

4

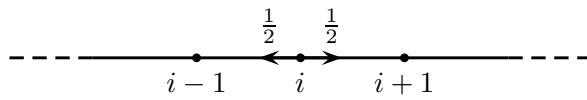
Further theory

In the first three chapters we have given an account of the elementary theory of Markov chains. This already covers a great many applications, but is just the beginning of the theory of Markov processes. The further theory inevitably involves more sophisticated techniques which, although having their own interest, can obscure the overall structure. On the other hand, the overall structure is, to a large extent, already present in the elementary theory. We therefore thought it worth while to discuss some features of the further theory in the context of simple Markov chains, namely, martingales, potential theory, electrical networks and Brownian motion. The idea is that the Markov chain case serves as a guiding metaphor for more complicated processes. So the reader familiar with Markov chains may find this chapter helpful alongside more general higher-level texts. At the same time, further insight is gained into Markov chains themselves.

4.1 Martingales

A martingale is a process whose average value remains constant in a particular strong sense, which we shall make precise shortly. This is a sort of balancing property. Often, the identification of martingales is a crucial step in understanding the evolution of a stochastic process.

We begin with a simple example. Consider the simple symmetric random walk $(X_n)_{n \geq 0}$ on \mathbb{Z} , which is a Markov chain with the following diagram



The average value of the walk is constant; indeed it has the stronger property that the average value of the walk at some future time is always simply the current value. In precise terms we have

$$\mathbb{E}X_n = \mathbb{E}X_0;$$

and the stronger property says that, for $n \geq m$,

$$\mathbb{E}(X_n - X_m \mid X_0 = i_0, \dots, X_m = i_m) = 0.$$

This stronger property says that $(X_n)_{n \geq 0}$ is in fact a martingale.

Here is the general definition. Let us fix for definiteness a Markov chain $(X_n)_{n \geq 0}$ and write \mathcal{F}_n for the collection of all sets depending only on X_0, \dots, X_n . The sequence $(\mathcal{F}_n)_{n \geq 0}$ is called the *filtration* of $(X_n)_{n \geq 0}$ and we think of \mathcal{F}_n as representing the state of knowledge, or history, of the chain up to time n . A process $(M_n)_{n \geq 0}$ is called *adapted* if M_n depends only on X_0, \dots, X_n . A process $(M_n)_{n \geq 0}$ is called *integrable* if $\mathbb{E}|M_n| < \infty$ for all n . An adapted integrable process $(M_n)_{n \geq 0}$ is called a *martingale* if

$$\mathbb{E}[(M_{n+1} - M_n)1_A] = 0$$

for all $A \in \mathcal{F}_n$ and all n . Since the collection \mathcal{F}_n consists of countable unions of elementary events such as

$$\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\},$$

this martingale property is equivalent to saying that

$$\mathbb{E}(M_{n+1} - M_n \mid X_0 = i_0, \dots, X_n = i_n) = 0$$

for all i_0, \dots, i_n and all n .

A third formulation of the martingale property involves another notion of conditional expectation. Given an integrable random variable Y , we define

$$\mathbb{E}(Y \mid \mathcal{F}_n) = \sum_{i_0, \dots, i_n} \mathbb{E}(Y \mid X_0 = i_0, \dots, X_n = i_n)1_{\{X_0=i_0, \dots, X_n=i_n\}}.$$

The random variable $\mathbb{E}(Y \mid \mathcal{F}_n)$ is called the *conditional expectation* of Y given \mathcal{F}_n . In passing from Y to $\mathbb{E}(Y \mid \mathcal{F}_n)$, what we do is to replace on each elementary event $A \in \mathcal{F}_n$, the random variable Y by its average value $\mathbb{E}(Y \mid A)$. It is easy to check that an adapted integrable process $(M_n)_{n \geq 0}$ is a martingale if and only if

$$\mathbb{E}(M_{n+1} \mid \mathcal{F}_n) = M_n \quad \text{for all } n.$$

Conditional expectation is a partial averaging, so if we complete the process and average the conditional expectation we should get the full expectation

$$\mathbb{E}(\mathbb{E}(Y \mid \mathcal{F}_n)) = \mathbb{E}(Y).$$

It is easy to check that this formula holds.

