$ matrix of real numbers. Then the function which sends $v \in \mathbb{R}^n$ to $Av \in \mathbb{R}^m$ is a linear transformation.*

Further, if $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then there is a matrix B such that $T(v) = Bv$ for all $v \in \mathbb{R}^n$. □

We call the above matrix B the matrix associated to the linear transformation.

Example 3.24 Consider the linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ by writing

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ x+3y \\ 2x-y \end{pmatrix}.$$

Let

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 3 \\ 2 & -1 \end{pmatrix} \qquad [v = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Then

$$Av = \begin{pmatrix} x \\ x+3y \\ 2x-y \end{pmatrix} = T(v).$$

3.5 Linear Independence

The concept of linear independence is central in linear algebra, and will prove useful to us later on.

Definition 3.25 We call a set of vectors $\{v_1, \ldots, v_r\}$ in $\mathbb{R}^n$ *linearly independent* if the *only* solution to the equation

$$\alpha_1 v_1 + \cdots + \alpha_r v_r \qquad \alpha_i \in \mathbb{R}$$

is given by $\alpha_1 = 0, \alpha_2 = 0, \ldots, \alpha_r = 0$.

Example 3.26 Consider the vectors

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad v_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

in $\mathbb{R}^3$. Is the set of vectors $\{v_1, v_2, v_3\}$ linearly independent ?

Solution:

Consider the equation

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = 0.$$

Looking at entries of the vector, this is a system of linear equations

$$\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 &= 0 \\ 2\alpha_1 + \alpha_3 &= 0 \\ \alpha_1 + \alpha_2 &= 0 \end{aligned}$$

This system has augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

So the only solution is

$$\alpha_1 = 0 \quad \alpha_2 = 0 \quad \alpha_3 = 0.$$

We conclude that the vectors v_1, v_2, v_3 are linearly independent.

Example 3.27 Consider the vectors

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \quad v_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

in $\mathbb{R}^3$. Is the set of vectors $\{v_1, v_2, v_3\}$ linearly independent ?

Solution: Consider the equation

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = 0.$$

Looking at entries of the vector, this is a system of linear equations

$$\begin{aligned} \alpha_1 + \alpha_2 &= 0 \\ 2\alpha_1 + \alpha_3 &= 0 \\ 2\alpha_1 + 2\alpha_2 &= 0 \end{aligned}$$

This system has augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 2 & 2 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

So we have the general solution $\alpha_3 = \alpha$, $\alpha_2 = \frac{1}{2}\alpha$, and $\alpha_1 = -\frac{1}{2}\alpha$, where $\alpha \in \mathbb{R}$.
In particular, we have non-zero solutions. This means the set of vectors $\{v_1, v_2, v_3\}$ is *not* linearly independent.

Sometimes, it is obvious when a set of vectors is not linearly independent.

Example 3.28 Consider the vectors

$$v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad v_2 = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$$

Then $v_2 = -2v_1$. This means that $2v_1 + v_2 = 0$. So the equation $\alpha_1 v_1 + \alpha_2 v_2 = 0$ has a non-zero solution. This means that the set $\{v_1, v_2\}$ is not linearly independent.

However, if not obvious, we can always use the above procedure to check.

4 Matrix Algebra

4.1 Matrix Multiplication

Just as for vectors, we can form the sum of two $m \times n$ matrices by adding the relevant entries.

Example 4.1

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

Similarly, given a real number $\alpha \in \mathbb{R}$, and an $m \times n$ matrix A, we define the *scalar multiple* αA to be the matrix obtained by multiplying each entry of the matrix A by α.

Example 4.2

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

However, we can also multiply matrices together of the appropriate size. The method for doing this is a generalisation of the method for multiplying a matrix and a vector.

Definition 4.3 Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mr} \end{pmatrix} \qquad B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{r1} & b_{r2} & \dots & b_{rn} \end{pmatrix}$$

be an $m \times r$ and a $r \times n$ matrix respectively. Then we define the *product* AB to be the $m \times n$ matrix, where the entry c_{ij} in row i and column j is the sum

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Note that multiplication of a matrix by a vector is a special case of this definition. Note that the number of columns of A must be the same as the number of rows of B for the product AB to make sense.

Example 4.4 Let

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

Then

$$AB = \begin{pmatrix} 1\times 1+2\times 2+3\times 1 & 1\times(-1)+2\times 1+3\times 2 \\ (-2)\times 1+1\times 2+0\times 1 & (-2)\times(-1)+1\times 1+0\times 2 \end{pmatrix} = \begin{pmatrix} 8 & 7 \\ 0 & 3 \end{pmatrix}$$

and

$$BA = \begin{pmatrix} 1\times 1+(-1)\times(-2) & 1\times 2+(-1)\times 1 & 1\times 3+(-1)\times 0 \\ 2\times 1+1\times(-2) & 2\times 2+1\times 1 & 2\times 3+1\times 0 \\ 1\times 1+2\times(-2) & 1\times 2+2\times 1 & 1\times 3+2\times 0 \end{pmatrix} = \begin{pmatrix} 3 & 1 & 3 \\ 0 & 5 & 6 \\ -3 & 4 & 3 \end{pmatrix}.$$

Remark 4.5 Matrix multiplication is NOT in general *commutative*, that is to say for matrices A and B, in general $AB \neq BA$ (even when, as above, both products make sense). This is different to the case of multiplication of real numbers or integers.

The reason to multiply matrices is that it corresponds to composition of linear transformations. To be precise, we have the following result.

Theorem 4.6 *Let $S\colon \mathbb{R}^r \to \mathbb{R}^m$ and $T\colon \mathbb{R}^n \to \mathbb{R}^r$ be linear transformations, with associated matrices A and B respectively. Then the composition $S \circ T\colon \mathbb{R}^n \to \mathbb{R}^m$ has associated matrix AB.* □

Example 4.7 As above, let

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

Define $S\colon \mathbb{R}^3 \to \mathbb{R}^2$ by $S(v) = Av$. Define $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ by $T(v) = Bv$. Then

$$S\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+2y+3z \\ -2x+y \end{pmatrix} \qquad T\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u-v \\ 2u+v \\ u+2v \end{pmatrix}.$$

Observe

$$ST\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} (u-v)+2(2u+v)+3(u+2v) \\ -2(u-v)+(2u+v) \end{pmatrix} = \begin{pmatrix} 8u+7v \\ 3v \end{pmatrix}.$$

But, by the previous example

$$AB = \begin{pmatrix} 8 & 7 \\ 0 & 3 \end{pmatrix}$$

which is the matrix associated to the linear transformation $R\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$R\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 8u+7v \\ 3v \end{pmatrix}.$$

But this is the linear transformation $S \circ T$. In other words, as claimed in the theorem, the matrix associated to $S \circ T$ is AB.

4.2 Inverses

Now, let us concentrate on *square matrices*, that is to say $n \times n$ matrices for some n. Any such matrix has the same number of rows and columns, and we can multiply two square matrices of the same size together. A linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$ has an associated $n \times n$ matrix.

Definition 4.8 The $n \times n$ matrix

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

is called the *identity matrix*.

We sometimes write just I instead of I_n. The reason for this name is the following result, which is easy to check.

Proposition 4.9 *Let $J\colon \mathbb{R}^n \to \mathbb{R}^n$ be the identity linear transformation, defined by the formula $J(v) = v$ for all $v \in \mathbb{R}^n$. Then J has associated matrix I_n.* □

Proposition 4.10 *Let A be an $n \times n$ matrix. Then $AI_n = I_nA = A$.*

Proof: Let A be associated to the linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$. Then $T \circ J = J \circ T = T$. Hence, looking at the associated matrices, as matrix multiplication corresponds to composition of linear transformations

$$AI_n = I_nA = A.$$

□

Definition 4.11 We call an $n \times n$ matrix A *invertible* if there is a matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_n$.

We call A^{-1} the *inverse* of A. If A is invertible, the inverse is unique. We now present an algorithm to find it. This method is called *Gaussian elimination.*

Theorem 4.12 *Let A be an $n \times n$ matrix. Consider the 'double matrix' $(A|I_n)$ defined by writing the identity matrix to the right of the matrix A. Suppose $(A|I_n) \sim (I_n|B)$. Then A is invertible, and $B = A^{-1}$.* □

Here, $(A|I_n) \sim (I_n|B)$ means we can get from one double matrix to the other by elementary row operations.

