Chapter 3

Difference Equations

In Section 1.1 we saw some examples of sequences that were specified by giving some of the terms of the sequence, and then defining further terms as a function of previous terms. This relationship between terms of the sequence is known as a *difference equation*, or *recurrence relation*. Together with the given *initial conditions* it uniquely determines the remaining terms of the sequence. For example, an arithmetic progression is defined by the difference equation $a_n = a_{n-1} + d$, along with an initial condition consisting of the value of a_0.

Difference equations have many applications, and can be used in modelling problems in various disciplines including biology, economics and computer science. We will begin this chapter by looking at some examples that demonstrate how sequences defined by difference equations can arise and illustrate the sorts of properties of these sequences that we are interested in studying. After discussing basic definitions and terminology we will then consider techniques for *solving* difference equations in Section 3.2, before considering various applications of difference equations in more detail in Section 3.3.

Example 3.1 (biology)**.** A single bacterium is introduced into a test tube. Under a simple mathematical model, suppose the number of bacteria doubles every hour. After how many hours will the number of bacteria exceed 10,000,000?

Solution. Let u_t be the number of bacteria after t hours. We require the smallest t for which $u_t \geq 10,000,000 = 10^7$. Now, $u_0 = 1$ and $u_{t+1} = 2u_t$. So we have $u_0 = 1$, $u_1 = 2$, $u_2 = 4$, $u_3 = 8$ *etc.* and in general $u_t = 2^t$. Thus

$$\begin{aligned}
u_t \geq 10,000,000 &\Leftrightarrow 2^t \geq 10^7 \\
&\Leftrightarrow \log_{10} 2^t \geq \log_{10} 10^7 \\
&\Leftrightarrow t \log_{10} 2 \geq 7 \\
&\Leftrightarrow t \geq \frac{7}{\log_{10} 2} = 23.5
\end{aligned}$$

So the number of bacteria will exceed 10^7 after 24 hours.

Example 3.2 (numerical analysis)**.** The equation $x^5 - 3x - 20 = 0$ cannot be solved explicitly. However we can attempt a solution with difference equations by using the following rearrangements, starting with the plausible $x_1 = 2$.

1. $x = \sqrt[5]{3x + 20}$ so try $x_{n+1} = (3x_n + 20)^{\frac{1}{5}}$, $x_1 = 2$.

2. $x = \frac{1}{3}(x^5 - 20)$ so try $x_{n+1} = \frac{1}{3}(x_n^5 - 20)$, $x_1 = 2$

3. $x = \frac{20}{x^4-3}$ so try $x_{n+1} = 20(x_n^4 - 3)^{-1}$, $x_1 = 2$

4. $x = \sqrt[4]{\frac{3x+20}{x}}$ so try $x_{n+1} = \sqrt[4]{\frac{3x_n+20}{x_n}}$, $x_1 = 2$

5. $x_{n+1} = x_n - \frac{x_n^5-3x_n-20}{5x_n^4-3}$, $x_1 = 2$ (Newton-Raphson)

Which relations converge? Which converge fast, so are most useful in practice? It may depend on the choice of x_1.
Now, 1. gives the sequence

$$1.918645, 1.915029, 1.914868, 1.914861, 1.914861$$

2. gives

$$2, 4, 334.667, 1.4 \times 10^{12}, 1.79 \times 10^{60}, 6.1 \times 10^{300}$$

3. gives

$$2, 1.538462, 7.686263, 0.005735, -6.6667, 0.0101$$

4. gives

$$2, 1.898829, 1.917993, 1.914253, 1.914979, 1.91483$$

5. gives

$$2, 1.401639, 1.918613, 1.91512, 1.914792$$

It seems that 2. and 3. diverge, 4. and 5. converge but 1. converges fastest and so is best in this case. If we could decide in advance of the iteration, we would have a powerful method for obtaining approximate roots of equations.

Example 3.3 (sequences)**.** One famous example of a difference equation is the *Fibonacci sequence*,

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55 \ldots$$

which is given by $u_t = u_{t-1} + u_{t-2}$ for $t \geq 3$, with $u_1 = u_2 = 1$. We will see later in this section how we can find an explicit formula for the n^{th} term of this sequence.

If we consider the ratio u_t/u_{t-1} of successive terms of this sequence we observe that they appear to be converging:

$$\frac{1}{1} = 1, \frac{2}{1} = 2, \frac{3}{2} = 1.5, \frac{5}{3} = 1.67, \frac{8}{5} = 1.6, \frac{13}{8} = 1.625, \frac{21}{13} = 1.615385$$

We seek to find $\lim_{t\to\infty} \frac{u_t}{u_{t-1}}$, which must equal $\lim_{t\to\infty} \frac{u_{t-1}}{u_{t-2}} = L$, say.

But $u_t = u_{t-1} + u_{t-2}$, which implies $\frac{u_t}{u_{t-1}} = 1 + \frac{u_{t-2}}{u_{t-1}}$. Thus, taking limits of both sides of this equation, we have that $L = 1 + \frac{1}{L}$ (making some assumptions about the algebra of limits...). Hence $L^2 = L + 1$ and thus $L = \frac{1+\sqrt{5}}{2} = 1.618034$ (taking the positive root).

3.1 Definitions

We now introduce some of the terminology used in describing difference equations.

Definition 3.1. The difference equation for the sequence $u_0, u_1, u_2, \ldots$ is said to have *order r* if u_k is expressed in terms of $u_{k-1}, u_{k-2}, \ldots, u_{k-r}$.

To be able to evaluate the terms of a sequence defined by a difference equation of order r uniquely, we need to know the values of $u_0, u_1, \ldots, u_{r-1}$.

Definition 3.2. A difference equation is *linear* if it is of the form

$$u_k = f_1(k)u_{k-1} + f_2(k)u_{k-2} + \cdots + f_r(k)u_{k-r} + F(k),$$

where $f_1, f_2, \ldots, f_r, F$ are functions of k.

If $F(k) = 0\ \forall k$, the difference equation is said to be *homogeneous*, otherwise it is *inhomogeneous*.

If $f_1, f_2, \ldots, f_r$ are constants, the difference equation is said to have *constant coefficients*.

So $u_k = u_{k-1}^2$ is a non-linear first order difference equation, $a_n = 3a_{n-1} + na_{n-2}$ is a homogeneous second order linear difference equation (whose coefficients are not constants,) $f_{n+1} = f_n + f_{n-1}$ is a homogeneous linear second order equation with constant coefficients, $b_{n+3} = b_{n+2} + 2b_n + n^2$ is an inhomogeneous linear difference equation with constant coefficients of order 3.

3.2 Solving Difference Equations

To *solve* a difference equation is to find an explicit formula that gives the n^{th} term of the corresponding sequence as a function of n (*i.e.* a formula that doesn't involve any of the previous terms of the sequence). There is no general method to solve all difference equations. However, we will see how first order linear difference equations and higher order linear equations with constant coefficients may be approached, plus some possible approaches for other equations.

3.2.1 First Order Linear Difference Equations

A first order linear difference equation is one that can be written in the form

$$u_n = f(n)u_{n-1} + g(n), \quad u_0 = U.$$

We will first consider the solution to the homogeneous case (where g=0), then go on to examine how to approach the inhomogeneous case, before seeing some special examples of inhomogeneous first order linear difference equations whose solutions can be expressed in a straightforward form.

Homogeneous First Order Linear Difference Equations

A homogeneous first order linear difference equation has the form

$$u_n = f(n)u_{n-1}, \quad u_0 = U.$$

Observe that

$$\begin{aligned}
u_0 &= U, \\
u_1 &= f(1)u_0 = f(1)U, \\
u_2 &= f(2)u_1 = f(2)f(1)U, \\
u_3 &= f(3)u_2 = f(3)f(2)f(1)U,
\end{aligned}$$

and hence, in general,

$$U_n = U \cdot \prod_{i=1}^{n} f(i). \tag{3.1}$$

Example 3.4. Solve the difference equation given by

$$u_n = nu_{n-1}, \quad u_0 = 1.$$

Solution. This is a homogeneous, first order, linear difference equation. Note that $f(i) = i$ for all i, so using (3.1) we see that $u_n = \prod_{i=1}^{n} i = n!$.

Inhomogeneous First Order Linear Difference Equations

An inhomogeneous first order linear difference equation has the form

$$u_n = f(n)u_{n-1} + g(n), \quad u_0 = U.$$

Observe that

$$\begin{aligned}
u_0 &= U, \\
u_1 &= f(1)u_0 + g(1) = f(1)U + g(1), \\
u_2 &= f(2)u_1 + g(2) = f(2)f(1)U + f(2)g(1) + g(2), \\
u_3 &= f(3)u_2 + g(3) = f(3)f(2)f(1)U + f(3)f(2)g(1) + f(3)g(2) + g(3), \\
&\vdots \\
u_n &= f(n)u_{n-1} + g(n) = U \cdot \prod_{i=1}^{n} f(i) + \sum_{i=1}^{n} g(i) \cdot \prod_{j=i+1}^{n} f(j). \qquad (3.2)
\end{aligned}$$

We have formulated this result by observing how the first few terms of the sequence arise; it could be proved formally using induction on n.

Example 3.5. Solve the difference equation given by

$$u_n = \frac{n+1}{n} u_{n-1} + (n+1)^3, \quad u_0 = 3.$$

Solution. Using (3.2), we have

$$\begin{aligned}
u_n &= 3 \cdot \prod_{i=1}^{n} \frac{i+1}{i} + \sum_{i=1}^{n} (i+1)^3 \cdot \prod_{j=i+1}^{n} \frac{j+1}{j} \\
&= 3(n+1) + \sum_{i=1}^{n} (i+1)^3 \frac{n+1}{i+1} \\
&= (n+1)\left(3 + \sum_{i=1}^{n} (i+1)^2\right) \\
&= (n+1)\left(3 + \sum_{i=1}^{n} i^2 + 2\sum_{i=1}^{n} i + \sum_{i=1}^{n} 1\right) \\
&= (n+1)\left(3 + \frac{1}{6}n(n+1)(2n+1) + n(n+1) + n\right) \\
&= \frac{1}{6}(n+1)\left(2n^3 + 9n^2 + 13n + 18\right).
\end{aligned}$$

Example 3.6. Solve the difference equation given by

$$u_n = 3nu_{n-1} + n!, \quad u_0 = 9.$$

Solution. Using (3.2), we have

$$\begin{aligned}
u_n &= 9 \cdot \prod_{i=1}^{n} 3i + \sum_{i=1}^{n} i! \cdot \prod_{j=i+1}^{n} 3j \\
&= 9 \cdot 3^n n! + \sum_{i=1}^{n} i! \cdot 3^{n-i} \frac{n!}{i!} \\
&= 3^n n! \left(9 + \sum_{i=1}^{n} 3^{-i}\right) \\
&= 3^n n! \left(9 + \frac{\frac{1}{3}\left(1 - \left(\frac{1}{3}\right)^n\right)}{1 - \frac{1}{3}}\right) \quad \text{(sum of a geometric progression)} \\
&= 3^n n! \left(9 + \frac{1}{2}(1 - 3^{-n})\right) \\
&= \frac{1}{2} n! \left(19 \cdot 3^n - 1\right).
\end{aligned}$$

Special Cases of First Order Linear Difference Equations

There are certain particular choices for the functions f and g that lead to equation (3.2) having an especially simple form. Here we list some of them.