In particular, for a martingale

$$\mathbb{E}(M_n) = \mathbb{E}(\mathbb{E}(M_{n+1} \mid \mathcal{F}_n)) = \mathbb{E}(M_{n+1})$$

so, by induction

$$\mathbb{E}(M_n) = \mathbb{E}(M_0).$$

This was already clear on taking $A = \Omega$ in our original definition of a martingale.

We shall prove one general result about martingales, then see how it explains some things we know about the simple symmetric random walk. Recall that a random variable

$$T : \Omega \rightarrow \{0, 1, 2, \dots\} \cup \{\infty\}$$

is a *stopping time* if $\{T = n\} \in \mathcal{F}_n$ for all $n < \infty$. An equivalent condition is that $\{T \leq n\} \in \mathcal{F}_n$ for all $n < \infty$. Recall from Section 1.4 that all sorts of hitting times are stopping times.

Theorem 4.1.1 (Optional stopping theorem). *Let $(M_n)_{n \geq 0}$ be a martingale and let T be a stopping time. Suppose that at least one of the following conditions holds:*

- (i) $T \leq n$ for some n ;
- (ii) $T < \infty$ and $|M_n| \leq C$ whenever $n \leq T$.

Then $\mathbb{E}M_T = \mathbb{E}M_0$.

Proof. Assume that (i) holds. Then

$$\begin{aligned} M_T - M_0 &= (M_T - M_{T-1}) + \dots + (M_1 - M_0) \\ &= \sum_{k=0}^{n-1} (M_{k+1} - M_k) 1_{k < T}. \end{aligned}$$

Now $\{k < T\} = \{T \leq k\}^c \in \mathcal{F}_k$ since T is a stopping time, and so

$$\mathbb{E}[(M_{k+1} - M_k)1_{k < T}] = 0$$

since $(M_k)_{k \geq 0}$ is a martingale. Hence

$$\mathbb{E}M_T - \mathbb{E}M_0 = \sum_{k=0}^{n-1} \mathbb{E}[(M_{k+1} - M_k)1_{k < T}] = 0.$$

If we do not assume (i) but (ii), then the preceding argument applies to the stopping time $T \wedge n$, so that $\mathbb{E}M_{T \wedge n} = \mathbb{E}M_0$. Then

$$|\mathbb{E}M_T - \mathbb{E}M_0| = |\mathbb{E}M_T - \mathbb{E}M_{T \wedge n}| \leq \mathbb{E}|M_T - M_{T \wedge n}| \leq 2C\mathbb{P}(T > n)$$

for all n . But $\mathbb{P}(T > n) \rightarrow 0$ as $n \rightarrow \infty$, so $\mathbb{E}M_T = \mathbb{E}M_0$. \square

Returning to the simple symmetric random walk $(X_n)_{n \geq 0}$, suppose that $X_0 = 0$ and we take

$$T = \inf\{n \geq 0 : X_n = -a \text{ or } X_n = b\}$$

where $a, b \in \mathbb{N}$ are given. Then T is a stopping time and $T < \infty$ by recurrence of finite closed classes. Thus condition (ii) of the optional stopping theorem applies with $M_n = X_n$ and $C = a \vee b$. We deduce that $\mathbb{E}X_T = \mathbb{E}X_0 = 0$. So what? Well, now we can compute

$$p = \mathbb{P}(X_n \text{ hits } -a \text{ before } b).$$

We have $X_T = -a$ with probability p and $X_T = b$ with probability $1 - p$, so

$$0 = \mathbb{E}X_T = p(-a) + (1 - p)b$$

giving

$$p = b/(a + b).$$

There is an entirely different, Markovian, way to compute p , using the methods of Section 1.4. But the intuition behind the result $\mathbb{E}X_T = 0$ is very clear: a gambler, playing a fair game, leaves the casino once losses reach a or winnings reach b , whichever is sooner; since the game is fair, the average gain should be zero.