Example 4.13 Let

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

We have

$$\begin{aligned} \big(A|I_3\big) &= \left(\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 & 0 & 0 \end{array}\right) \\ &\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array}\right). \end{aligned}$$

It follows from the above that A is invertible, and

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

Further, we can check our arithmetic. Multiplying matrices, we see that $AA^{-1} = I_3$ and $A^{-1}A = I_3$.

4.3 Determinants

Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

be a 2×2 matrix. Then we define the *determinant*

$$\det A = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

The definition of the determinant of a 3×3 matrix uses the definition of the determinant of a 2×2 matrix.

To be more precise, let

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

be a 3×3 matrix. For $i, j = 1, 2, 3$, the *(i, j)-minor of A* is the determinant of the 2×2 matrix obtained by deletion of the row i and column j from A, and the *(i, j)-cofactor of A*, denoted A_{ij}, is the result of multiplying the (i, j)-minor of A by the sign $(-1)^{i+j}$. This sign can be seen in the matrix of alternating plus and minus symbols:

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}.$$

For example, with

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

we have

$$A_{12} = - \begin{vmatrix} \\ 3 & -1 \\ 2 & 6 \end{vmatrix} .$$

The determinant $\det A$ can be calculated by expansion along any row or down any column. Thus

$$\begin{aligned} \det A &= a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} && \text{(expansion along the first row)} \\ &= a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23} && \text{(expansion along the second row)} \\ &= a_{31}A_{31} + a_{32}A_{32} + a_{33}A_{33} && \text{(expansion along the third row)} \\ &= a_{11}A_{11} + a_{21}A_{21} + a_{31}A_{31} && \text{(expansion down the first column)} \\ &= a_{12}A_{12} + a_{22}A_{22} + a_{32}A_{32} && \text{(expansion down the second column)} \\ &= a_{13}A_{13} + a_{23}A_{23} + a_{33}A_{33} && \text{(expansion down the third column).} \end{aligned}$$

Example 4.14 Find $\det A$, where

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 3 & 1 & -1 \\ 2 & 5 & 6 \end{pmatrix} .$$

Solution: By expansion along the first row,

$$\begin{aligned} \det A &= \begin{vmatrix} 1 & 2 & 4 \\ 3 & 1 & -1 \\ 2 & 5 & 6 \end{vmatrix} \\ &= 1 \begin{vmatrix} \\ 1 & -1 \\ 5 & 6 \end{vmatrix} - 2 \begin{vmatrix} \\ 3 & -1 \\ 2 & 6 \end{vmatrix} + 4 \begin{vmatrix} \\ 3 & 1 \\ 2 & 5 \end{vmatrix} \\ &= (6 - (-5)) - 2(18 - (-2)) + 4(15 - 2) = 11 - 40 + 52 = 23. \ \square \end{aligned}$$

More generally, for an $n \times n$ matrix A, the definition of the determinant $\det A$ is recursive; it uses the definition of the determinant of an $(n-1) \times (n-1)$ matrix, in the same way as the determinant of a 3×3 matrix is defined in terms of determinants of 2×2 matrices.

This is fine; we eventually get to a calculation in terms of determinants of 2×2-matrices, and we have an explicit formula for the determinant of a 2×2 matrix.

Definition 4.15 Let $A = (a_{ij}) \in M_n(\mathbb{R})$, where $n \geq 2$. For $1 \leq i, j \leq n$, the (i, j)-*minor* of A is the determinant of the $(n-1) \times (n-1)$ matrix obtained by deletion of row i and column j in the matrix A.

The (i, j)-*cofactor* of A, written A_{ij}, is the result of multiplying the (i, j)-minor by the sign $(-1)^{i+j}$.

We define the *determinant* of A by the formula

$$\det A = a_{11}A_{11} + a_{12}A_{12} + \cdots + a_{1n}A_{1n}$$

We call the above expression 'expansion along the first row' of the matrix A.

Remark 4.16 It will occasionally be useful to consider the determinant of a 1×1 matrix. Such a determinant is defined trivially:

$$\det(a) = a$$

Example 4.17 Find $\det A$, where

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

Solution: Observe

$$\begin{aligned} \det A &= \begin{vmatrix} 2 & 4 & -3 & 1 \\ 0 & 0 & 1 & 4 \\ -1 & 1 & 0 & 1 \\ 3 & -1 & 1 & 1 \end{vmatrix} \\ &= 2 \begin{vmatrix} 0 & 1 & 4 \\ 1 & 0 & 1 \\ -1 & 1 & 1 \end{vmatrix} - 4 \begin{vmatrix} 0 & 1 & 4 \\ -1 & 0 & 1 \\ 3 & 1 & 1 \end{vmatrix} \\ &\quad + (-3) \begin{vmatrix} 0 & 0 & 4 \\ -1 & 1 & 1 \\ 3 & -1 & 1 \end{vmatrix} + -1 \begin{vmatrix} 0 & 0 & 1 \\ -1 & 1 & 0 \\ 3 & -1 & 1 \end{vmatrix} \\ &= 2\big(-(1-(-1)) + 4\big) - 4\big(-(-1-3) + 4(-1)\big) - 3\big(4(1-3)\big) - 1\big(1(1-3)\big) \\ &= 2 \times 2 + 0 + 3 \times 8 + 1 \times 2 = 4 + 24 + 2 = 30. \end{aligned}$$

In fact we can expand along any row or column to calculate determinants. To be precise, we have the following result.

Theorem 4.18 *Let $A = (a_{ij}) \in M_n(\mathbb{R})$, where $n \geq 2$. Then the determinant $\det A$ can be expanded along any row or down any column; in other words, for each $i = 1, \ldots, n$,*

$$\det A = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = \textstyle\sum_{j=1}^{n} a_{ij}A_{ij} \quad \textit{(expansion along the i-th row)}$$

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

$\det A = a_{1j}A_{1j}+a_{2j}A_{2j}+\cdots+a_{nj}A_{nj} = \sum_{i=1}^{n} a_{ij}A_{ij}$ *(expansion down the j-th column).*

□

Definition 4.19 A square matrix $B = (b_{ij})$ is said to be *lower triangular* if it takes the form

$$B = \begin{pmatrix} b_{11} & 0 & \ldots & 0 & 0 \\ b_{21} & b_{22} & \ldots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ b_{n-1\,1} & b_{n-1\,2} & \ldots & b_{n-1\,n-1} & 0 \\ b_{n1} & b_{n2} & \ldots & b_{n\,n-1} & b_{nn} \end{pmatrix}.$$

The following result is shown using mathematical induction and the definition of the determinant.

Proposition 4.20 *Let B be a lower triangular matrix. Then the determinant* $\det B$ *is the product of the diagonal entries.* □

A similar result is also true for upper triangular matrices, which are defined in the obvious way.

Corollary 4.21 *Let*

$$D = \begin{pmatrix} d_1 & 0 & \ldots & 0 & 0 \\ 0 & d_2 & \ldots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & d_{n-1} & 0 \\ 0 & 0 & \ldots & 0 & d_n \end{pmatrix}$$

Then

$$\det D = d_1 d_2 \ldots d_{n-1} d_n.$$

□

A matrix of the above type is called a *diagonal matrix*.

Example 4.22 Let I_n be the $n \times n$ identity matrix. Then $\det I_n = 1$.

We conclude with two properties of determinants that are the key to many of their uses in calculations. We omit the proofs.

Theorem 4.23 *Let A and B be $n \times n$ matrices. Then* $\det(AB) = \det(A)\det(B)$. □

Theorem 4.24 *Let A be an $n \times n$ matrix. Then A is invertible if and only if* $\det A \neq 0$. □

Example 4.25 Let

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

We have

$$\det A = -\begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} + \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1 \neq 0$$

so A is invertible.

4.4 Eigenvalues and Eigenvectors

Definition 4.26 Let A be an $n \times n$ (square) matrix. A number $\lambda \in \mathbb{R}$ is called an *eigenvalue* of the matrix A if there is a *non-zero* vector v such that $Av = \lambda v$.