1. $\mathbf{f(k) = 1}\ \forall \mathbf{k}$
   In this case we have that $U_n = U + \sum_{i=1}^{n} g(i)$.

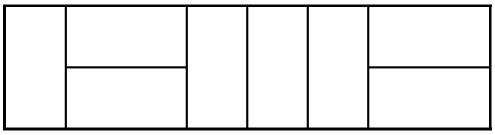


Figure 3.1: A paving of a 2m × 8m path.

2. $\mathbf{g(k) = 0}\ \forall \mathbf{k}$
   This is simply the case of a homogeneous linear first order difference equation, and thus we have that $U_n = U \cdot \prod_{i=1}^{n} f(i)$.

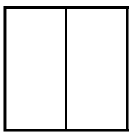
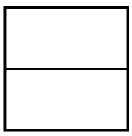
3. $\mathbf{f(k) = R},\ \mathbf{g(k) = 0}$
   In this case the resulting sequence is a geometric progression, and $U_n = R^n U$.

4. $\mathbf{f(k) = R}$ **and** $\mathbf{g(k) = d}$
   We considered difference equations of this form in Section 1.1, and we deduced that $U_n = U \cdot R^n + d\frac{R^n - 1}{R-1}$. We could also obtain this result by substituting $f(k) = R$ and $g(k) = d$ into (3.2).

3.2.2 Second Order Linear Difference Equations with Constant Coefficients

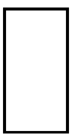
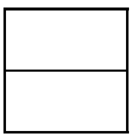
We have seen some methods for finding solutions for first order linear difference equations. We will now see how solutions can be found for higher order linear difference equations with constant coefficients, focusing mainly on the case of second order equations. We begin with an example of a counting problem that gives rise to such an equation.

Example 3.7. Suppose a 2 m × n m rectangular path is paved with 2 m × 1 m rectangular pavers (Fig. 3.1 gives an example of this for the case $n = 8$). How many different pavings are possible?
Solution. Let u_n be the number of pavings of a 2 m × n m rectangular path. When $n = 1$, there is just one way in which the path can be paved, so $u_1 = 1$. When $n = 2$, there are two ways, namely

 and  ,

so $u_2 = 2$.

Now, consider a paving of a 2 m × n m path. It must start in one of two ways, either

a)  or b)  .

If it begins in manner a), then the rest of the pavement is a paving of a 2 m × $(n-1)$ m path. If it starts as in b), then the rest of the pavement is a paving of a 2 m × $(n-2)$ m path. Thus, we have that $u_n = u_{n-1} + u_{n-2}$.

So, we see that the solution to this problem can be described by a homogeneous second order linear difference equation with constant coefficients. Evaluating the first few terms of the corresponding sequence gives

$$1, 2, 3, 5, 8, 13, 21, 34, \ldots$$

which is the familiar Fibonacci sequence (but starting with $u_0 = 1$, $u_1 = 2$ rather than the usual $u_0 = u_1 = 1$).

We will now explore a technique that will let us find an expression for the n^{th} term of this sequence, and other similar sequences.

The Homogeneous Case

Consider a homogeneous second order linear difference equation with constant coefficients, given by

$$\begin{aligned} u_n &= au_{n-1} + bu_{n-2}, \\ u_0 &= U, \ u_1 = V. \end{aligned} \tag{3.3}$$

(Recall that a second order difference equation requires two initial conditions.)

Suppose we look for a solution of the form $u_n = w^n$, for some constant $w \neq 0$. Substituting this into (3.3), we obtain

$$w^n = aw^{n-1} + bw^{n-2}.$$

Dividing through by w^{n-2} tells us that $w^2 - aw - b = 0$. That is, w^n satisfies (3.3) if and only if w is a zero of the polynomial $\lambda^2 - a\lambda - b$.

Definition 3.3. Given a homogeneous second order linear difference equation with constant coefficients described by the equation $u_n = au_{n-1} + bu_{n-2}$, we say the *characteristic polynomial* (sometimes referred to as the *auxiliary polynomial*) of this difference equation is the quadratic polynomial given by

$$\lambda^2 - a\lambda - b.$$

Since the characteristic polynomial is a quadratic polynomial, it either has two distinct real zeros, one repeated real zero, or two complex zeros that are complex conjugates of each other. There are separate techniques for solving the corresponding difference equation in each case.

distinct real zeros The characteristic polynomial $\lambda^2 - a\lambda - b$ has distinct real zeros when $a^2 + 4b > 0$. Denote these zeros by w_1 and w_2. We have established that w_1^n and w_2^n satisfy (3.3). In fact, since (3.3) describes a linear difference equation, any linear combination of w_1^n and w_2^n also satisfies (3.3), as for any $c_1, c_2 \in \mathbb{R}$ we have

$$\begin{aligned} a(c_1w_1^{n-1} + c_2w_2^{n-1}) + b(c_1w_1^{n-2} + c_2w_2^{n-2}) &= c_1(aw_1^{n-1} + bw_1^{n-2}) + c_2(aw_2^{n-1} + bw_2^{n-2}), \\ &= c_1w_1^n + c_2w_2^n. \end{aligned}$$

However, to solve the difference equation we need to find a linear combination that also satisfies both initial conditions. That is, we want to find constants c_1 and c_2 such that $u_n = c_1w_1^n + c_2w_2^n$ satisfies both initial conditions. Each linear condition gives rise to an equation in the unknowns c_1 and c_2, and the fact that we have a linear difference equation implies that they will be linear equations. Specifically, we require that

$$c_1 + c_2 = U \tag{3.4}$$

$$c_1 w_1 + c_2 w_2 = V \tag{3.5}$$

(The first of these equations comes from substituting $u_n = c_1 w_1^n + c_2 w_2^n$ into the first initial condition, and the second equation comes from substituting that expression into the second initial condition.) We observe that since $w_1 \neq w_2$ this system of equations always gives rise to a unique solution for c_1 and c_2. Once we have determined these values then we have a complete solution for our difference equation. A word of caution: the initial conditions may not always be given for terms u_0 and u_1. For instance, if the sequence is labelled starting from u_1 you may be given conditions on u_1 and u_2. In any case, each condition will give rise to a linear equation in c_1 and c_2, with a condition for u_i giving rise to an equation of the form $c_1 w_1^i + c_2 w_2^i = u_i$.

Example 3.8. Find a formula for the n^{th} term of the Fibonacci sequence.
Solution. The Fibonacci sequence is defined by the difference equation $u_n = u_{n-1} + u_{n-2}$, with the initial conditions $u_0 = 1$, $u_1 = 1$. The characteristic polynomial for this difference equation is thus $\lambda^2 - \lambda - 1$. Using the quadratic formula, we see that the zeros of this polynomial are $\frac{1 \pm \sqrt{1+4}}{2}$, so we can take $w_1 = \frac{1+\sqrt{5}}{2}$ and $w_2 = \frac{1-\sqrt{5}}{2}$. In order to complete the solution to this problem, we need to solve equations (3.4) and (3.5). This leads to the following system of equations:

$$\begin{aligned} c_1 + c_2 &= 1, \\ c_1 \frac{1+\sqrt{5}}{2} + c_2 \frac{1-\sqrt{5}}{2} &= 1. \end{aligned}$$

Solving these equations gives $c_1 = \frac{\sqrt{5}+1}{2\sqrt{5}}$ and $c_2 = \frac{\sqrt{5}-1}{2\sqrt{5}}$. Thus the complete solution to this difference equation is given by

$$\begin{aligned} u_n &= \frac{\sqrt{5}+1}{2\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n + \frac{\sqrt{5}-1}{2\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n \\ &= \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} \right]. \end{aligned}$$

This seems like a surprising result, since this expression involves $\sqrt{5}$, whereas we know that the Fibonacci numbers are all integers. However, if you compute the first few terms of this sequence, you will notice that the square roots all cancel out. If you need further convincing that this expression is indeed correct, you can try proving it for yourself using induction.

This method can be immediately generalised for solving higher order homogeneous linear difference equations with constant coefficients: such a difference equation of order d gives rise to a characteristic polynomial of degree d. If it has d distinct real zeros $w_1, w_2, \ldots, w_d$ then the difference equation has a solution of the form $u_n = c_1 w_1^n + c_2 w_2^n + \cdots + c_d w_d^n$, where the constants c_i for $i = 1, 2, \ldots, d$ can be determined by solving the system of d linear equations arising from the d initial conditions for the difference equation.

complex zeros The characteristic polynomial $\lambda^2 - a\lambda - b$ has two complex zeros that are conjugates of each other when $a^2 + 4b < 0$. In this case we will denote the zeros by $z, \bar{z} \in \mathbb{C}$.

To solve this difference equation, we express z in polar form, *i.e.* we write $z = r\cos(\theta) + ir\sin(\theta)$ where $r, \theta \in \mathbb{R}$. Once we have done this, it turns out that a general solution to the difference equation can be written in the form

$$u_n = Ar^n \cos(n\theta) + Br^n \sin(n\theta).$$

We then use the initial conditions to solve for the real constants A and B.