We discussed in Section 1.3 the counter-intuitive case of a gambler who keeps on playing a fair game against an infinitely rich casino, with the certain outcome of ruin. This game ends at the finite stopping time

$$T = \inf\{n \geq 0 : X_n = -a\}$$

where a is the gambler's initial fortune. Since $X_T = -a$ we have

$$\mathbb{E}X_T = -a \neq 0 = \mathbb{E}X_0$$

but this does not contradict the optional stopping theorem because neither condition (i) nor condition (ii) is satisfied. Thus, while intuition might suggest that $\mathbb{E}X_T = \mathbb{E}X_0$ is rather obvious, some care is needed as it is not always true.

The example just discussed was rather special in that the chain $(X_n)_{n \geq 0}$ itself was a martingale. Obviously, this is not true in general; indeed a martingale is necessarily real-valued and we do not in general insist that the state-space I is contained in \mathbb{R} . Nevertheless, to every Markov chain is associated a whole collection of martingales, and these martingales characterize the chain. This is the basis of a deep connection between martingales and Markov chains.

We recall that, given a function $f : I \rightarrow \mathbb{R}$ and a Markov chain $(X_n)_{n \geq 0}$ with transition matrix P , we have

$$(P^n f)(i) = \sum_{j \in I} p_{ij}^{(n)} f_j = \mathbb{E}_i(f(X_n)).$$

Theorem 4.1.2. Let $(X_n)_{n \geq 0}$ be a random process with values in I and let P be a stochastic matrix. Write $(\mathcal{F}_n)_{n \geq 0}$ for the filtration of $(X_n)_{n \geq 0}$. Then the following are equivalent:

- (i) $(X_n)_{n \geq 0}$ is a Markov chain with transition matrix P ;
- (ii) for all bounded functions $f : I \rightarrow \mathbb{R}$, the following process is a martingale:

$$M_n^f = f(X_n) - f(X_0) - \sum_{m=0}^{n-1} (P - I)f(X_m).$$

Proof. Suppose (i) holds. Let f be a bounded function. Then

$$|(Pf)(i)| = \left| \sum_{j \in I} p_{ij} f_j \right| \leq \sup_j |f_j|$$

so

$$|M_n^f| \leq 2(n+1) \sup_j |f_j| < \infty$$

showing that M_n^f is integrable for all n .

Let $A = \{X_0 = i_0, \dots, X_n = i_n\}$. By the Markov property

$$\mathbb{E}(f(X_{n+1}) \mid A) = \mathbb{E}_{i_n}(f(X_1)) = (Pf)(i_n)$$

so

$$\mathbb{E}(M_{n+1}^f - M_n^f \mid A) = \mathbb{E}[f(X_{n+1}) - (Pf)(X_n) \mid A] = 0$$

and so $(M_n^f)_{n \geq 0}$ is a martingale.

On the other hand, if (ii) holds, then

$$\mathbb{E}[f(X_{n+1}) - (Pf)(X_n) \mid X_0 = i_0, \dots, X_n = i_n] = 0$$

for all bounded functions f . On taking $f = 1_{\{i_{n+1}\}}$ we obtain

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n) = p_{i_n i_{n+1}}$$

so $(X_n)_{n \geq 0}$ is Markov with transition matrix P . \square

Some more martingales associated to a Markov chain are described in the next result. Notice that we drop the requirement that f be bounded.

Theorem 4.1.3. *Let $(X_n)_{n \geq 0}$ be a Markov chain with transition matrix P . Suppose that a function $f : \mathbb{Z}^+ \times I \rightarrow \mathbb{R}$ satisfies, for all $n \geq 0$, both*

$$\mathbb{E}|f(n, X_n)| < \infty$$

and

$$(Pf)(n+1, i) = \sum_{j \in I} p_{ij} f(n+1, j) = f(n, i).$$

Then $M_n = f(n, X_n)$ is a martingale.