Such a non-zero column vector v is called an *eigenvector* of A with eigenvalue λ, or more briefly a λ-*eigenvector* of the matrix A.

Observe $Av = \lambda v$ if and only if

$$Av - \lambda v = Av - \lambda I v = (A - \lambda I)v = 0.$$

So $\lambda \in \mathbb{R}$ is an eigenvalue of the matrix A if and only if the system of linear equations

$$(A - \lambda I)v = 0$$

has a *non-zero* solution. Each non-zero solution to the above equation is a λ-eigenvector.

Definition 4.27 Let A be an $n \times n$ matrix. The *characteristic polynomial* of A is the polynomial $\chi_A(t)$ defined by the formula

$$\chi_A(t) = \det(A - tI_n)$$

Theorem 4.28 *Let A be an $n \times n$ matrix. Then a real number $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if $\chi_A(\lambda) = 0$.*

Proof: Let $\lambda \in \mathbb{R}$. Then λ is an eigenvalue of the matrix A if and only if we have a non-zero solution to the system of linear equations $(A - \lambda I)v = 0$.

Now, if $A - \lambda I$ were invertible, the above equation would imply $v = (A - \lambda I)^{-1}0 = 0$, so we would have no non-zero solutions. Thus $(A - \lambda I)v = 0$ has a non-zero solution precisely when $A - \lambda I$ is not invertible.

By the above, this occurs when $\det(A - \lambda I) = \chi_A(\lambda) = 0$. □

So the eigenvalues of the matrix A are precisely the roots of the characteristic polynomial $\chi_A(t)$.

Once we know that $\lambda \in \mathbb{R}$ is an eigenvalue of the matrix A, we can find all λ-eigenvectors by solving the system of linear equations $(A - \lambda I_n)X = 0$; every non-trivial solution (and there will be some, since λ is an eigenvalue) will be a λ-eigenvector.

Note that any non-zero multiple of a λ-eigenvector is also a λ-eigenvector.

Example 4.29 Let

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

Find the eigenvalues and eigenvectors of A.

Solution: We begin by finding the roots of the characteristic polynomial $\chi_A(t)$ of A:

$$\chi_A(t) = \det(A - tI_2) = \begin{vmatrix} 1-t & 2 \\ -1 & 4-t \end{vmatrix} = (1-t)(4-t)+2 = 6-5t+t^2 = (3-t)(2-t).$$

Thus the eigenvalues of A are 3 and 2.

To find the eigenvectors corresponding to the eigenvalue 3, we solve the system of linear equations $(A - 3I_2)\begin{pmatrix} x \\ y \end{pmatrix} = 0$. The augmented matrix of this system

$$\big(A - 3I_2|0\big) = \left(\begin{array}{cc|c} -2 & 2 & 0 \\ -1 & 1 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} -2 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right).$$

Therefore the general solution of the system is given by $y = \alpha$, $x = \alpha$, where α can be any real number; therefore the set of eigenvectors of A corresponding to the eigenvalue 3 is

$$\left\{\alpha \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mid \alpha \in \mathbb{R},\ \alpha \neq 0\right\}.$$

(Remember that zero is not an eigenvector).

To find the eigenvectors corresponding to the eigenvalue 2, we solve the system of linear equations $(A - 2I_2)\begin{pmatrix} x \\ y \end{pmatrix} = 0$. The augmented matrix of this system is

$$\big(A - 2I_2|0\big) = \left(\begin{array}{cc|c} -1 & 2 & 0 \\ -1 & 2 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} -1 & 2 & 0 \\ 0 & 0 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -2 & 0 \\ 0 & 0 & 0 \end{array}\right).$$

Therefore the general solution of the system is given by $y = \mu$, $x = 2\beta$, where β can be any real number; therefore the set of eigenvectors of A corresponding to the eigenvalue 2 is

$$\left\{\mu \begin{pmatrix} 2 \\ 1 \end{pmatrix} \mid \beta \in \mathbb{R},\ \beta \neq 0\right\}.$$

4.5 Diagonalisation

Definition 4.30 Let $A, B \in M_n(\mathbb{R})$. We say that the matrices A and B are *similar* if there is an invertible matrix P such that $B = P^{-1}AP$.

Proposition 4.31 *Similar matrices have the same characteristic polynomial.*

Proof: Let $B = P^{-1}AP$. Then

$$\chi_B(t) = \det(P^{-1}AP - tI_n) = \det(P^{-1}AP - P^{-1}tI_nP) = \det(P^{-1}(A - tI_n)P)$$

Hence, by theorem **??**

$$\chi_B(t) = \det(P^{-1})\det(A - tI_n)\det P = \det(P^{-1})\chi_A(t)\det P$$

But determinants are just real numbers, so the above order of multiplication does not matter. Hence, since taking the determinant preserves multiplication, we have

$$\chi_B(t) = \det(P^{-1})\det P\chi_A(t) = \det(P^{-1}P)\chi_A(t) = \det I_n\chi_A(t) = \chi_A(t)$$

which completes the proof. □

Corollary 4.32 *Similar matrices have the same eigenvalues.* □

Recall that a matrix of the form

$$D = \begin{pmatrix} d_1 & 0 & \dots & 0 & 0 \\ 0 & d_2 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & d_{n-1} & 0 \\ 0 & 0 & \dots & 0 & d_n \end{pmatrix}$$

is called a *diagonal matrix*.

Such a matrix has characteristic polynomial

$$\chi_D(t) = \det\begin{pmatrix} d_1 - t & 0 & \dots & 0 & 0 \\ 0 & d_2 - t & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & d_{n-1} - t & 0 \\ 0 & 0 & \dots & 0 & d_n - t \end{pmatrix}$$

that is

$$\chi_D(t) = (d_1 - \lambda)(d_2 - \lambda)\cdots(d_n - \lambda)$$

Thus the eigenvalues of the matrix D are precisely the diagonal entries.

Definition 4.33 We call a matrix A *diagonalisable* if it is similar to a diagonal matrix.

In other words, a matrix A is diagonal if we have a diagonal matrix D and an invertible matrix P such that $P^{-1}AP$. The diagonal elements of the matrix D will be the eigenvalues of A.

Remark 4.34 Let

$$D = \begin{pmatrix} d_1 & 0 & \dots & 0 & 0 \\ 0 & d_2 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & d_{n-1} & 0 \\ 0 & 0 & \dots & 0 & d_n \end{pmatrix}$$

be a diagonal matrix. Then we can check by induction that

$$D^k = \begin{pmatrix} d_1^k & 0 & \dots & 0 & 0 \\ 0 & d_2^k & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & d_{n-1}^k & 0 \\ 0 & 0 & \dots & 0 & d_n^k \end{pmatrix}$$

for all $k \in \mathbb{N}$.

Suppose we have a matrix A and an invertible matrix P such that $P^{-1}AP = D$. Then $A = PDP^{-1}$, and

$$A^k = (PDP^{-1})(PDP^{-1}) \cdots (PDP^{-1}) = PD^kP^{-1}$$

so

$$A^k = P \begin{pmatrix} d_1^k & 0 & \dots & 0 & 0 \\ 0 & d_2^k & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & d_{n-1}^k & 0 \\ 0 & 0 & \dots & 0 & d_n^k \end{pmatrix} P^{-1}$$

Diagonalisation is the process of finding, for a matrix A, an invertible matrix P and diagonal matrix D such that $A = PDP^{-1}$. We call the matrix A *diagonalisable* if such matrices exist.

We do not really go into the question of whether or not a matrix is diagonalisable in this course. However, note that ”most” matrices are diagonalisable.

The following result presents a procedure for diagonalisation.

Theorem 4.35 *Let A be an $n \times n$ matrix, with n distinct eigenvalues $\lambda_1, \dots, \lambda_n$, and associated eigenvectors $v_1, \dots, v_n$. Set*

$$P = \left(\begin{array}{c|c|c|c} & & & \\ v_1 & v_2 & \dots & v_n \\ & & & \end{array} \right) \in M_n(\mathbb{R}) \qquad D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Then $A = PDP^{-1}$. □

Thus an $n \times n$ matrix with n distinct eigenvalues is diagonalisable. One can also make the above procedure work when we have a repeated eigenvalue; if an eigenvalue is repeated (for example) 3 times, then we need 3 linearly independent eigenvectors. However, we will not look at repeated eigenvalues in any more detail in this course.