Example 3.9. Solve the difference equation $u_n = 2u_{n-1} - 2u_{n-2}$ subject to initial conditions $u_0 = 2$, $u_1 = 3$.
Solution. The characteristic polynomial for this difference equation is $\lambda^2 - 2\lambda + 2$, which has zeros $\frac{2 \pm \sqrt{4-8}}{2} = 1 \pm i$. We begin by writing $1 + i = \sqrt{2}\cos(\pi/4) + i\sqrt{2}\sin(\pi/4)$. Next we write down a general solution for this difference equation:

$$u_n = A\sqrt{2}^n \cos(n\pi/4) + B\sqrt{2}^n \sin(n\pi/4).$$

Finally, we solve for A and B:

$$\begin{aligned}
u_0 &= A\sqrt{2}^0 \cos(0) + B\sqrt{2}^0 \sin(0), \\
&= A, \\
&= 2. \\
u_1 &= A\sqrt{2}\cos(\pi/4) + B\sqrt{2}\sin(\pi/4), \\
&= 2\sqrt{2}\frac{1}{\sqrt{2}} + B\sqrt{2}\frac{1}{\sqrt{2}}, \\
&= 2 + B, \\
&= 3.
\end{aligned}$$

Thus we have $A = 2$, $B = 1$, and so the final solution is:

$$u_n = 2\sqrt{2}^n \cos(n\pi/4) + \sqrt{2}^n \sin(n\pi/4).$$

repeated real zero A quadratic polynomial $\lambda^2 - a\lambda - b$ has a repeated zero w if $a^2 + 4b = 0$. In this case, the polynomial is of the form $(\lambda - w)^2 = \lambda^2 - 2w\lambda + w^2$, and hence the corresponding difference equation is of the form $u_n = 2wu_{n-1} - w^2 u_{n-2}$. If we try to solve this difference equation using the same technique as we did when we had distinct roots, we find that the equations (3.4) and (3.5) will not have a solution in general, since $w_1 = w_2$. The trouble is that we have two equations in just one unknown.

However, we observe that in this case we have that nw^n also satisfies (3.3), since

$$\begin{aligned}
2w(n-1)w^{n-1} - w^2(n-2)w^{n-2} &= 2w^n(n-1) - w^n(n-2) \\
&= nw^n.
\end{aligned}$$

Thus we seek a solution of the form $u_n = c_1 w^n + c_2 n w^n$ that satisfies both initial conditions, by solving the simultaneous equations

$$c_1 = u_0, \tag{3.6}$$
$$c_1 w + c_2 w = u_1. \tag{3.7}$$

It is clear that for $w \neq 0$, we have the solution $c_1 = u_0$ and $c_2 = \frac{u_1}{w} - c_1$.

Example 3.10. Find a solution to the difference equation $u_n = 4u_{n-1} - 4u_{n-2}$ where $u_0 = 1$ and $u_1 = 4$.
Solution. The characteristic polynomial for this difference equation is $\lambda^2 - 4\lambda + 4$, which factors as $(\lambda - 2)^2$, and hence 2 is a repeated zero of this polynomial. We solve the equations

$$c_1 = 1, \tag{3.8}$$
$$2c_1 + 2c_2 = 4 \tag{3.9}$$

to obtain $c_1 = 1$ and $c_2 = 1$. Thus we conclude that the solution to this difference equation is

$$\begin{aligned} u_n &= 2^n + n2^n \\ &= (1+n)2^n. \end{aligned}$$

This technique also applies to higher order homogeneous linear difference equations: if the characteristic polynomial of a homogeneous linear difference equation of order $v_1 + v_2 + \cdots + v_k$ has (real or complex) zeros $w_1, w_2, \ldots w_k$ that occur with multiplicity $v_1, v_2, \ldots, v_k$ respectively, then it has a solution of the form

$$\begin{aligned} (c_{10} + c_{11}n + c_{12}n^2 + \cdots + c_{1(v_1-1)}n^{v_1-1})w_1^n + \\ (c_{20} + c_{21}n + c_{22}n^2 + \cdots + c_{2(v_2-1)}n^{v_2-1})w_2^n + \\ \vdots \\ (c_{k0} + c_{k1}n + c_{k2}n^2 + \cdots + c_{k(v_k-1)}n^{v_k-1})w_k^n, \end{aligned}$$

where the c_{ij} are found by solving the system of $v_1 + v_2 + \cdots + v_k$ linear equations determined by the initial conditions of the difference equation.

Exercise 3.11.

1. Find a solution to the difference equation $u_n = 5u_{n-1} - 6u_{n-2}$, where $u_0 = 4$ and $u_1 = 1$.
2. Find a solution to the difference equation $u_n = 2u_{n-1} - 4u_{n-2}$, where $u_0 = 2$ and $u_1 = -1$.
3. Find a solution to the difference equation $u_n = 6u_{n-1} - 9u_{n-2}$, where $u_0 = 2$ and $u_1 = 2$.

The Inhomogeneous Case

We have seen that homogeneous linear difference equations with constant coefficients can be solved by finding a general solution to the difference equation, then using the initial conditions to determine the value of the constants in the solution. We now turn our attention to difference equations of the form

$$\begin{aligned} u_n &= au_{n-1} + bu_{n-2} + f(n), \qquad (3.10)\\ u_0 &= U,\\ u_1 &= V. \end{aligned}$$

Let $G(n)$ denote the general solution of the homogeneous part of this difference equation, *i.e.* the general solution to the equation

$$u_n = au_{n-1} + bu_{n-2}. \qquad (3.11)$$

We observe that if $P(n)$ is a *particular solution* satisfying (3.10), then $P(n)+G(n)$ also satisfies (3.10), since

$$\begin{aligned} &a(P(n-1) + G(n-1)) + b(P(n-2) + G(n-2)) + f(n)\\ &= [aP(n-1) + bP(n-2) + f(n)] + [aG(n-1) + bG(n-2)],\\ &= P(n) + G(n). \end{aligned}$$

This suggests that we can find a solution to a second order linear inhomogeneous difference equation with constant coefficients in the following manner:

1. **Find a general solution $G(n)$ to the corresponding homogeneous difference equation (3.11).**
2. **Find a particular solution $P(n)$ satisfying (3.10).**
3. **A general solution to (3.10) is given by $P(n) + G(n)$.**
4. **Use the initial conditions to determine the values of the constants in this solution.**

(Note that it is essential that these steps be carried out in this specific order!)

We have already seen techniques suitable for carrying out steps 1, 3 and 4 of the above method, so now we address the problem of how to perform step 2. Unfortunately this is hard in general, but there are several special cases for which appropriate techniques are known. Figure 3.2 contains a list of these techniques.

1. If $\mathbf{f(n) = c\boldsymbol{\alpha}^n}$ and $\boldsymbol{\alpha}$ **is *not* a zero of the characteristic polynomial**, try a solution of the form $M\alpha^n$, and determine M by substituting in (3.10).

2. If $\mathbf{f(n) = c\boldsymbol{\alpha}^n}$ and $\boldsymbol{\alpha}$ ***is* a zero of the characteristic polynomial of multiplicity k**, choose a particular solution $Mn^k\alpha^n$ and substitute to find M.

3. If $\mathbf{f(n)}$ **is a polynomial of degree s in n** and **1 is not a zero of the characteristic polynomial**, try a polynomial of degree s, say $M_0 + M_1 n + \cdots + M_s n^s$ as a particular solution, and substitute to find $M_0, M_1, \ldots, M_s$.

4. If $\mathbf{f(n)}$ **is a polynomial of degree s in n** and **1 *is* a zero of the characteristic polynomial of multiplicity k**, try $n^k(M_0 + M_1 n + \cdots + M_s n^s)$ and proceed as before.

5. If $\mathbf{f(n)}$ **contains several terms**, the particular solution will be the sum of the solutions for each term separately.

Figure 3.2: Techniques for finding particular solutions to special cases of inhomogeneous linear second order difference equations with constant coefficients.

Example 3.12. Find a solution to the difference equation $v_{n+2} - 7v_{n+1} + 10v_n = 30 \cdot 7^n$, with $v_0 = 2, v_1 = 1$.
Solution. The characteristic polynomial for the homogeneous part of this difference equation is

$$\lambda^2 - 7\lambda + 10;$$

its zeros are 2 and 5, so the general solution for the homogeneous part is $v_n = A \cdot 5^n + B \cdot 2^n$.
For a particular solution, we try $v_n = M \cdot 7^n$ (case 1. in Figure 3.2). Then we have

$$\begin{aligned} M \cdot 7^{n+2} - 7M \cdot 7^{n+1} + 10M \cdot 7^n &= 30 \cdot 7^n \\ 49M - 49M + 10M &= 30 \end{aligned}$$

so $10M = 30$ and $M = 3$.
A general solution for the inhomogeneous equation is thus given by $v_n = A \cdot 5^n + B \cdot 2^n + 3 \cdot 7^n$. The initial conditions imply that

$$\begin{aligned} A + B + 3 &= 2, \\ 5A + 2B + 21 &= 1, \end{aligned}$$

from which we deduce that $A = -6$ and $B = 5$. So we have

$$v_n = 5 \cdot 2^n - 6 \cdot 5^n + 3 \cdot 7^n.$$

Example 3.13. Find a solution to the difference equation $v_{n+2} - 7v_{n+1} + 10v_n = 12n + 9$, with $v_0 = 2, v_1 = 1$.

Solution. The general solution for the homogeneous part is $v_n = A \cdot 5^n + B \cdot 2^n$, as before. We try for a particular solution of the form $M_0 + M_1 n$ (case 3.).

$$\begin{aligned} M_0 + M_1(n+2) - 7M_0 - 7M_1(n+1) + 10M_0 + 10M_1 n &= 12n + 9 \\ 4M_0 - 5M_1 + 4M_1 n &= 12n + 9 \end{aligned}$$

so we have

$$\begin{aligned} 4M_1 &= 12, \\ 4M_0 - 5M_1 &= 9, \end{aligned}$$

which gives $M_1 = 3$ and $M_0 = 6$, leading to the general solution $v_n = A \cdot 5^n + B \cdot 2^n + 6 + 3n$. Thus we have

$$\begin{aligned} A + B + 6 &= 2, \\ 5A + 2B + 9 &= 1, \end{aligned}$$

which gives $A = 0$ and $B = -4$. So $v_n = -4 \cdot 2^n + 6 + 3n = 6 + 3n - 2^{n+2}$.

Example 3.14. Find a general solution to the difference equation $a_{n+2} + a_{n+1} - 2a_n = 3(-2)^n$.
Solution. The characteristic polynomial is

$$\lambda^2 + \lambda - 2,$$

which has zeros 1 and -2. Thus the general solution of the homogeneous part is $A(-2)^n + B1^n = A(-2)^n + B$. We seek a particular solution of the form $Mn(-2)^n$ (case 2). Then we have

$$\begin{aligned} M(n+2)(-2)^{n+2} + M(n+1)(-2)^{n+1} - 2Mn(-2)^n &= 3(-2)^n \\ 4M(n+2) - 2M(n+1) - 2Mn &= 3 \\ 6M &= 3, \end{aligned}$$

so $M = 1/2$. The general solution is thus $a_n = A(-2)^n + B + \frac{1}{2}n(-2)^n$.