Proof. We have assumed that M_n is integrable for all n . Then, by the Markov property

$$\begin{aligned} \mathbb{E}(M_{n+1} - M_n \mid X_0 = i_0, \dots, X_n = i_n) &= \mathbb{E}_{i_n}[f(n+1, X_1) - f(n, X_0)] \\ &= (Pf)(n+1, i_n) - f(n, i_n) = 0. \end{aligned}$$

So $(M_n)_{n \geq 0}$ is a martingale. \square

Let us see how this theorem works in the case where $(X_n)_{n \geq 0}$ is a simple random walk on \mathbb{Z} , starting from 0. We consider $f(i) = i$ and $g(n, i) = i^2 - n$. Since $|X_n| \leq n$ for all n , we have

$$\mathbb{E}|f(X_n)| < \infty, \quad \mathbb{E}|g(n, X_n)| < \infty.$$

Also

$$(Pf)(i) = (i-1)/2 + (i+1)/2 = i = f(i),$$

$$(Pg)(n+1, i) = (i-1)^2/2 + (i+1)^2/2 - (n+1) = i^2 - n = g(n, i).$$

Hence both $X_n = f(X_n)$ and $Y_n = g(n, X_n)$ are martingales.

In order to put this to some use, consider again the stopping time

$$T = \inf\{n \geq 0 : X_n = -a \quad \text{or} \quad X_n = b\}$$

where $a, b \in \mathbb{N}$. By the optional stopping theorem

$$0 = \mathbb{E}(Y_0) = \mathbb{E}(Y_{T \wedge n}) = \mathbb{E}(X_{T \wedge n}^2) - \mathbb{E}(T \wedge n).$$

Hence

$$\mathbb{E}(T \wedge n) = \mathbb{E}(X_{T \wedge n}^2).$$

On letting $n \rightarrow \infty$, the left side converges to $\mathbb{E}(T)$, by monotone convergence, and the right side to $\mathbb{E}(X_T^2)$ by bounded convergence. So we obtain

$$\mathbb{E}(T) = \mathbb{E}(X_T^2) = a^2 p + b^2(1-p) = ab.$$

We have given only the simplest examples of the use of martingales in studying Markov chains. Some more will appear in later sections. For an excellent introduction to martingales and their applications we recommend *Probability with Martingales* by David Williams (Cambridge University Press, 1991).

Exercise

4.1.1 Let $(X_n)_{n \geq 0}$ be a Markov chain on I and let A be an absorbing set in I . Set

$$T = \inf\{n \geq 0 : X_n \in A\}$$

and

$$h_i = \mathbb{P}_i(X_n \in A \text{ for some } n \geq 0) = \mathbb{P}_i(T < \infty).$$

Show that $M_n = h(X_n)$ is a martingale.

4.2 Potential theory

Several physical theories share a common mathematical framework, which is known as potential theory. One example is Newton's theory of gravity, but potential theory is also relevant to electrostatics, fluid flow and the diffusion of heat. In gravity, a distribution of mass, of density ρ say, gives rise to a gravitational potential ϕ , which in suitable units satisfies the equation

$$-\Delta\phi = \rho,$$

where $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$. The potential ϕ is felt physically through its gradient

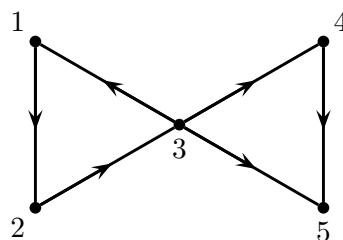
$$\nabla\phi = \left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z}\right)$$

which gives the force of gravity acting on a particle of unit mass. Markov chains, where space is discrete, obviously have no direct link with this theory, in which space is a continuum. An indirect link is provided by Brownian motion, which we shall discuss in Section 4.4.