Example 4.36 Diagonalise

$$A = \begin{pmatrix} 3 & -5 & 5 \\ 2 & -4 & 5 \\ 2 & -2 & 3 \end{pmatrix}$$

Solution: We can calculate that A has eigenvalues 1, -2 and 3, that $v_1 = \left(\begin{array}{c} 0 \\ 1 \\ 1 \end{array} \right)$ is an eigenvector of A corresponding to the eigenvalue 1, that $v_2 = \left(\begin{array}{c} 1 \\ 1 \\ 0 \end{array} \right)$ is an eigenvector of A corresponding to the eigenvalue -2, and that $v_3 = \left(\begin{array}{c} 1 \\ 1 \\ 1 \end{array} \right)$ is an eigenvector of A corresponding to the eigenvalue 3.

Since v_1, v_2, v_3 are eigenvectors of A corresponding to distinct eigenvalues of A, by the above we have an invertible matrix

$$P = \left(\begin{array}{c|c|c} v_1 & v_2 & v_3 \end{array} \right) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \text{ is such that } A = P \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{pmatrix} P^{-1}.$$

5 Combinatorics and Probability

5.1 Counting Elements of Sets

We call a set A *finite* if it contains only finitely many elements. For example, $\{1, 2, 3\}$ is finite whereas $\mathbb{Z}$ is not. We write $|A|$ to denote the number of elements in a finite set A.

Theorem 5.1 *Let A and B be finite sets. Then*

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Proof: Let

$$A \cap B = \{c_1, \ldots, c_k\}.$$

Then $A \cap B$ has n elements, so $|A \cap B| = k$. Every element of $A \cap B$ belongs to A, so we can add elements $a_1, \ldots, a_m$ to $A \cap B$ to get A. In other words, write

$$A = \{c_1, \ldots, c_k, a_1, \ldots, a_m\}$$

so $|A| = k + m$.

Similarly, we have $b_1, \ldots, b_n$ for which

$$B = \{c_1, \ldots, c_k, b_1, \ldots, b_n\}$$

so $|B| = k + n$.

Observe

$$A \cup B = \{c_1, \ldots, c_k, a_1, \ldots, a_m, b_1, \ldots, b_n\}$$

as we only write down an element of a set once. We see that $|A \cup B| = k + m + n$.

Now

$$|A| + |B| - |A \cap B| = (k + m) + (k + n) - k = k + m + n = |A \cup B|.$$

□

In particular, note that if $A \cap B = \emptyset$, then

$$|A \cup B| = |A| + |B|.$$

Our next result counting principle involves the Cartesian product of two sets. Recall that for sets A and B, the *Cartesian product* is the set of ordered pairs

$$A \times B = \{(a, b) \mid a \in A,\ b \in B\}.$$

If A and B are finite sets, then we have the formula

$$|A \times B| = |A| \cdot |B|.$$

Example 5.2 Let

$$A = \{1, 2, 3\} \qquad B = \{1, 2\}.$$

Then

$$A \times B = \{(1, 1),\ (1, 2),\ (2, 1),\ (2, 2),\ (3, 1),\ (3, 2)\}.$$

Observe

$$|A| = 3 \quad |B| = 2 \quad |A \times B| = 3 \times 2 = 6.$$

Similarly, if we have a set A and a positive integer r, we define A^r to be the set of ordered r-tuples of elements of A:

$$A^r = \{(a_1, a_2, \ldots, a_r) \mid a_i \in A\}.$$

If A is a finite set, with n elements, then we have

$$|A^r| = n^r.$$

Example 5.3 Consider a combination lock, with three digits, each of which is a number from 0 to 9. Then the set of all combinations is

$$\{0, 1, 2, \ldots, 9\}^3 = \{(a, b, c) \mid a, b, c = 0, 1, 2, \ldots, 9\}.$$

The total number of possible combinations in the lock is the number of elements of this set, which is

$$10^3 = 1000.$$

5.2 Permutations and Combinations

Example 5.4 How many different anagrams are there of the word BARK ? Here, any rearrangement of the letters counts as an anagram, whether or not it is a valid word.

Solution: To solve this problem, suppose we are writing down these four letters in some order. Observe:

- There are 4 different choices for the first letter.
- There remain $4 - 1 = 3$ different choices for the second letter.
- There are 2 choices for the third letter.
- Only 1 choice is left for the fourth letter.

So there are

$$4 \times 3 \times 2 \times 1 = 24$$

different anagrams.

Recall that for a positive integer r, we define r factorial:

$$r! = 1 \times 2 \times 3 \times \cdots \times r.$$

As a special case (to make certain formulae work), we write $0! = 1$.

Generalising the above example, if we have a set with r different elements, there are $r!$ different orders in which they can be arranged.

Example 5.5 Suppose we have a race with eight contestants. The one who comes first in the race gets a gold medal, the one who comes second gets a silver, and the one who comes third gets a bronze.

How many possible ways can medals be awarded ?

Solution:

To solve this problem, observe:

- There are 8 choices for the person who comes first, and gets gold.
- Knowing who gets gold, there are 7 choices for who gets silver.
- Knowing who gets gold and silver, there are 6 choices for who gets bronze.

So the total number of possibilities for how medals are awarded is:

$$8 \times 7 \times 6 = 336.$$

Theorem 5.6 *Suppose we have a set of n elements. The number of ways of ordering k of these n elements is*

$$P(n,k) = \frac{n!}{(n-k)!}.$$

□

To use this theorem in the above example, we want to order 3 of 8 elements. The number of ways of doing this is

$$P(8,3) = \frac{8!}{(8-3)!} = \frac{8!}{5!} = 336.$$

The reason the theorem works here (and in general) is that the above fraction cancels some of the factors in the top factorial. So here (and more generally)

$$\frac{8!}{5!} = \frac{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8}{1 \times 2 \times 3 \times 4 \times 5} = 6 \times 7 \times 8.$$

Theorem 5.7 *Suppose we have a set of n elements. The number of ways of choosing a subset of k elements is*

$$C(n,k) = \frac{n!}{k!(n-k)!}.$$

Proof: Let $C(n,k)$ be the number of different ways of choosing a subset of k elements. The number of ways of choosing an ordered set of k elements is

$$P(n,k) = \frac{n!}{(n-k)!}.$$

If we have a set of k elements, there are $k!$ different ways of putting it into order. So, each subset of k elements accounts for $k!$ different ordered sets.

In other words, $k!C(n,k) = P(n,k)$. Dividing

$$C(n,k) = \frac{P(n,k)}{k!} = \frac{n!}{k!(n-k)!}.$$

□

It is easiest to remember the above formula and use it directly in problems.

Example 5.8 How many ways are there of choosing a team of three people from a group of twelve?

Solution: We see we are choosing a subset of 3 elements from a set of 12, so the number of choices is just

$$C(12,3) = \frac{n!}{k!(n-k)!} = \frac{12 \times 11 \times 10}{3 \times 2 \times 1} = 220.$$

5.3 Probability Space

To study probability mathematically we need to formalise the concept.

Definition 5.9 A *probability space* (Ω, P) consists of a *sample space* Ω, and a *probability function*, P, which assigns a number $P(A)$, where $0 \leq P(A) \leq 1$, to each subset $A \subseteq \Omega$. This probability function is required to satisfy the axioms:

- $P(\emptyset) = 0$ and $P(\Omega) = 1$.
- If $A, B \subseteq \Omega$, and $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.

Subsets of the sample space Ω are called *events*. Given an event A, the number $P(A)$ is called the *probability* of A.

In the case of a set $A = \{a\}$, informally we write $P(a)$ instead of $P(A)$ or $P(\{a\})$.