Example 3.15. Find a solution to the difference equation $a_{n+2} + a_{n+1} - 2a_n = 162n - 3$, with $a_0 = 1, a_1 = 3$.
Solution. The general solution to the homogeneous part is $A(-2)^n + B$, as before. We seek a particular solution of the form $n(M_0 + M_1 n)$ (case 4.)

$$\begin{aligned} M_0(n+2) + M_1(n+2)^2 + M_0(n+1) + M_1(n+1)^2 - 2M_0 n - 2M_1 n^2 &= 162n - 3 \\ n^2(M_1 + M_1 - 2M_1) + n(M_0 + 4M_1 + M_0 + 2M_1 - 2M_0) + (2M_0 + 4M_1 + M_0 + M_1) &= 162n - 3 \\ 6M_1 n + 3(M_0 + 5M_1) &= 162n - 3, \end{aligned}$$

so we have

$$\begin{aligned} 6M_1 &= 162 \\ 3M_0 + 5M_1 &= -3, \end{aligned}$$

which yields $M_1 = 27$ and $M_0 = -46$. The general solution is thus $a_n = A(-2)^n + B + 27n^2 - 46n$. Substituting into the initial conditions gives

$$\begin{aligned} A + B &= 1, \\ -2A + B + 27 - 46 &= 3 \end{aligned}$$

so we have $A = -7$ and $B = 8$. Thus $a_n = 8 - 7(-2)^n + 27n^2 - 46$.

The techniques we have just discussed also apply directly in the case of higher order inhomogeneous linear difference equations with constant coefficients.

Example 3.16. Consider the third order linear difference equation with constant coefficients $a_n + 6a_{n-1} + 12a_{n-2} + 8a_{n-3} = \Theta(n)$. The characteristic polynomial is $\lambda^3 + 6\lambda^2 + 12\lambda + 8 = (\lambda+2)^3$, and so the general solution to the homogeneous part is $(An^2+Bn+C)(-2)^n$. Depending on the form of Θ, we now look for a particular solution.

- If $\Theta(n) = 9$, try $a_n = M$ as a particular solution.
- If $\Theta(n) = 3^n$, try $a_n = M \cdot 3^n$ as a particular solution.
- If $\Theta(n) = (-2)^n$, try $a_n = Mn^3(-2)^n$.
- If $\Theta(n) = 9 + 3^n$, try $a_n = M_0 + M_1 \cdot 3^n$
- *etc.*

In addition, first order linear difference equations with constant coefficients can also be approached by these methods.

Example 3.17. Solve the difference equation $a_n = a_{n-1} + 12n^2$, with $a_0 = 5$.
Solution. *EITHER* (using the method described in Section 3.2.1) We have $f(n) = 1$, $g(n) = 12n^2$, and $a_0 = 5$. So

$$\begin{aligned} a_n &= U + \sum_{i=1}^{n} 12i^2 \\ &= 5 + 12\frac{1}{6}n(n+1)(2n+1) \\ &= 5 + 2n(n+1)(2n+1). \end{aligned}$$

OR The homogeneous equation $a_n - a_{n-1} = 0$ has characteristic polynomial $\lambda - 1$, whose only zero is 1. Thus we have the general solution $a_n = A \cdot 1^n = A$.

Now we look for a particular solution. Since 1 is a root of the characteristic polynomial, try $a_n = n(M_0 + M_1 n + M_2 n^2)$. Substituting this into the difference equation we obtain

$$\begin{gathered} M_0 n + M_1 n^2 + M_2 n^3 - M_0(n-1) - M_1(n-1)^2 - M_2(n-1)^3 = 12n^2 \\ M_0 + 2M_1 n - M_1 + 3M_2 n^2 - 3M_2 n + M_2 = 12n^2. \end{gathered}$$

Equating like terms and solving, we obtain $M_2 = 4$, $M_1 = 6$ and $M_0 = 2$. Thus a_n is the sum of A and $n(2 + 6n + 4n^2)$, *i.e.* $a_n = A + 2n(1 + 3n + 2n^2)$. But $a_0 = 5$ implies $A = 5$, so we obtain the solution $5 + 2n(n+1)(2n+1)$, which agrees with our earlier result.

In this instance the first method was faster than the second, although this will not always be the case.

3.2.3 Additional Techniques for Solving Difference Equations

We have seen methods for solving various types of linear difference equation. Here we mention a couple of techniques that can sometimes be used to solve difference equations to which the above techniques cannot be applied directly.

Looking for a pattern Sometimes we can deduce a solution by examining the first few terms of the sequence arising from a difference equation.

Example 3.18. Find a solution to the difference equation $u_n = u_{n-1}^2 - u_{n-2}^2$, with $u_1 = 1$ and $u_2 = 0$.
Solution. This is not a linear difference equation, and hence the techniques we have seen so far cannot be used. However, we observe that the sequence arising from this difference equation begins

$$1, 0, -1, 1, 0, -1, 1, 0, -1, 1, \ldots$$

So we can guess that the sequence continues in the same pattern; we can confirm this by using induction.

Solving via a substitution Sometimes we can use a *substitution* to turn a non-linear difference equation into a linear one that we can solve.

Example 3.19. Suppose $a_n - a_{n-1} = 3a_n a_{n-1}$, and $a_0 = 1$. This is a first order, nonlinear difference equation. If we divide through by a_n and a_{n-1} we can write it in the form

$$\frac{1}{a_{n-1}} - \frac{1}{a_n} = 3.$$

If we let $b_n = \frac{1}{a_n}$ (so $b_{n-1} = \frac{1}{a_{n-1}}$,) then we can express the difference equation in the form $b_{n-1} - b_n = 3$, with $b_0 = a_0^{-1} = 1$. This is now a linear first order equation (in fact it describes an arithmetic progression) and hence we can solve it to find $b_n = 1 - 3n$. We can then conclude that $a_n = \frac{1}{1-3n}$, thus solving the original difference equation.

Exercise 3.20. Solve the following difference equations:

1. $u_n = n^2 u_{n-1} + (n!)^2$, $u_0 = 2$
2. $a_{n+2} - 2\sqrt{3}a_{n+1} + 4a_n + 2^n = 0$, $a_0 = a_1 = 1$
3. $u_n = 2u_{n-1} - u_{n-2} - 2n + 2$, $u_0 = 1$, $u_1 = 2$

Exercise 3.21. Express the solution to each of the following counting problems in terms of a difference equation, then use a generating function to solve the difference equation to give an explicit solution.

1. I am making a border for the front of a garden bed out of small bricks that are 10cm long and larger bricks that are 20 cm long. If the small bricks come in two colours, and the larger bricks come in three possible colours, how many different borders of length $10n$ cm can I make?

2. Every week my grandmother knits either a hat that uses one ball of wool, or a scarf that uses two balls of wool. If the hats are all black and the scarves can be blue, red, or green, yellow, purple or orange, how many different ways are there for her to use up n balls of wool? (We assume that the order in which she knits the items matters.)

3.2.4 Generating Functions and Difference Equations

In Chapter 2 we saw that generating functions can be useful in the study of sequences; this is equally true in the case of the sequences associated with linear homogeneous difference equations with constant coefficients. We have seen how to solve such difference equations by making use of the characteristic polynomial. An alternative approach is to use generating functions to solve these equations.

Example 3.22. Consider the difference equation $u_n - 4u_{n-1} + 4u_{n-2} = 0$ with $u_0 = 1$ and $u_1 = 4$. Let $g(x) = \sum_{i=0}^{\infty} u_i x^i$ be the generating function for the corresponding sequence $(u_i)_{i=0}^{\infty}$. We observe that

$$\begin{array}{rcccc} g(x) = & u_0 & +u_1 x & +u_2 x^2 + \cdots \\ -4xg(x) = & & -4u_0 x & -4u_1 x^2 + \cdots \\ 4x^2 g(x) = & & & 4u_0 x^2 + \cdots \end{array}$$

The sum of the left hand sides of these three equations is $(1 - 4x + 4x^2)g(x)$. The sum of the right hand sides is $u_0 + (u_1 - 4u_0)x$, since for $n \geq 2$ the coefficient of x^n in this sum is $u_n - 4u_{n-1} + 4u_{n-2} = 0$. Thus we conclude that

$$(1 - 4x + 4x^2)g(x) = u_0 + (u_1 - 4u_0)x$$

and so

$$\begin{aligned} g(x) &= \frac{u_0 + (u_1 - 4u_0)x}{1 - 4x + 4x^2} \\ &= \frac{1}{1 - 4x + 4x^2}. \end{aligned}$$

Thus we have found the generating function corresponding to the sequence arising from the given difference equation.

If we can find an expression for the coefficients of this generating function we will have a solution to our difference equation. We observe that

$$\begin{aligned} \frac{1}{1 - 4x + 4x^2} &= \frac{1}{(1 - 2x)^2} \\ &= (1 + 2x + (2x)^2 + (2x)^3 + \cdots)^2 \text{ (Identity 3)} \\ &= \sum_{r=0}^{\infty} \binom{2 - 1 + r}{r} (2x)^r \text{ (Identity 4)} \\ &= \sum_{r=0}^{\infty} (r + 1)2^r x^r. \end{aligned}$$

Thus we conclude that the solution to the difference equation is $u_n = (n + 1)2^n$. (Fortunately this agrees with the answer we obtained to the same problem in Example 3.10!)

This approach can be applied to any homogeneous linear difference equation with constant coefficients. Be aware that in some cases it may be necessary to use *partial fractions* to find the coefficients of the resulting generating function.

Example 3.23. Use generating functions to find a formula for the n^{th} term of the Fibonacci sequence.
Solution. We need to solve the difference equation $u_n - u_{n-1} - u_{n-2} = 0$ with $u_0 = u_1 = 1$. Following the approach of Example 3.22, we observe that the generating function $g(x)$ for the Fibonacci sequence must satisfy

$$\begin{aligned}(1 - x - x^2)g(x) &= u_0 + (u_1 - u_0)x \\ &= 1\end{aligned}$$

and hence

$$\begin{aligned}g(x) &= \frac{1}{1 - x - x^2} \\ &= \frac{1}{\left(1 - \frac{1-\sqrt{5}}{2}x\right)\left(1 - \frac{1+\sqrt{5}}{2}x\right)}\end{aligned} \tag{3.12}$$

The partial fraction decomposition of (3.12) has the form

$$= \frac{A}{1 - \frac{1-\sqrt{5}}{2}x} + \frac{B}{1 - \frac{1+\sqrt{5}}{2}x}, \tag{3.13}$$

where A and B are constants. Equating (3.12) and (3.13) and solving for the constants we can determine that $A = \frac{\sqrt{5}-1}{2\sqrt{5}}$ and $B = \frac{\sqrt{5}+1}{2\sqrt{5}}$. Hence the generating function can be written

$$\begin{aligned}g(x) &= \left(\frac{\sqrt{5}-1}{2\sqrt{5}}\right)\sum_{i=0}^{\infty}\left(\frac{1-\sqrt{5}}{2}x\right)^i + \left(\frac{\sqrt{5}+1}{2\sqrt{5}}\right)\sum_{i=0}^{\infty}\left(\frac{1+\sqrt{5}}{2}x\right)^i \quad \text{(Identity 3)} \\ &= \sum_{i=0}^{\infty}\left[\left(\frac{\sqrt{5}-1}{2\sqrt{5}}\right)\left(\frac{1-\sqrt{5}}{2}\right)^i + \left(\frac{\sqrt{5}+1}{2\sqrt{5}}\right)\left(\frac{1+\sqrt{5}}{2}\right)^i\right]x^i.\end{aligned}$$

We observe that the coefficients of the generating function thus agree with the expression we derived previously for the n^{th} term of the Fibonacci sequence.