In this section we are going to consider *potential theory for a countable state-space*, which has much of the structure of the continuum version. This discrete theory amounts to doing Markov chains without the probability, which has the disadvantage that one loses the intuitive picture of the process, but the advantage of wider applicability. We shall begin by introducing the idea of potentials associated to a Markov chain, and by showing how to calculate these potentials. This is a unifying idea, containing within it other notions previously considered such as hitting probabilities and expected hitting times. It also finds application when one associates costs to Markov chains in modelling economic activity: see Section 5.4.

Once we have established the basic link between a Markov chain and its associated potentials, we shall briefly run through some of the main features of potential theory, explaining their significance in terms of Markov chains. This is the easiest way to appreciate the general structure of potential theory, unobscured by technical difficulties. The basic ideas of boundary theory for Markov chains will also be introduced.

Before we embark on a general discussion of potentials associated to a Markov chain, here are two simple examples. In these examples the potential ϕ has the interpretation of expected total cost.



Example 4.2.1

Consider the discrete-time random walk on the directed graph shown above, which at each step chooses among the allowable transitions with equal probability. Suppose that on each visit to states $i = 1, 2, 3, 4$ a cost c_i is incurred, where $c_i = i$. What is the fair price to move from state 3 to state 4?

The fair price is always the difference in the expected total cost. We denote by ϕ_i the expected total cost starting from i . Obviously, $\phi_5 = 0$ and

by considering the effect of a single step we see that

$$\begin{aligned}\phi_1 &= 1 + \phi_2, \\ \phi_2 &= 2 + \phi_3, \\ \phi_3 &= 3 + \frac{1}{3}\phi_1 + \frac{1}{3}\phi_4, \\ \phi_4 &= 4.\end{aligned}$$

Hence $\phi_3 = 8$ and the fair price to move from 3 to 4 is 4.

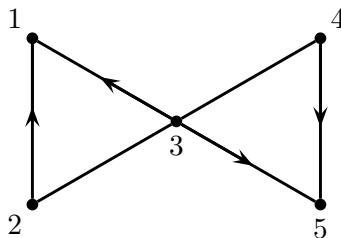
We shall now consider two variations on this problem. First suppose our process is, instead, the continuous-time random walk $(X_t)_{t \geq 0}$ on the same directed graph which makes each allowable transition at rate 1, and suppose cost is incurred at rate $c_i = i$ in state i for $i = 1, 2, 3, 4$. Thus the total cost is now

$$\int_0^\infty c(X_s)ds.$$

What now is the fair price to move from 3 to 4? The expected cost incurred on each visit to i is given by c_i/q_i and $q_1 = 1, q_2 = 1, q_3 = 3, q_4 = 1$. So we see, as before

$$\begin{aligned}\phi_1 &= 1 + \phi_2, \\ \phi_2 &= 2 + \phi_3, \\ \phi_3 &= \frac{3}{3} + \frac{1}{3}\phi_1 + \frac{1}{3}\phi_4, \\ \phi_4 &= 4.\end{aligned}$$

Hence $\phi_3 = 5$ and the fair price to move from 3 to 4 is 1.



In the second variation we consider the discrete-time random walk $(X_n)_{n \geq 0}$ on the modified graph shown above. Where there is no arrow, transitions are allowed in both directions. Obviously, states 1 and 5 are absorbing. We impose a cost $c_i = i$ on each visit to i for $i = 2, 3, 4$, and a final cost f_i on arrival at $i = 1$ or 5 , where $f_1 = f_5 = 1$. Thus the total cost is now

$$\sum_{n=0}^{T-1} c(X_n) + f(X_T)$$

where T is the hitting time of $\{1, 5\}$. Write, as before, ϕ_i for the expected total cost starting from i . Then $\phi_1 = 1$, $\phi_5 = 5$ and

$$\begin{aligned}\phi_2 &= 2 + \frac{1}{2}(\phi_1 + \phi_3), \\ \phi_3 &= 3 + \frac{1}{4}(\phi_1 + \phi_2 + \phi_4 + \phi_5), \\ \phi_4 &= 4 + \frac{1}{2}(\phi_3 + \phi_5).\end{aligned}$$

On solving these equations we obtain $\phi_2 = 7$, $\phi_3 = 9$ and $\phi_4 = 11$. So in this case the fair price to move from 3 to 4 is -2 .