Let $A_1, A_2, \ldots, A_n \subseteq \Omega$ be such that $A_i \cap A_j = \emptyset$. Then iterating the second of the above properties tells us

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

In particular, if $A = \{a_1, \ldots, a_n\}$ is a finite subset of Ω, then

$$P(A) = P(a_1) + P(a_2) + \cdots + P(a_n).$$

We think of the event $A \cup B$ as the event where A *or* B occurs.

Example 5.10 Consider tossing a coin. Then the sample space Ω is the set of possible outcomes; we have $\Omega = \{H, T\}$, where H stands for Heads and T stands for Tails.

Assuming the chance of each is equally likely, the probability function is defined by saying $P(H) = \frac{1}{2}$ and $P(T) = \frac{1}{2}$.

Example 5.11 Consider rolling a fair six-sided die. Then the sample space, Ω, is the set of possible dice rolls; we have $\Omega = \{1, 2, 3, 4, 5, 6\}$. The probability function is determined by saying that $P(k) = \frac{1}{6}$ for each $k \in \Omega$.

By the above, we have, for example

$$P(\{1, 3\}) = P(1) + P(3) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

The above examples illustrates an important principle. If the sample space Ω is finite, and each $\omega \in \Omega$ is equally likely, then $P(\omega) = \frac{1}{|\Omega|}$. More generally, for an event A,

$$P(A) = \frac{|A|}{|\Omega|}.$$

Example 5.12 Suppose a coin is tossed five times. Work out the probability of having precisely two heads in the sequence of tosses.

Solution: Note that each possible sequence, such as $HTTHH$ is equally likely. There are $2^5 = 32$ such sequences.

The total number of sequences with precisely two heads is the number we get from choosing 2 elements from a set of 5, that is

$$C(5, 2) = \frac{5!}{2!3!} = \frac{5 \times 4}{2} = 10.$$

So the probability of exactly two heads is

$$P(\text{Two heads}) = \frac{10}{32} = \frac{5}{16}.$$

Example 5.13 Suppose a coin is tossed five times. Work out the probability of having two or fewer heads in the sequence of tosses.

Solution: A slightly harder question is to work out the probability of there being two or fewer heads. Well, above, we worked out that the probability of exactly two heads is $\frac{5}{16}$.

Similarly, there are

$$C(5, 1) = \frac{5!}{1!}4! = 5$$

sequences with just one head, so

$$P(\text{One head}) = \frac{5}{32}.$$

There is precisely one sequence with no heads- the sequence of all tails- so

$$P(\text{No heads}) = \frac{1}{32}.$$

Now, the set of sequences with two or fewer heads is the union of the sets of sequences with no, one, and two heads. The intersection of any two of these sets is zero, so

$$P(\text{Two or fewer heads}) = \frac{1}{32} + \frac{5}{32} + \frac{5}{16} = \frac{16}{32} = \frac{1}{2}.$$

Given events A and B in a probability space, we think of the probability $P(A \cap B)$ as the probability of event A *and* event B occurring.

Definition 5.14 We call two events, A and B, in a probability space *independent* if $P(A \cap B) = P(A)P(B)$.

Example 5.15 Consider tossing a coin twice. Let A be the set of sequences of two tosses where the first toss is a head. Let B be the set of sequences of two tosses where the second toss is a head. Then $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, and $P(A \cap B) = \frac{1}{4}$.

Thus A and B are independent.

Recall that if we have a subset $X \subseteq Y$, we define the *difference*

$$Y \backslash X = \{y \in Y \mid y \notin X\}.$$

The following is sometimes useful in working out examples. The proof is a straightforward exercise.

Proposition 5.16 *Let* (Ω, P) *be a probability space. Let* A *be an event. Then*

$$P(\Omega \backslash A) = 1 - P(A).$$

□

Example 5.17 Suppose a coin is tossed five times. Work out the probability of having at least one head in the sequence of tosses.

Solution: Let Ω be the sample space consisting of all sequences of five coin tosses. Let A be the event where we have at least one head. Then $\Omega \backslash A$ is the event where we have no heads, that is all tails.

There are $2^5 = 32$ possible sequences of coin tosses, all equally likely. There is only one sequence where all we have is tails. So

$$P(\Omega \backslash A) = \frac{1}{32}$$

and by the above

$$P(A) = \frac{31}{32}.$$

5.4 Conditional Probability

Let A and B be events in a probability space (Ω, P). We can think of $A \cap B$ as the event where both A and B occur. If it is impossible for A and B to occur simultaneously, then $A \cap B = \emptyset$, and we have our usual formula

$$P(A \cup B) = P(A) + P(B).$$

More generally, we have the following.

Proposition 5.18 *Let (Ω, P) be a probability space. Let A and B be events. Then*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Proof: Let $C = A \backslash (A \cap B)$. Then $A = (A \cap B) \cup C$, and $(A \cap B) \cap C = \emptyset$. Hence

$$P(A) = P(A \cap B) + P(C).$$

Similarly, if $D = B \backslash (A \cap B)$, then

$$P(B) = P(A \cap B) + P(D).$$

Now,

$$A \cup B = C \cup D \cup (A \cap B)$$

and the intersection of any two of the sets in the union on the right is empty. So

$$P(A \cup B) = P(C) + P(D) + P(A \cap B) = P(A) + P(B) - P(A \cap B)$$

by the above. □

Definition 5.19 Let (Ω, P) be a probability space. Let A and B be events, where $P(B) > 0$. Then we define the *conditional probability* of A given B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

We think of the conditional probability $P(A|B)$ as the probability that A occurs if we already know that B occurs.

Observe that if A and B are independent, with non-zero probabilities, then $P(A \cap B) = P(A)P(B)$, so

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

and similarly $P(B|A) = P(B)$. Thus for independent events, the presence of one does not affect the probability of the other.

The following is easy to check.

Proposition 5.20 *The pair $(B, P(-|B))$ is a probability space.* □

Example 5.21 A fair six-sided die is rolled twice. Find the probability that the first number thrown is no larger than 3 given that that sum of the two numbers thrown is 6.

Solution: Let A be the event where the first number thrown is 3 or lower. Let B be the event that the sum of the two numbers thrown is 6. Then we need to calculate $P(A|B)$.

In this problem, our event space is

$$\Omega = \{(i, j) \mid i, j \in \{1, 2, 3, 4, 5, 6\}\}.$$

Observe $|\Omega| = 6 \times 6 = 36$.

We have

$$A = \{(i, j) \mid i \in \{1, 2, 3\},\ j \in \{1, 2, 3, 4, 5, 6\}$$

and

$$B = \{(1, 5),\ (2, 4),\ (3, 3),\ (4, 2),\ (5, 1)\}$$

so

$$A \cap B = \{(1, 5),\ (2, 4),\ (3, 3)\}.$$

We see that

$$P(A \cap B) = \frac{3}{36} = \frac{1}{12} \qquad P(B) = \frac{5}{36}$$

so

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3}{5}.$$

Definition 5.22 Let (Ω, P) be a probability space. A *partition* of Ω is a collection $B_1, B_2, \ldots, B_n$ of events such that $B_i \cap B_j = \emptyset$ for all i, j, and $B_1 \cup \cdots \cup B_n = \Omega$.

Note that if $B_1, B_2, \ldots, B_n$ is a partition of Ω, then

$$P(B_1) + P(B_2) + \cdots + P(B_n) = 1.$$

The following result is sometimes called the *partition theorem.*

Theorem 5.23 *Let (Ω, P) be a probability space. Let $B_1, B_2, \ldots, B_n$ be a partition of Ω. Then for any event A*

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n).$$

Proof: Observe

$$(A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n) = A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) = A \cap \Omega = A.$$

We have

$$P(A|B_i)P(B_i) = P(A \cap B_i)$$

and $(A \cap B_i) \cap (A \cap B_j) = \emptyset$ if $i \neq j$. Therefore, from the above

$$P(A) = P(A\cap B_1)+P(A\cap B_2)+\cdots+P(A\cap B_n) = P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\cdots+P(A|B_n)P(B_n)$$

as required. □

Example 5.24 The probability of someone having a particular disease is 0.001. There is a test for the disease, where:

- If the test is given to a person with the disease, the test is positive with probability 0.99.
- If the test is given to a person without the disease, the test is negative with probability 0.95.

If a person's test is positive, what is the probability that the person has the disease ?