General Form for the Generating Function of a Homogeneous Linear Difference Equation with Constant Coefficients

The arguments used in the previous section can be applied to a general homogeneous linear difference equation with constant coefficients of order k. Such an equation can be written in the form

$$u_n + a_1u_{n-1} + a_2u_{n-2} + \cdots + a_ku_{n-k} = 0.$$

To determine the generating function for this difference equation we observe that

$$\begin{array}{rlllll}
g(x) = & u_0 & +u_1x & +u_2x^2 & +u_3x^3 + \cdots & +u_kx^k + \cdots \\
a_1xg(x) = & & a_1u_0x & +a_1u_1x^2 & +a_1u_2x^3 + \cdots & +a_1u_{k-1}x^k + \cdots \\
a_2x^2g(x) = & & & a_2u_0x^2 & +a_2u_1x^3 + \cdots & +a_2u_{k-2}x^k + \cdots \\
\vdots & & & & & \vdots \\
a_kx^kg(x) = & & & & & a_ku_0x^k + \cdots
\end{array}$$

The sum of the left hand sides of these equations is $(1 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_kx^k)g(x)$ and the sum of the right hand sides is $u_0 + (u_1 + a_1u_0)x + (u_2 + a_1u_1 + a_2u_0)x^2 + \cdots + \left(u_{k-1} + \sum_{i=1}^{k-1} a_iu_{k-1-i}\right) x^{k-1}$. From this analysis we can deduce the following theorem:

Theorem 3.1. *The generating function for the sequence defined by the difference equation*

$$u_n + a_1u_{n-1} + a_2u_{n-2} + \cdots + a_ku_{n-k} = 0 \tag{3.14}$$

can be written in the form $\frac{p(x)}{q(x)}$, *where*

$$p(x) = u_0 + \sum_{r=1}^{k-1} \left(u_r + \sum_{i=1}^{r} a_iu_{r-i} \right) x^r,$$
$$q(x) = 1 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_kx^k.$$

In particular, p is a polynomial of degree less than k, and q is a polynomial of degree k whose constant term is nonzero.

Conversely, any generating function of this form gives rise to a sequence that can be described by a homogeneous linear difference equation with constant coefficients such as (3.14).

It remains to note the connection between the characteristic polynomial for a difference equation and the polynomial $q(x)$. If we let $r(x)$ be the characteristic polynomial of this difference equation, then $r(x) = x^k + a_1x^{k-1} + a_2x^{k-2} + \cdots + a_k$. So we see that the polynomial q is closely related to the characteristic polynomial.

Exercise 3.24. Show that $q(x) = x^k r\left(\frac{1}{x}\right)$.

Exercise 3.25. Use the generating function technique to solve the difference equation given in part 1 of Exercise 3.11.

3.2.5 Catalan Numbers

The *Catalan numbers* are a sequence of numbers that turn up as the solution to many different counting problems. In order to introduce these numbers, and see how they can be studied with the help of difference equations and generating functions, we consider the following problem:

> *How many ways are there of joining $2n$ points on a line by n nonintersecting arcs, so that each point lies on precisely one arc, and the arcs all lie above the line?*

For example, in the case of six points there are five ways to add three arcs, as illustrated in Figure 3.3. We denote the number of ways of adding n arcs in this manner to $2n$ points by C_n.

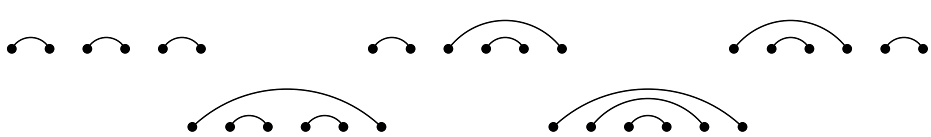


Figure 3.3: The five ways of adding three nonintersecting arcs to sets of six points on a line

We will now derive a difference equation describing the sequence $(C_n)_{n=0}^{\infty}$. In order to do this we make the following observations:

1. The arc that contains the leftmost point passes over $2k$ points for some k with $0 \leq k \leq n-1$. For if an odd number of points lay under this arc, then it would be necessary to connect some point under the arc to a point to the right of the arc, which would result in the corresponding arcs crossing.

2. If there are $2k$ points lying under the arc containing the leftmost point then the number of ways of joining these points with arcs is C_k.

3. If there are $2k$ points lying under the arc containing the leftmost point then there are $2n-2k-2$ points that occur to the right of this arc. The number of ways of joining these points with arcs is C_{n-k-1}.

Combining these equations, we observe that the C_n satisfy the difference equation

$$C_n = \sum_{k=0}^{n-1} C_k C_{n-k-1}, \qquad C_0 = 1. \tag{3.15}$$

The number of ways of assigning arcs in which the arc containing the leftmost point passes over k points is equal to the number of ways of allocating arcs to the points under the arc multiplied by the number of ways of allocating arcs to the points to the right of this arc, by the rule of product. To count the total number of possibilities we then simply sum over all possible values of k. This difference equation has a quite different form to the ones we have studied previously; in particular, it is not linear. However, we can solve it with the use of generating function techniques.

Let $g(x) = \sum_{r=0}^{\infty} C_r x^r$ be the generating function corresponding to this problem. Recall that in Chapter 2 we saw that the power series $\sum_{r=0}^{\infty} a_r x^r$ and $\sum_{r=0}^{\infty} b_r x^r$ could be multiplied to obtain a new power series in which the coefficient of x^r is $\sum_{i=0}^{r} a_i b_{r-i}$. Comparing this expression with the righthand side of (3.15), we observe that $\sum_{k=0}^{n-1} C_k C_{n-k-1}$ is in fact the coefficient of x^{n-1} in the power series $g(x)^2$. From this we deduce that for $n \geq 1$, the sum $\sum_{k=0}^{n-1} C_k C_{n-k-1}$ gives the coefficient of x^n in the power series $xg(x)^2$. Finally, we observe that the coefficient of x^0 in the power series $xg(x)^2$ is 0, whereas the coefficient of x^0 in $g(x)$ is 1. From this we conclude that

$$g(x) = 1 + xg(x)^2$$

and hence, by the quadratic formula,

$$g(x) = \frac{1 - \sqrt{1-4x}}{2x}.$$

(Note that in the quadratic formula we take the solution that has a negative sign in order to avoid having a term of the form $\frac{1}{x}$ in the corresponding power series.)

Thus we have found a generating function that describes the solutions to our problem. In order to find the coefficients of this generating function, it is helpful to recall that the Maclaurin series for $\sqrt{1+x}$ is given by

$$\sqrt{1+x} = \sum_{r=0}^{\infty} \frac{(-1)^r (2r)!}{(1-2r)(r!)^2 4^r} x^r = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 + \cdots \qquad |x| < 1.$$

Hence we have that

$$\begin{aligned}
g(x) &= \frac{1}{2x}\left[1 - \sum_{r=0}^{\infty} \frac{(-1)^r (2r)!}{(1-2r)(r!)^2 4^r}(-4x)^r\right] \\
&= \frac{1}{2x}\left[\sum_{r=1}^{\infty} \frac{(-1)^{r+1} (2r)!}{(1-2r)(r!)^2 4^r}(-4x)^r\right] \\
&= \frac{1}{2x}\left[\sum_{r=1}^{\infty} \frac{(-1)(2r)!}{(1-2r)(r!)^2}x^r\right] \\
&= \sum_{r=1}^{\infty} \frac{(-1)(2r)!}{2(1-2r)(r!)^2}x^{r-1} \\
&= \sum_{r=0}^{\infty} \frac{(-1)(2r+2)!}{2(-1-2r)((r+1)!)^2}x^{r} \\
&= \sum_{r=0}^{\infty} \frac{(2r)!(2r+2)(2r+1)}{2(1+2r)(r!)^2(r+1)^2}x^{r} \\
&= \sum_{r=0}^{\infty} \frac{(2r)!}{(r!)^2(r+1)}x^{r} \\
&= \sum_{r=0}^{\infty} \frac{1}{r+1}\binom{2r}{r}x^{r}.
\end{aligned}$$

And so we conclude (at last!) that the solution to our problem is

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

The numbers of the sequence $(C_n)_{n=0}^{\infty}$ are known as the *Catalan numbers*. The first few terms are given by

$$1, 1, 2, 5, 14, 42, 139, 429, \ldots$$

They occur in many different situations; Volume 2 of *Enumerative Combinatorics* by Richard P. Stanley (Cambridge University Press 1999, 511.62 STA) contains an exercise listing 66 different counting problems whose solution is given by the Catalan numbers! Here we give just a couple of examples.

Example 3.26. We say that a string of parentheses (*i.e.* round brackets) is *legal* if, when read from left to right, at any point the number of right parentheses seen so far is never greater than the number of left parentheses already seen. So (()(())) is a legal string, but ())(() is not. Legal strings represent valid ways in which parentheses could be arranged within a mathematical expression; for each left parenthesis in a legal string there is a unique corresponding right parenthesis. How many ways are there of arranging n pairs of parentheses into a legal string?
Solution. In fact this problem is equivalent to the one previously studied: if the dot on the leftmost side of any arc is replaced by a left parenthesis and the corresponding dot on the righthand side is replaced by a right parenthesis then the sequence of $2n$ dots is replaced by

a legal string of parentheses. This is because each arc corresponds to a pair of parentheses, and you never see the right parenthesis in a pair without having previously seen the left one, hence the resulting string must be legal. Conversely, given a legal string of parentheses, we can start with the innermost pairs of parentheses and work outwards, associating each pair of parentheses with an arc, and thus obtain a bijection between the ways of joining points with arcs and the legal strings of parentheses (see Figure 3.4, for example). Thus we conclude that

Figure 3.4: Converting between a set of arcs and a string of parentheses

the solution to this problem is given by the n^{th} Catalan number $\frac{1}{n+1}\binom{2n}{n}$.