Example 4.2.2

Consider the simple discrete-time random walk on \mathbb{Z} with transition probabilities $p_{i,i-1} = q < p = p_{i,i+1}$. Let $c > 0$ and suppose that a cost c^i is incurred every time the walk visits state i . What is the expected total cost ϕ_0 incurred by the walk starting from 0?

We must be prepared to find that $\phi_0 = \infty$ for some values of c , as the total cost is a sum over infinitely many times. Indeed, we know that the walk $X_n \rightarrow \infty$ with probability 1, so for $c \geq 1$ we shall certainly have $\phi_0 = \infty$.

Let ϕ_i denote the expected total cost starting from i . On moving one step to the right, all costs are multiplied by c , so we must have

$$\phi_{i+1} = c\phi_i.$$

By considering what happens on the first step, we see

$$\phi_0 = 1 + p\phi_1 + q\phi_{-1} = 1 + (cp + q/c)\phi_0.$$

Note that $\phi_0 = \infty$ always satisfies this equation. We shall see in the general theory that ϕ_0 is the minimal non-negative solution. Let us look for a finite solution: then

$$-(c^2p - c + q)\phi_0 = c$$

so

$$\phi_0 = \frac{c}{c - c^2p - q}.$$

The quadratic $c^2p - c + q$ has roots at q/p and 1, and takes negative values in between. Hence the expected total cost is given by

$$\phi_0 = \begin{cases} c/(c - c^2p - q) & \text{if } c \in (q/p, 1) \\ \infty & \text{otherwise.} \end{cases}$$

It was clear at the outset that $\phi_0 = \infty$ when $c \geq 1$. It is interesting that $\phi_0 = \infty$ also when c is too small: in this case the costs rapidly become large to the left of 0, and although the walk eventually drifts away to the right, the expected cost incurred to the left of 0 is infinite.

In the examples just discussed we were able to calculate potentials by writing down and solving a system of linear equations. This situation is familiar from hitting probabilities and expected hitting times. Indeed, these are simple examples of potentials for Markov chains. As the examples show, one does not really need a general theory to write down the linear equations. Nevertheless, we are now going to give some general results on potentials. These will help to reveal the scope of the ideas used in the examples, and will reveal also what happens when the linear equations do not have a unique solution. We shall discuss the cases of discrete and continuous time side-by-side. Throughout, we shall write $(X_n)_{n \geq 0}$ for a discrete-time chain with transition matrix P , and $(X_t)_{t \geq 0}$ for a continuous-time chain with generator matrix Q . As usual, we insist that $(X_t)_{t \geq 0}$ be minimal.

Let us partition the state-space I into two disjoint sets D and ∂D ; we call ∂D the *boundary*. We suppose that functions $(c_i : i \in D)$ and $(f_i : i \in \partial D)$ are given. We shall consider the associated *potential*, defined by

$$\phi_i = \mathbb{E}_i \left(\sum_{n < T} c(X_n) + f(X_T) 1_{T < \infty} \right)$$

in discrete time, and in continuous time

$$\phi_i = \mathbb{E}_i \left(\int_0^T c(X_t) dt + f(X_T) 1_{T < \infty} \right),$$

where T denotes the hitting time of ∂D . To be sure that the sums and integrals here are well defined, we shall assume for the most part that c and f are *non-negative*, that is, $c_i \geq 0$ for all $i \in D$ and $f_i \geq 0$ for all $i \in \partial D$. More generally, ϕ is the difference of the potentials associated with the positive and negative parts of c and f , so this assumption is not too restrictive. In the explosive case we always set $c(\infty) = 0$, so no further costs are incurred after explosion.