Solution: Consider a person being tested. Let A be the event where they have the disease, B be the event where they test positive. Write $A^c = \Omega \backslash A$ and $B^c = \Omega \backslash B$.

We are being asked to calculate $P(A|B)$. We know that $P(A) = 0.001$, $P(B|A) = 0.99$ and $P(B^c|A^c) = 0.95$. Hence $P(B^c|A) = 0.01$, and $P(B|A^c) = 0.05$.

We know that A, A^c is a partition of the sample space. So by the partition theorem

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Now by definition of conditional probability and the above

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Hence

$$P(A|B) = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.05 \times 0.999} \approx 0.019.$$

Thus, if the test is positive, the chance of having the disease is 0.019. This is, on the surface, a surprising result. It reflects the rarity of the disease and comparative unreliability of the test.

6 Variables and Processes

6.1 Random Variables

Definition 6.1 A *random variable* is a function $X\colon \Omega \to \mathbb{R}$, where (Ω, P) is a probability space.

Example 6.2 We can define a random variable X to be the sum of two six-sided dice. To square this with the formal definition, when rolling two six-sided dice, the sample space is

$$\Omega = \{(i,j) \mid i,j \in \{1,2,3,4,5,6\}\}$$

and the random variable X is defined by the formula

$$X(i,j) = i + j.$$

The interesting things about random variables are their possible values, and the probabilities they take those values.

Definition 6.3 Let (Ω, P) be a probability space, and let $X\colon \Omega \to \mathbb{R}$ be a random variable. Let $x \in \mathbb{R}$. Then we write

$$P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\}).$$

For a subset $A \subseteq \mathbb{R}$, we write

$$P(X \in A) = P(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

Example 6.4 Consider the above random variable, X, obtained by adding two six-sided dice together. Let

$$A = \{(i,j) = \mid i,j \in \{1,2,3,4,5,6\},\ i+j = 6\}.$$

Then $P(X = 6) = P(A)$. Observe

$$A = \{(1,5),\ (2,4),\ (3,3),\ (4,2),\ (5,1)\}$$

so $|A| = 5$. On the other hand, $|\Omega| = 36$, and the dice are fair, so all outcomes are equally likely. Hence

$$P(X = 6) = P(A) = \frac{5}{36}.$$

Definition 6.5 We call a random variable X *discrete* if for all $A \subseteq \mathbb{R}$ we have

$$P(X \in A) = \sum_{a \in A} P(X = a).$$

For a discrete random variable, the function $p_X\colon \mathbb{R} \to \mathbb{R}$ defined by the formula $p_X(x) = P(X = x)$ is called the *probability mass function*.

The probability mass function determines everything we need to know about a discrete random variable. We look at some detailed examples in the next section.

Proposition 6.6 *Let X be a discrete random variable with probability mass function p_X. Then*

$$\sum_{x \in \mathbb{R}} p_X(x) = 1.$$

Proof: Note that

$$P(X \in \mathbb{R}) = P(\{\omega \in \Omega \mid X(\omega) \in \mathbb{R}\}) = P(\Omega) = 1$$

since every value of X is in $\mathbb{R}$.

By definition of a discrete random variable

$$\sum_{x \in \mathbb{R}} p_X(x) = P(X \in \mathbb{R}) = 1.$$

□

Definition 6.7 Let X be a discrete random variable, with probability mass function p_X. Then we define the *expectation*

$$E(X) = \sum_{x \in \mathbb{R}} x p_X(x).$$

Note that in the above sum, we need only add the elements $xp_X(x)$ together when the probability $P_X(x) = P(X = x)$ is not zero, that is to say when x is a possible value of the random variable X.

We think of $E(X)$ as the average value of X.

Example 6.8 Let X be the random variable obtained by taking the result of rolling a fair 6-sided die. Then we have probability mass function

$$p_X(x) = \begin{cases} \frac{1}{6} & x \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases}$$

Hence the range of possible values is $\{1,2,3,4,5,6\}$ and we have expectation

$$E(X) = \sum_{n=1}^{6} \frac{n}{6} = \frac{21}{6} = \frac{7}{2}.$$

Note that if X is a random variable, and $f \colon \mathbb{R} \to \mathbb{R}$ is a function, then we have a new random variable $f(X) = f \circ X$ defined by composing these functions.

By definition

$$E(f(X)) = \sum_{x \in \mathbb{R}} f(x) p_X(x).$$

Definition 6.9 Let X be a discrete random variable, with expectation $\mu = E(X)$. Then we define the *variance*

$$var(X) = E((X - \mu)^2).$$

We define the *standard deviation* $sd(X) = \sqrt{var(X)}$.

The standard deviation of X is a measure of its randomness. We think of the standard deviation of X as the average distance of X from its expectation. There are some short-cuts in working out the variance.

Proposition 6.10 *Let X be a discrete random variable. Then*

$$var(X) = E(X^2) - E(X)^2.$$

Proof: Observe

$$var(X) = \sum_{x\in\mathbb{R}} (x - \mu)^2 p_X(x) = \sum_{x\in\mathbb{R}} x^2 p_X(x) - 2\mu \sum_{x\in\mathbb{R}} x p_X(x) + \mu^2 \sum_{x\in\mathbb{R}} p_X(x).$$

Now

$$\sum_{x\in\mathbb{R}} x^2 p_X(x) = E(X^2) \quad \sum_{x\in\mathbb{R}} x p_X(x) = \mu \quad \sum_{x\in X} p_X(x) = 1.$$

Putting this together

$$var(X) = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2.$$

□

Example 6.11 Let X be the random variable obtained by taking the result of rolling a fair 6-sided die. Then we have probability mass function

$$p_X(x) = \left\{ \begin{array}{ll} \frac{1}{6} & x \in \{1,2,3,4,5,6\} \\ 0 & \text{otherwise} \end{array} \right.$$

By the previous example, we have expectation $E(X) = \frac{7}{2}$. Now

$$E(X^2) = \sum_{n=1}^{6} \frac{n^2}{6} = \frac{91}{6}$$

so

$$var(X) = E(X^2) = E(X)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12} \approx 2.917$$

Taking the square root, we see

$$sd(X) \approx 1.708.$$

6.2 Examples

Here we note some facts about some standard discrete random variables. All calculations are left as exercises.

Definition 6.12 Let $0 \leq p \leq 1$. We say a random variable X has the *Bernoulli distribution* with parameter p if the range of possible values of X is $\{0, 1\}$, and $P(X = 1) = p$, $P(X = 0) = 1 - p$.

Example 6.13 Suppose we have a coin that lands on the heads side with probability p, and the tails side with probability $1 - p$. Define a random variable X by tossing the coin, and seting $X = 1$ when we get heads, and $X = 0$ when we get tails. Then X has the Bernoulli distribution.

Proposition 6.14 *Let X be a random variable with the Bernoulli distribution with parameter p. Then $E(X) = p$, and $var(X) = p(1 - p)$.* □

Definition 6.15 Let $n \in \{1, 2, 3, \ldots\}$ and $0 \leq p \leq 1$. We say the random variable X has the *binomial distribution* with parameters n and p if the range of possible values of X is $\{0, 1, 2, \ldots, n\}$, and

$$P(X = k) = C(n, k)p^k(1 - p)^{n-k}.$$

Example 6.16 Suppose we have a (perhaps biased) coin that lands on the heads side with probability p, and the tails side with probability $1 - p$. Define a random variable X by tossing the coin n times and counting the number of heads.

Then $P(X = k)$ is the probability that our sequence of n tosses contains precisely k heads. There are $C(n, k)$ sequences of tosses which have k out of n heads. The chance of each of them occurring is $p^k(1 - p)^{n-k}$, as we need heads k times (each with probability p), and tails $n - k$ times (each with probability $1 - p$). Therefore

$$P(X = k) = C(n, k)p^k(1 - p)^{n-k}$$

and the random variable has a binomial distribution.

Proposition 6.17 *Let X be a random variable with the binomial distribution with parameters n and p. Then $E(X) = np$, and $var(X) = np(1 - p)$.* □

More generally, if we add n independent random variables with Bernoulli distribution with parameter p together, we get a random variable with binomial distribution with parameters n and p. We will not make this statement precise in this course, however.