Example 3.27. Given $2n$ points on the circumference of a circle, how many ways are there of connecting pairs of points by n chords that do not cross, such that each point lies on precisely one chord?
Solution. This is another problem that is equivalent to the first problem we considered: by "unrolling" the circle into a straight line we can convert a set of chords for this problem into a set of arcs for the previous problem, and *vice versa*. So again, we conclude that the solution is

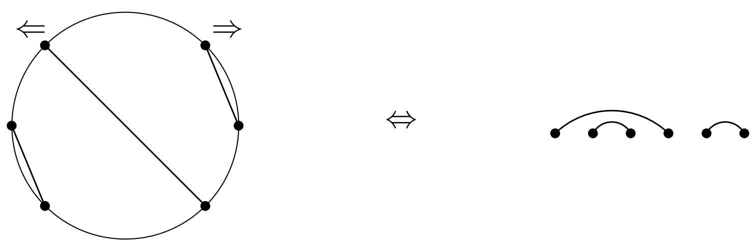


Figure 3.5: Converting between chords of a circle and arcs connecting points on a line

the n^{th} Catalan number $\frac{1}{n+1}\binom{2n}{n}$.

Exercise 3.28. Use the formula for the n^{th} Catalan number to show that $C_n = \frac{2\cdot 6\cdot 10\cdots 2(2n-1)}{(n+1)!}$, and hence deduce that the Catalan numbers satisfy the difference equation

$$(n+1)C_n = (4n-2)C_{n-1}.$$

Exercise 3.29. Show that C_n can be written in the form $C_n = \binom{2n}{n} - \binom{2n}{n-1}$.

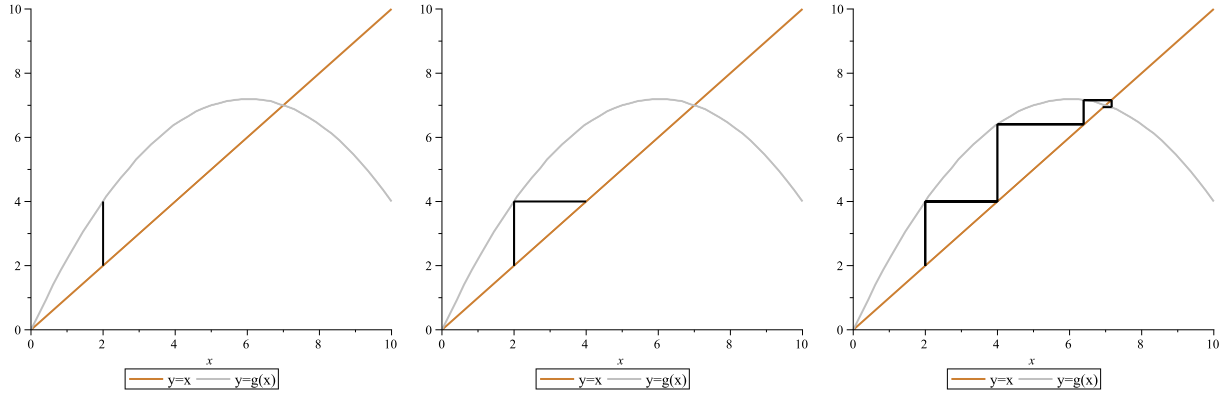


Figure 3.6: Graphs showing steps 1 and 2 in the construction of a cobweb diagram for the function $y = g(x)$ with initial value $u_0 = 2$, as well as the appearance of the cobweb diagram after 4 iterations of this process

3.3 Applications of Difference Equations

Now that we have seen some techniques for working with difference equations, we are going to explore how they arise in models of real world systems. This section draws on material from *Modelling with Differential and Difference Equations* by Fulford, Forrester and Jones (Cambridge University Press 1997, 511.8 FUL).

Firstly we will introduce a method of visualising the behaviour of the solutions to certain first order difference equations.

Cobweb Diagrams

A *cobweb diagram* is a graphical method of displaying the solution to a first order difference equation of the form $u_n = g(u_{n-1})$. To construct a cobweb diagram, we first construct plots of $y = g(x)$ and $y = x$ on the same axes, then we display the solution to the difference equation as follows:

1. The initial value is shown by means of the vertical line of equation $x = u_0$ that is drawn until it meets the graph of $y = g(x)$. The point at the top of this line has coordinates $(u_0, g(u_0))$, which is equal to (u_0, u_1).

2. We draw a horizontal line from this point (u_0, u_1) until it meets the graph of $y = x$ in the point (u_1, u_1).

3. We then draw a vertical line from this point that meets the graph $y = g(x)$ in the point (u_1, u_2).

4. We continue repeating these steps, with the i^{th} iteration giving rise to a vertical line that intersect the graph $y = g(x)$ in the point (u_{i-1}, u_i) and a horizontal line meeting the line $y = x$ in the point (u_i, u_i).

An example of this process is illustrated in Figure 3.6. The x values of the vertical lines in a cobweb diagram represent the values of u_i for $i = 0, 1, 2, \ldots$, and so these diagrams can provide an informative illustration of how the sequence $(u_i)_{i=0}^{\infty}$ behaves as n increases.

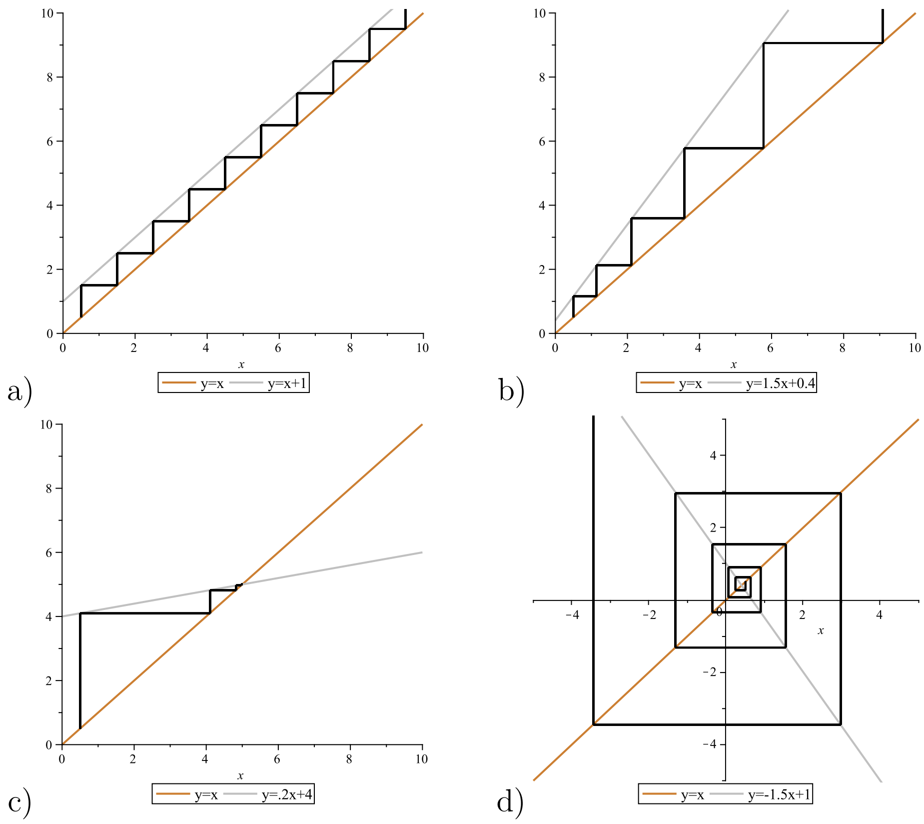


Figure 3.7: Cobweb diagrams for difference equations of the form $u_n = au_{n-1} + b$.

Example 3.30. Consider the linear difference equation $u_n = au_{n-1} + b$, with $u_0 = U$. We have seen in Chapter 1 that this difference equation has solution

$$u_n = a^n U + \frac{b(a^n - 1)}{a - 1}$$

unless $a = 1$, in which case it is simply an arithmetic progression, and it has the solution

$$u_n = U + nb.$$

That is, when $a \neq 1$ the sequence $(u_i)_{i=0}^{\infty}$ grows exponentially with i, whereas when $a = 1$ the growth is linear. We can observe this behaviour by studying the corresponding cobweb diagrams, as shown in Figure 3.7. Diagram a) shows an example with $a = 1$, and we can see the controlled way in which the values of u_i increases, in contrast to the rapidly increasing rate of growth in diagram b), for which $a = 1.5$. Diagram c) illustrates the case where $a = .5$; in this case we have $0 < a < 1$, and so the sequence $(u_i)_{i=0}^{\infty}$ converges. Finally, diagram d) has $a = -1.5$, and so we see the value of u_i varies between negative and positive values, but $|u_i|$ increases exponentially, as illustrated by the ever-increasing spiral of the diagram.

Example 3.31. Consider the difference equation $u_n = -u_{n-1} + 10$, with $u_0 = 7$. By examining the cobweb diagram for this difference equation (Figure 3.8), we see that the corresponding sequence is periodic: the value of u_i is 7 for even i, and 3 for odd i, so the sequence has period 2.

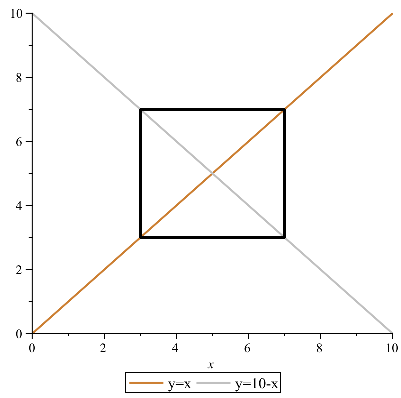


Figure 3.8: A cobweb diagram corresponding to a sequence with period 2

Thus we see that cobweb diagrams can provide a useful means of visualising the different behaviours that even a linear first order difference equation can produce. We will make further use of these diagrams later in this chapter.

3.3.1 Modelling Simple Problems in Finance and Economics Using Difference Equations

Difference equations arise in the modelling of systems for which the behaviour in a particular time period depends on the conditions in previous time periods. There are many instances of problems in finance and economics for which this is the case.