The most obvious interpretation of these potentials is in terms of cost: the chain wanders around in D until it hits the boundary: whilst in D , at state i say, it incurs a *cost* c_i per unit time; when and if it hits the boundary, at j say, a *final cost* f_j is incurred. Note that we do not assume the chain will hit the boundary, or even that the boundary is non-empty.

Theorem 4.2.3. Suppose that $(c_i : i \in D)$ and $(f_i : i \in \partial D)$ are non-negative. Set

$$\phi_i = \mathbb{E}_i \left(\sum_{n < T} c(X_n) + f(X_T) 1_{T < \infty} \right)$$

where T denotes the hitting time of ∂D . Then

(i) the potential $\phi = (\phi_i : i \in I)$ satisfies

$$\begin{cases} \phi = P\phi + c & \text{in } D \\ \phi = f & \text{in } \partial D; \end{cases} \quad (4.1)$$

(ii) if $\psi = (\psi_i : i \in I)$ satisfies

$$\begin{cases} \psi \geq P\psi + c & \text{in } D \\ \psi \geq f & \text{in } \partial D \end{cases} \quad (4.2)$$

and $\psi_i \geq 0$ for all i , then $\psi_i \geq \phi_i$ for all i ;

(iii) if $\mathbb{P}_i(T < \infty) = 1$ for all i , then (4.1) has at most one bounded solution.

Proof. (i) Obviously, $\phi = f$ on ∂D . For $i \in D$ by the Markov property

$$\begin{aligned} \mathbb{E}_i \left(\sum_{1 \leq n < T} c(X_n) + f(X_T) 1_{T < \infty} \middle| X_1 = j \right) \\ = \mathbb{E}_j \left(\sum_{n < T} c(X_n) + f(X_T) 1_{T < \infty} \right) = \phi_j \end{aligned}$$

so we have

$$\begin{aligned} \phi_i &= c_i + \sum_{j \in I} p_{ij} \mathbb{E} \left(\sum_{1 \leq n < T} c(X_n) + f(X_T) 1_{T < \infty} \middle| X_1 = j \right) \\ &= c_i + \sum_{j \in I} p_{ij} \phi_j \end{aligned}$$

as required.

(ii) Consider the expected cost up to time n :

$$\phi_i(n) = \mathbb{E}_i \left(\sum_{k=0}^n c(X_k) 1_{k < T} + f(X_T) 1_{T \leq n} \right).$$

By monotone convergence, $\phi_i(n) \uparrow \phi_i$ as $n \rightarrow \infty$. Also, by the argument used in part (i), we find

$$\begin{cases} \phi(n+1) = c + P\phi(n) & \text{in } D \\ \phi(n+1) = f & \text{in } \partial D. \end{cases}$$

Suppose that ψ satisfies (4.2) and $\psi \geq 0 = \phi(0)$. Then $\psi \geq P\psi + c \geq P\phi(0) + c = \phi(1)$ in D and $\psi \geq f = \phi(1)$ in ∂D , so $\psi \geq \phi(1)$. Similarly and by induction, $\psi \geq \phi(n)$ for all n , and hence $\psi \geq \phi$.

(iii) We shall show that if ψ satisfies (4.2) then

$$\psi_i \geq \phi_i(n-1) + \mathbb{E}_i(\psi(X_n)1_{T \geq n}),$$

with equality if equality holds in (4.2). This is another proof of (ii). But also, in the case of equality, if $|\psi_i| \leq M$ and $\mathbb{P}_i(T < \infty) = 1$ for all i , then as $n \rightarrow \infty$

$$|\mathbb{E}_i(\psi(X_n)1_{T \geq n})| \leq M\mathbb{P}_i(T \geq n) \rightarrow 0$$

so $\psi = \lim_{n \rightarrow \infty} \phi(n) = \phi$, proving (iii).