Definition 6.18 Let $\lambda > 0$. We say a random variable X has the *Poisson distribution* with parameter λ if the range of possible values of X is $\{0, 1, 2, \ldots\}$, and

$$P(X = k) = \frac{1}{k!}\lambda^k e^{-\lambda}.$$

Proposition 6.19 *Let X be a random variable with the Poisson distribution with parameter λ. Then $E(X) = \lambda$, and $var(X) = \lambda$.* □

6.3 Difference Equations

Definition 6.20 A *difference equation* is a vector equation of the form $v_{k+1} = Av_k$, wher (v_k) is a sequence of vectors in $\mathbb{R}^n$, and $A \in M_n(\mathbb{R})$ is a fixed square matrix.

Solving a difference equation means finding the vector v_k when we know v_0. Since $v_1 = Av_0$, $v_2 = Av_1 = A^2v_0$, and so on; we have $v_k = A^k v_0$ for all $k \in \mathbb{N}$.

So we want to find the matrix A^k. This can be hard to do directly, especially for a larger matrix- finding A^10 for example, means multiplying the matrix A by itself 10 times.

As we commented in section 4.5, the key to solving this problem is to diagonalise A. Suppose we have a diagonal matrix

$$D = \begin{pmatrix} d_1 & 0 & \dots & 0 & 0 \\ 0 & d_2 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & d_{n-1} & 0 \\ 0 & 0 & \dots & 0 & d_n \end{pmatrix}$$

be a diagonal matrix. Then

$$D^k = \begin{pmatrix} d_1^k & 0 & \dots & 0 & 0 \\ 0 & d_2^k & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & d_{n-1}^k & 0 \\ 0 & 0 & \dots & 0 & d_n^k \end{pmatrix}.$$

If we diagonalise A, we find a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$, and

$$A^k == PD^kP^{-1}.$$

As we commented above, D^k is easy to find, even for large D and large k.

Example 6.21 Consider the difference equation $v_{k+1} = Av_k$, where

$$A = \begin{pmatrix} 6 & -3 \\ 5 & -2 \end{pmatrix} \qquad v_0 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Find v_9.

Solution:

- Step 1: We find the eigenvalues of the matrix A.

  We have characteristic polynomial

  $$\chi_A(t) = \begin{vmatrix} 6-t & -3 \\ 5 & -2-t \end{vmatrix} = (6-t)(-2-t)+15 = t^2-4t+3 = (t-1)(t-3).$$

  So we have eigenvalues 1 and 3.

- Step 2: Find corresponding eigenvectors.

  Let X be an eigenvector with eigenvalue 1. Then $(A-I)X = 0$. This system of linear equations has matrix

  $$\left(\begin{array}{cc|c} 5 & -3 & 0 \\ 5 & -3 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 5 & -3 & 0 \\ 0 & 0 & 0 \end{array}\right).$$

  We see that $\begin{pmatrix} 3 \\ 5 \end{pmatrix}$ is a 1-eigenvector.

  Let X be an eigenvector with eigenvalue 3. Then $(A-3I)X = 0$. This system of linear equations has matrix

  $$\left(\begin{array}{cc|c} 3 & -3 & 0 \\ 5 & -5 & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right).$$

  We see that $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is a 3-eigenvector.

- Step 3: Diagonalise. Set

  $$D = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \qquad P = \begin{pmatrix} 3 & 1 \\ 5 & 1 \end{pmatrix}.$$

  Then $A = PDP^{-1}$.

- Step 4: Find P^{-1}.

  We can use Gaussian elimination here

  $$\begin{aligned} (P|I) &\sim \left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 5 & 1 & 0 & 1 \end{array}\right) \\ &\sim \left(\begin{array}{cc|cc} 1 & \frac{1}{3} & \frac{1}{3} & 0 \\ 5 & 1 & 0 & 1 \end{array}\right) \\ &\sim \left(\begin{array}{cc|cc} 1 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & -\frac{2}{3} & -\frac{5}{3} & 1 \end{array}\right) \\ &\sim \left(\begin{array}{cc|cc} 1 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 1 & \frac{5}{2} & -\frac{3}{2} \end{array}\right) \end{aligned}$$

$$\sim \left(\begin{array}{cc|cc} 1 & 0 & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{5}{2} & -\frac{3}{2} \end{array} \right)$$

so

$$P^{-1} = \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{5}{2} & -\frac{3}{2} \end{pmatrix}$$

- Step 5: Work out A^k.

  We have

$$A^k = PD^kP^{-1} = \begin{pmatrix} 3 & 1 \\ 5 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 3^k \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{5}{2} & -\frac{3}{2} \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 3^k \\ 5 & 3^k \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{5}{2} & -\frac{3}{2} \end{pmatrix} = \begin{pmatrix} -\frac{3}{2} + \frac{5}{2}(3^k) & \frac{3}{2} - \frac{3}{2}(3^k) \\ -\frac{5}{2} + \frac{5}{2}(3^k) & \frac{5}{2} = \frac{3}{2}(3^k) \end{pmatrix}.$$

- Step 6: Work out v_k.

  We have

$$v_k = A^k v_0 = \begin{pmatrix} -\frac{3}{2} + \frac{5}{2}(3^k) & \frac{3}{2} - \frac{3}{2}(3^k) \\ -\frac{5}{2} + \frac{5}{2}(3^k) & \frac{5}{2} = \frac{3}{2}(3^k) \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -3 + 3^k \\ -5 + 3^k \end{pmatrix}.$$

  In particular, we were asked for v_9, which is

$$v_9 = \begin{pmatrix} -3 + 5^9 \\ -5 + 3^9 \end{pmatrix} = \begin{pmatrix} 19680 \\ 19678 \end{pmatrix}.$$

6.4 Markov Processes

Let us begin our discussion of Markov processes with an example.

Example 6.22 Suppose that bus passengers in Sheffield are studied. After examining several years of data, it was found that 30% of people who regularly use buses in a given year do not regularly use the bus in the next year. Also it was found that 20% of the people who do not regularly use the bus in one year begin to use the bus next year. If 30% of people used the bus regularly in 2012, what proportion can we predict will use the bus regularly in 2014 ? What about k years after 2012 ?

Solution:

First we will determine the proportion people who will use the bus in 2013. Of the 30% of people using the bus in 2012, 70% of them will continue to do so. Of the remaining 70% of people who dont use the bus, 20% of them will begin to use the bus.

Let b_1 be the proportion using the bus regularly in 2013. Then

$$b_1 = 0.7 \times 0.3 + 0.2 \times 0.7$$

Let b_2 be the proportion not using the bus regularly in 2013. Then by the same argument as above,

$$b_2 = 0.3 \times 0.3 + 0.8 \times 0.7$$

This is equivalent to the matrix equation $Mx = b$ where

$$M = \begin{pmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{pmatrix} \quad x = \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Note that we calculate

$$b = \begin{pmatrix} 0.35 \\ 0.65 \end{pmatrix}.$$

So in 2013, 35% of people will use the bus regularly. For computing the result after 2 years, in 2014, we just use the same matrix M, however we use b in place of x. Thus the distribution after 2 years is given by the vector

$$Mb = M^2x.$$

In fact, after k years, the distribution is given by the vector M^kx. We can solve this as for any other difference equation, by diagonalising M.

Definition 6.23 A *stochastic process* is a sequence of events in which the outcome at any stage depends on some probability.

Definition 6.24 A *Markov process* is a stochastic process with the following properties:

- The number of possible outcomes or states is finite.
- The outcome at any stage depends only on the outcome of the previous stage.
- The probabilities are constant over time.

In a Markov process, let $\{\omega_1, \omega_2, \ldots, \omega_n\}$ be the set of states. After k stages, let p_i be the probability we are in state ω_i, and let

$$v_k = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix}.$$

We call v_k the *probability vector* at stage k. Notice that the numbers p_i must add up to 1.