Interest and Loan Repayments

When money is placed in an account with a bank or other financial institution in order to earn interest, the amount in the account at any given time is equal to the amount that was first placed in the account (referred to as the *principal*), plus the interest earned on that amount (assuming no additional money has been added to or withdrawn from the account during the interim). There are two main ways in which interest may be paid:

simple interest The amount of interest earned during a given time period is a fixed percentage p of the principal A. Denoting by A_t the amount of money in the account t time periods after it is opened, we observe that the amount of money in the account can be described by the following difference equation:

$$A_t = A_{t-1} + \frac{p}{100}A.$$

The sequence arising from this difference equation is simply an arithmetic progression, and hence we obtain the following solution to the difference equation.

$$\begin{aligned} A_t &= A + \frac{p}{100}tA, \\ &= A\left(1 + \frac{p}{100}t\right). \end{aligned}$$

The amount of money in the account grows linearly with time; this is analogous to the scenario illustrated in graph a) of Figure 3.7.

compound interest The amount of interest earned during a given period is a fixed percentage p of the amount of money in the account during that time period. In this case we have the following difference equation:

$$\begin{aligned} A_t &= A_{t-1} + \frac{p}{100}A_{t-1}, \\ &= A_{t-1}\left(1+\frac{p}{100}\right). \end{aligned}$$

This difference equation describes a geometric progression, and hence has the solution

$$A_t = A\left(1+\frac{p}{100}\right)^t.$$

Here the amount in the account grows exponentially with time, as in case b) of Figure 3.7. (Note that we are in case b) rather than case c) as the interest p is nonnegative.)

One point to note: the interest paid on an account is usually described as the interest *per annum*, which is the amount of interest you would earn in a year if it were paid as simple interest. However, in practice the interest may actually be paid at more frequent intervals, *e.g.* monthly or quarterly. If the *per annum* interest rate is $p\%$, and interest is compounded k times per year, then the actual amount of interest paid after each time period is $\frac{p}{k}\%$. So if the interest is compounded monthly, then each month you will be paid $\frac{p}{12}\%$ of interest, and if it is compounded quarterly then each quarter you will be paid $\frac{p}{4}\%$.

If you are borrowing money rather than investing it then you will have to pay interest on the loan. Here we consider one common model that is used to schedule the repayment of a loan.

amortisation Payments on a loan are said to be *amortised* when a fixed amount is paid at the end of each time period in order to pay off both the amount owed, and the interest charged on that amount. At the end of each time period, interest is charged on the amount that is still owed.

If we suppose the borrower pays off the loan in installments of size R, and the interest charged is p percent per time period, then the amount that is still owed after t time periods satisfies the difference equation

$$\begin{aligned} A_t &= A_{t-1} + \frac{p}{100}A_{t-1} - R, \\ &= A_{t-1}\left(1+\frac{p}{100}\right) - R. \end{aligned}$$

This is a first order linear difference equation with constant coefficients; it has the solution

$$\begin{aligned} A_t &= A\left(1+\frac{p}{100}\right)^t - R\frac{\left(1+\frac{p}{100}\right)^t - 1}{\left(1+\frac{p}{100}\right) - 1}, \\ &= A\left(1+\frac{p}{100}\right)^t - R\frac{100}{p}\left(1+\frac{p}{100}\right)^t + R\frac{100}{p}. \end{aligned}$$

We observe that if the value of R is chosen such that $R = \frac{p}{100}A$ then the amount of interest charged during each time period is equal to amount paid off, and so the amount owed never actually decreases. If $R < \frac{p}{100}A$ then the interest charged is greater than the amount paid off, and so the debt will in fact increase.

Exercise 3.32.

1. Suppose you wish to invest £1000 for the period of two years. Is it better to choose an account that pays 11% simple interest *per annum* or 10% compound interest *per annum* if the interest is compounded monthly?

2. If I place a certain amount of money in an account, how long will it take for my money to double if the account pays 6% interest that is compounded annually? What about if the interest is compounded monthly?

3. Suppose that I borrow £50 000 at an interest rate of 6% *per annum*. How much do I need to pay back each month in order to pay off the loan in 20 years?

Supply and Demand

The basic principle of supply and demand is that demand for a product decreases as price increases, the amount that producers are willing to supply increases as the price increases, and the market price for the product will be the equilibrium price for which the quantity supplied equals the quantity demanded. However, for certain products that take time to produce, such as many forms of agricultural produce, there is a limit to how quickly suppliers can react to a change in price. For example, the decision of how much wheat to plant must be made without access to the knowledge of what the price of wheat will be at harvest time.

Here we consider a simple model for investigating the effect that such a time lag has on the the way the price of a good varies with time. This model is known as the "cobweb model" of supply and demand, and is based on the following assumptions:

- The quantity that producers are willing to supply during time period t is a linear function of the price in the previous time period $t-1$, and supply increases as price increases.

- The quantity that consumers are willing to purchase during time period t is a linear function of the price during that time period, and the demand decreases as price increases.

- The market price at time t is determined by the price that consumers are willing to pay for the quantity that is supplied at that time.

Thus, if we use the notation s_t to represent the quantity supplied at time t and d_t to represent the quantity demanded, and p_t to represent the price, then we have

$$\begin{aligned} s_t &= ap_{t-1} + b, \\ d_t &= -cp_t + d, \end{aligned}$$

where a, b, c, d are real constants with $a, c > 0$. The price at time t is determined by the amount that consumers are willing to pay given the quantity that is being supplied, and is hence obtained by equating s_t and d_t. This leads to the following difference equation for p_t:

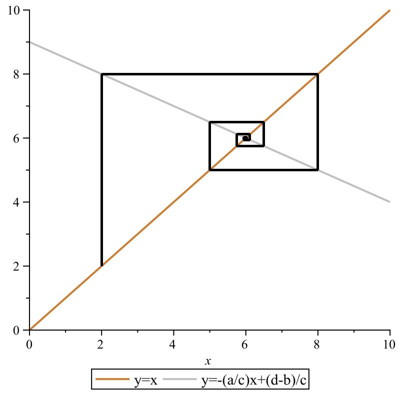
$$\begin{aligned} s_t &= d_t \\ ap_{t-1} + b &= -cp_t + d, \\ p_t &= -\frac{a}{c}p_{t-1} + \frac{d-b}{c} \end{aligned}$$

This is another example of a linear first order difference equation.

Exercise 3.33. Show that if at time t_0 the price is equal to $\frac{d-b}{c+a}$ then the price will remain constant over time.

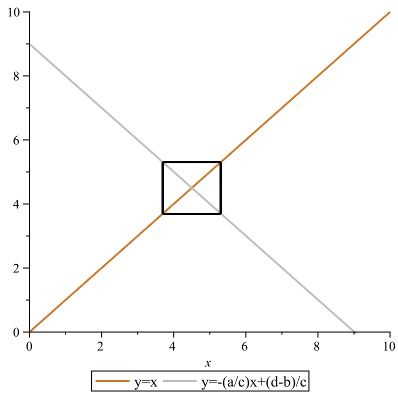
If the price at time t_0 is something other than $\frac{d-b}{c+a}$, then the behaviour of the system depends on whether $|\frac{a}{c}|$ is greater than, less than, or equal to 1.

$|\mathbf{\frac{a}{c}}| < \mathbf{1}$



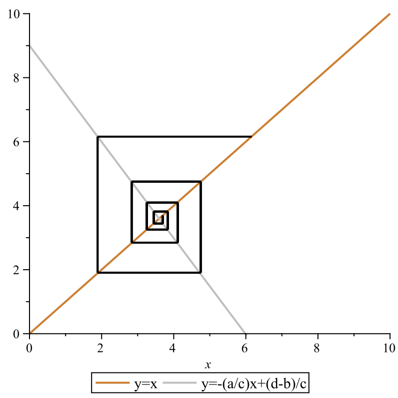
In this case the cobweb diagram for the difference equation spirals inwards, and the price eventually converges to $\frac{d-b}{c+a}$. This can therefore be regarded as a stable fixed point.

$|\mathbf{\frac{a}{c}}| = \mathbf{1}$



In this case, we find that the price fluctuates periodically between two values.

$|\mathbf{\frac{a}{c}}| > \mathbf{1}$



Finally, we observe that when $a > c$ the cobweb diagram spirals outwards, and the price diverges away from the fixed point.

3.3.2 Population Models

Here we consider some possible models for the number of individuals in a given population. While these models could be regarded as somewhat oversimplistic, omitting various factors that could affect the size of a population in the real world, nevertheless they serve to illustrate trends that have been observed in actual population studies.

Linear Models

In Example 3.1, we have already seen an instance of a simple population model in which the number of bacteria present in a test tube after a given time period is given by a linear function of the number of bacteria present after the previous time period. This model is slightly limited, in that it does not take account of the death of members of the population. We can construct a simple model that does, by making the following assumptions (using N_t to denote the size of the population at the start of time interval t):

- There is a fixed *per capita* birth rate rate α, so that the number of individuals born during time period t is equal to αN_t.
- There is a fixed *per capita* death rate β, so that the number of individuals that die during time period t is βN_t.
- No individuals join or leave the population at any time (other than through being born or dying).

These assumptions imply that the population at the start of the $t+1^{th}$ time interval can be described by the difference equation

$$\begin{aligned} N_{t+1} &= N_t - \beta N_t + \alpha N_t, \\ &= (1 + \alpha - \beta) N_k. \end{aligned}$$

Exercise 3.34.

1. What type of sequence is produced by this difference equation?
2. What happens to the population if $\alpha > \beta$?
3. What happens if $\alpha = \beta$?
4. What happens if $\alpha < \beta$?
5. Is this a realistic model?

Exercise 3.35. The half life of polonium-218 is approximately 3 minutes. This means that a given mass of polonium-218 will decay to leave half that amount remaining after 3 minutes. Suppose I place 10g of polonium-218 in a box, and then every 3 minutes I add another gram of polonium-218 to the box. Draw a cobweb diagram that illustrates how the amount of polonium-218 in the box varies with time, and describe what happens to the amount in the long term.

The Discrete Logistic Equation

We have seen that the models we have been studying give rise to populations that continue to grow indefinitely in the case where the birth rate exceeds the death rate. This seems unrealistic in practice, and suggests that our assumption of a constant birth rate and death rate does not accurately reflect that which occurs in nature. What this model neglects is the impact of environmental factors on the size of the population: the limited resources available within a particular environment will necessarily limit the size of the population that can be supported.

Instead, it seems reasonable to assume that as the population size increases, the birth rate will decrease and the death rate will increase, due to increased competition for available resources. For a given population, we define the *carrying capacity* C of its environment to be the population size for which the birth rate is equal to the death rate. This represents the largest stable population that the environment is capable of sustaining. If the population attains size C then the overall growth will be 0; if the population is larger then the death rate will exceed the birth rate and the population size will decrease in consequence, whereas if the population is less than C the rate of births will be higher than the rate of deaths, and the population will increase.