When moving from one stage to another of a Markov process, the probability of moving to state ω_i when we are in state ω_j, $P(\omega_i|\omega_j)$, stays constant at each stage. Let

$$M = \begin{pmatrix} P(\omega_1|\omega_1) & P(\omega_1|\omega_2) & \cdots & P(\omega_1|\omega_n) \\ P(\omega_2|\omega_1) & P(\omega_2|\omega_2) & \cdots & P(\omega_2|\omega_n) \\ \vdots & & & \vdots \\ P(\omega_n|\omega_1) & P(\omega_k|\omega_2) & \cdots & P(\omega_n|\omega_n) \end{pmatrix}$$

be the *transition matrix*.

Notice that the numbers in each column of M add up to 1. The definition of a Markov process immediately gives us the following.

Theorem 6.25 *In a Markov process with probability vectors v_k and transition matrix M, we have $v_{k+1} = Mv_k$ for all k.* □

Thus a Markov process is described by a difference equation. It follows from our work in the previous section that our solution is given by

$$v_k = M^k v_0.$$

We can find M^k by diagonalising M. The following result can help us find the eigenvalues.

Theorem 6.26 *Let M be the transition matrix of a Markov process. Then*

1. *1 is an eigenvalue of M.*
2. *If λ is an eigenvalue of M, then $|\lambda| \leq 1$.*

□

Our next result involves the long-term behaviour of Markov processes.

Theorem 6.27 *Consider a Markov process with probability vectors (v_k) and transition matrix M. Suppose that the eigenvalue, 1, of M is not repeated. Let v be a 1-eigenvector where the entries add up to 1. Then for large values of k, $v_k \approx v$.*

Proof: We have $A^k = PD^kP^{-1}$ where

$$D = \begin{pmatrix} 1 & 0 & & 0 \\ 0 & \lambda_1 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

where $|\lambda_i| < 1$. For large k,

$$D^k \approx \begin{pmatrix} 1 & 0 & & 0 \\ 0 & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{pmatrix} =: C$$

meaning

$$A^k v_0 \approx PCP^{-1}v_0.$$

Now, the entries of $A^k v_0 = v_k$ always add up to 1. So the same is true for v. Further,

$$APCP^{-1}v_0 = PDCP^{-1}v_0 = PCP^{-1}v_0$$

so $v = PCP^{-1}v_0$ is a 1-eigenvector. There is only one 1-eigenvector whose entries add up to 1, so we are done. □

Example 6.28 Let us return to our example of bus travellers in Sheffield. Recall that 30% of people who regularly use buses in a given year do not regularly use the bus in the next year. 20% of the people who do not regularly use the bus in one year begin to use the bus next year.

What proportion of people can we predict will use the bus in many years time ?

Solution: We have a Markov process with transition matrix

$$M = \begin{pmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{pmatrix}.$$

We know that 1 is an eigenvalue of M. By the above, we need to find a 1-eigenvector of M whose entries add up to 1.

An eigenvector of M corresponding to the eigenvalue 1 is a non-zero vector $v \in \mathbb{R}^3$ such that $(M - 1I)v = 0$. We have

$$M - 1I = \begin{pmatrix} -0.3 & 0.2 \\ 0.3 & 0.2 \end{pmatrix}$$

so it is clear that

$$u = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

is an eigenvector of M corresponding to the eigenvalue 1. The entries of u do not add up to 1- they add up to 2. So we divide M by 2 to get a new eigenvector

$$v = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

The entries of the vector v do add up to 1. So v is the probability vector governing the proportion of people using the bus in the long term.

In other words, in the long term, it is predicted that 50% of people will regularly use the bus, and 50/

Example 6.29 In the first year, every member of a certain sample of multinational European corporations required its board members to travel to board meetings by air.

However, with increasing awareness of the contributions to climate change caused by air travel, it is expected that, over the coming years, the firms will change their instructions in this respect over time. In detail, it is expected that, by the end of each period of twelve months:

- of those firms that required air travel to board meetings at the beginning of the period, 90% will still be using air travel, 9% will have changed to high-speed rail travel, and 1% will have eliminated the need for any travel to board meetings by use of video-conferences;
- of those firms that required high-speed rail travel to board meetings at the beginning of the period, 4% will have changed back to air travel, 90% will still be using high-speed rail travel, and 6% will have eliminated the need for any travel to board meetings by use of video-conferences; and
- each firm that was using video-conferences for its board meetings at the beginning of the period would still be doing so at the end.

In the long term, after many years, what proportion of the firms will be using air travel, high-speed rail travel, and video conferences ?

Solution: This is a Markov process, with probability vectors

$$v_k = \begin{pmatrix} \text{Proportion using air travel after } k \text{ years.} \\ \text{Proportion using high speed raid travel after } k \text{ years.} \\ \text{Proportion using video conferencing after } k \text{ years.} \end{pmatrix}$$

and transition matrix

$$M = \begin{pmatrix} 0.9 & 0.04 & 0 \\ 0.09 & 0.9 & 0 \\ 0.01 & 0.06 & 1 \end{pmatrix}.$$

We know that 1 is an eigenvalue of M. By the above, we need to find a 1-eigenvector of M whose entries add up to 1.

An eigenvector of M corresponding to the eigenvalue 1 is a non-zero vector $v \in \mathbb{R}^3$ such that $(M - 1I)v = 0$. We have

$$M - 1I = \begin{pmatrix} -0.1 & 0.04 & 0 \\ 0.09 & -0.1 & 0 \\ 0.01 & 0.06 & 0 \end{pmatrix},$$

So it is clear that

$$v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

is an eigenvector of M corresponding to the eigenvalue 1. The entries of the vector v add up to 1. So v is the probability vector governing the proportion of firms using the various forms of travel in the long term.

In other words, in the long term, it is predicted that 100% of firms will be using video conferences, with none using air travel or high-speed rail travel.

6.5 PageRank

The PageRank algorithm was developed by Larry Page in 1996, and used by the Google search engine. It is a system for ranking web pages according to their use, and so determine which is returned first in a query from a search engine. The key insight is that the higher ranked pages are more likely to be linked to.

Consider a collection of web pages, $S_1, S_2, \ldots, S_n$. Imagine surfing the web between these pages randomly. At each step, choose a link on the current page at random and click it. Let r_i be the average long term proportion of time we spend on a particular page, S_i. Then $r_i \geq 0$ for all i, and $\sum_{i=1}^{n} r_i = 1$.

We use this number r_i to rank page S_i. So how do we calculate it ?

Assume that:

- Each page S_j contains at least one link to another page in the collection.
- No page contains more than one link to the same target.

Let N_j be the total number of links on page S_j. Then $N_j > 0$, and N_j is the number of pages S_j links to. Define a matrix P by saying

$$P = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{nn} \\ \vdots & & & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{pmatrix}$$

where $P_{ij} = 1/N_j$ if there is a link from S_j to S_i, and 0 otherwise.

Thus, when surfing the web randomly, we have a Markov process with transition matrix P. Let r be the probability vector with entries r_i. Then r is the probability vector governing the long-term behaviour of the process.

In other words, r is an eigenvector of P with eigenvalue 1, and where the sum of the entries is 1. We calculate this to find the rank, r_i, of each page.

Unfortunately, when the number n is very large (millions, or even billions), as it will be when considering the number of web pages, r will be unfeasably hard to calculate.

Of course, since r governs the long term behaviour of the Markov process, we have the following.

Proposition 6.30 *Let q be the vector where the entries are all $1/n$. Then for large k, $P^k q \approx r$.* □

Actually, we don't need k very large for the above to settle down. And $P^k q$ is quite quick to calculate when k is small, at least for a computer.

Now, let us introduce a refinement to the algorithm. This refinement uses the fact that a web surfer will eventually stop clicking on links. The probability at each step that a person will continue is a number $0 < d < 1$ called the *damping factor*. Various studies have tested damping factors- it is generally set at 0.85.

Anyway, in this case we have a Markov process where a link is clicked with probability d, as above. With probability $1 - d$, the surfing stops. The next time the surfer starts, they choose a page at random. If we do this, we have a Markov process with transition matrix

$$Q = dP + (1 - d)R$$

where R is the $n \times n$ matrix where *every* entry is n. The page

As before, let r be the vector where the entry r_i is the long term proportion of time spent on page S_i. Then again we calculate r to be the 1-eigenvector of Q whose entries add up to 1.