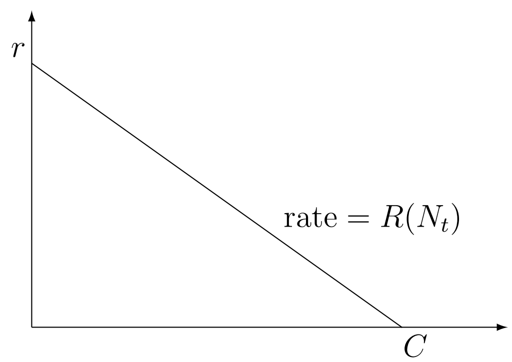
We can construct a model taking these concepts into account by means of a difference equation of the form

$$N_{t+1} = N_t + R(N_t)N_t, \tag{3.16}$$

where $R(N_t)$ represents the overall growth rate of the population, as a function of the population size. We would expect this function to satisfy the following properties:

- $R(N_t)$ decreases as N_t increases
- $R(C) = 0$
- If the population size is small relative to the carrying capacity then the environment places little constraint on the growth of the population. If we let r denote the rate at which the population would grow in the absence of environmental constraints, then we would expect $R(0) = r$. We refer to r as the *unconstrained growth rate.*

There are many possible different models with these properties, but we will consider the simple case in which $R(N_t)$ is a linear function. This is known as the *discrete logistic model*



The requirements that $R(0) = r$ and $R(C) = 0$ imply that R satisfies

$$R(N_t) = -\frac{r}{C}N_t + r. \tag{3.17}$$

Together with (3.16), this implies our model is described by the difference equation

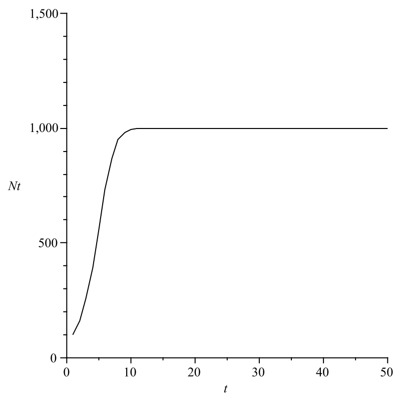
$$N_{t+1} = N_t + rN_t\left(1 - \frac{N_t}{C}\right). \tag{3.18}$$

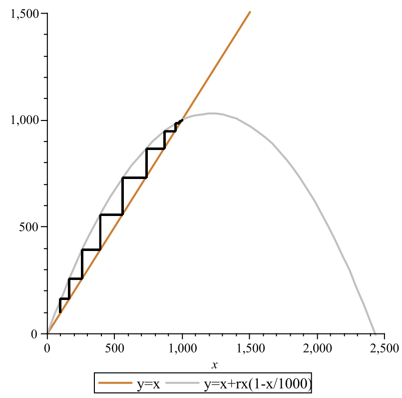
This difference equation is nonlinear, and none of the techniques we have studied can be used to solve it. In fact there is no known way to express the solution to this equation in closed form! However, it does give rise to a range of interesting behaviours that have been observed in genuine population studies. We will investigate some of its properties through the use of cobweb diagrams.

For our investigations we will suppose we have a carrying capacity of $C = 1000$, and an initial population size of 100. We will then consider the behaviour of our model for different values of the unconstrained growth rate r -this can be seen as a measure of the average fertility of the individuals in the population. We note that one drawback to the discrete logistic model is that for $r \geq 3$ it can give rise to negative population sizes, which is clearly unrealistic! Hence we will constrain our attention to cases with $0 < r < 3$.

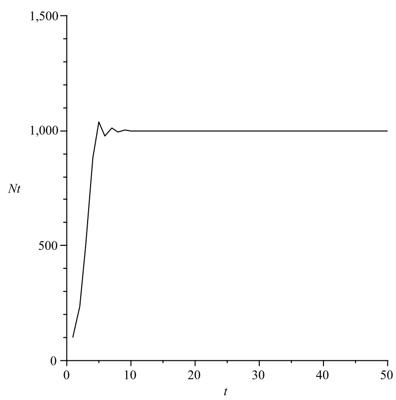
We note that if the population has size 0 or C then it remains constant; these are the only two constant solutions to this difference equation in general.

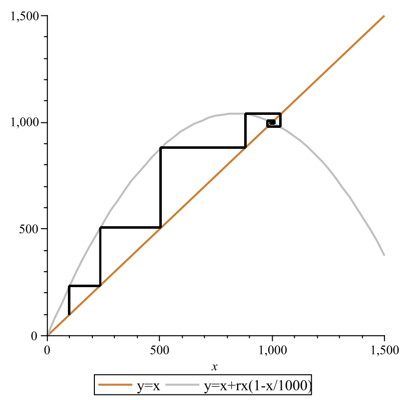
$\mathbf{0 < r \leq 2}$ (stable growth) Here we plot how the population size evolves with time, as well as giving the corresponding cobweb diagram.



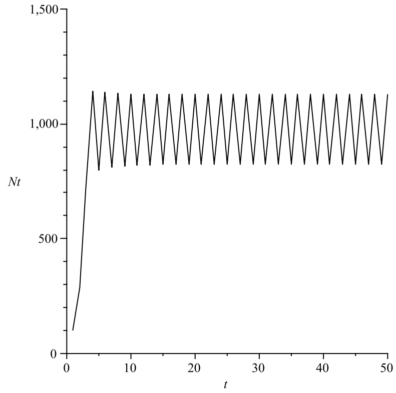


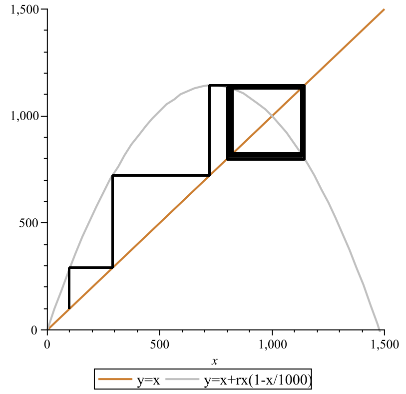
We see that for $r = 0.7$ the population grows steadily until the carrying capacity is reached. This is typical for values of r in the range $0 < r \leq 1$. For $1 < r \leq 2$ there is a damped oscillation of the population that eventually converges to the carrying capacity.



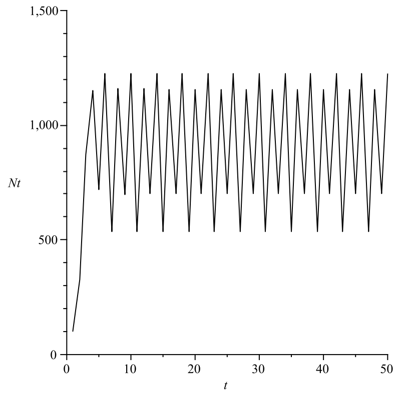


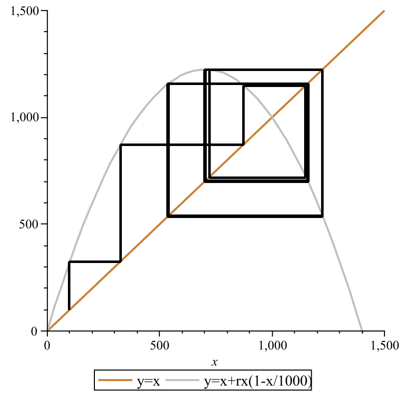
2 < r < 2.57 (periodic behaviour) As the value of r increases past 2, we find that the behaviour of our system changes to give a periodic solution. For $r = 2.1$ the resulting solution has period 2:





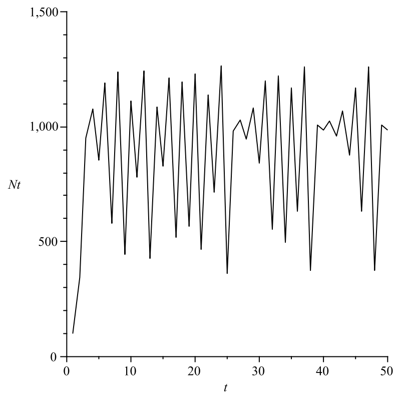
As r increases to around 2.5 the resulting solution switches to a period of length 4:

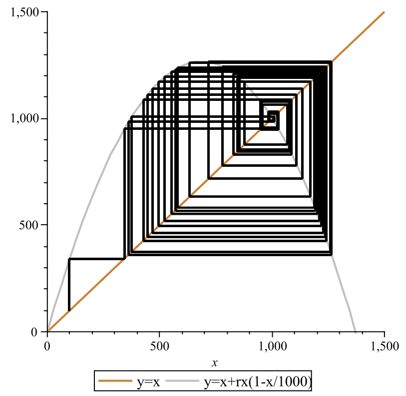




This is known as *period doubling*, and as r continues to increase we can find values for which the solution has period $8, 16, 32, \ldots$.

r > 2.57 (chaotic behaviour) Once r becomes greater than about 2.57 the population size appears to fluctuate in an unpredictable manner.





This type of seemingly random behaviour arising from an entirely deterministic system is said to be *chaotic*. (Those who would like to learn more about chaos theory might be interested in reading either *Does God Play Dice? : the New Mathematics of Chaos* by Ian Stewart, Penguin 1997, 530.1 STE, or *Chaos: Making A New Science* by James Gleick, Vintage 1987.) It has many unexpected and interesting features. For example, after a certain point the behaviour changes from being unpredictable to suddenly being regular again, only with period 3, and then we again observe period doubling, with increasing r leading to sequences of period $6, 12, \ldots$ before chaotic behaviour is again resumed. I will place a copy of the maple worksheet used to generate these examples on the course website so that you can investigate some of this behaviour for yourselves.

Learning Outcomes

After completing this chapter and the related problems you should be able to:

- determine the order of a difference equation, and state whether it is homogeneous, whether it is linear, and whether it is linear with constant coefficients;
- solve a homogeneous first order linear difference equation;
- solve an inhomogeneous first order linear difference equation;
- solve homogeneous second order linear difference equations with constant coefficients;
- solve certain cases of inhomogeneous second order linear difference equations with constant coefficients;
- use induction to prove that a conjectured solution to a difference equation is correct;
- recognise that a substitution may be used to convert a difference equation into a form in which it may be solved;
- find the generating function associated with a linear homogeneous difference equation with constant coefficients;
- make use of partial fractions to determine the coefficients of the generating function associated with a linear homogeneous difference equation with constant coefficients;
- write down a difference equation corresponding to a given counting problem;
- write down a difference equation to model the behaviour of a real-world system;
- draw a cobweb diagram to illustrate the behaviour of the solution to a difference equation of the form $u_n = g(u_{n-1})$;
- given a solution presented in the form of a cobweb diagram, determine whether it converges, and whether it is periodic.