For $i \in D$ we have

$$\psi_i \geq c_i + \sum_{j \in \partial D} p_{ij}f_j + \sum_{j \in D} p_{ij}\psi_j$$

and, by repeated substitution for ψ on the right

$$\begin{aligned} \psi_i &\geq c_i + \sum_{j \in \partial D} p_{ij}f_j + \sum_{j \in D} p_{ij}c_j \\ &\quad + \dots + \sum_{j_1 \in D} \dots \sum_{j_{n-1} \in D} p_{ij_1} \dots p_{j_{n-2}j_{n-1}} c_{j_{n-1}} \\ &\quad + \sum_{j_1 \in D} \dots \sum_{j_{n-1} \in D} \sum_{j_n \in \partial D} p_{ij_1} \dots p_{j_{n-1}j_n} f_{j_n} \\ &\quad + \sum_{j_1 \in D} \dots \sum_{j_n \in D} p_{ij_1} \dots p_{j_{n-1}j_n} \psi_{j_n} \\ &= \mathbb{E}_i \left(c(X_0)1_{T>0} + f(X_1)1_{T=1} + c(X_1)1_{T>1} \right. \\ &\quad \left. + \dots + c(X_{n-1})1_{T>n-1} + f(X_n)1_{T=n} + \psi(X_n)1_{T>n} \right) \\ &= \phi_i(n-1) + \mathbb{E}_i(\psi(X_n)1_{T \geq n}) \end{aligned}$$

as required, with equality when equality holds in (4.2). \square

It is illuminating to think of the calculation we have just done in terms of martingales. Consider

$$M_n = \sum_{k=0}^{n-1} c(X_k)1_{k < T} + f(X_T)1_{T < n} + \psi(X_n)1_{n \leq T}.$$

Then

$$\begin{aligned} \mathbb{E}(M_{n+1} | \mathcal{F}_n) &= \sum_{k=0}^{n-1} c(X_k)1_{k < T} + f(X_T)1_{T < n} \\ &\quad + (P\psi + c)(X_n)1_{T > n} + f(X_n)1_{T=n} \\ &\leq M_n \end{aligned}$$

with equality if equality holds in (4.2). We note that M_n is not necessarily integrable. Nevertheless, it still follows that

$$\psi_i = \mathbb{E}_i(M_0) \geq \mathbb{E}_i(M_n) = \phi_i(n-1) + \mathbb{E}_i(\psi(X_n)1_{T \geq n})$$

with equality if equality holds in (4.2).

For continuous-time chains there is a result analogous to Theorem 4.2.3. We have to state it slightly differently because when ϕ takes infinite values the equations (4.3) may involve subtraction of infinities, and therefore not make sense. Although the conclusion then appears to depend on the finiteness of ϕ , which is *a priori* unknown, we can still use the result to determine ϕ_i in all cases. To do this we restrict our attention to the set of states J accessible from i . If the linear equations have a finite non-negative solution on J , then $(\phi_j : j \in J)$ is the minimal such solution. If not, then $\phi_j = \infty$ for some $j \in J$, which forces $\phi_i = \infty$, since i leads to j .

Theorem 4.2.4. Assume that $(X_t)_{t \geq 0}$ is minimal, and that $(c_i : i \in D)$ and $(f_i : i \in \partial D)$ are non-negative. Set

$$\phi_i = \mathbb{E}_i \left(\int_0^T c(X_t)dt + f(X_T)1_{T < \infty} \right)$$

where T is the hitting time of ∂D . Then $\phi = (\phi_i : i \in I)$, if finite, is the minimal non-negative solution to

$$\begin{cases} -Q\phi = c & \text{in } D \\ \phi = f & \text{in } \partial D. \end{cases} \quad (4.3)$$

If $\phi_i = \infty$ for some i , then (4.3) has no finite non-negative solution. Moreover, if $\mathbb{P}_i(T < \infty) = 1$ for all i , then (4.3) has at most one bounded solution.

Proof. Denote by $(Y_n)_{n \geq 0}$ and S_1, S_2, \dots the jump chain and holding times of $(X_t)_{t \geq 0}$, and by Π the jump matrix. Then

$$\int_0^T c(X_t)dt + f(X_T)1_{T < \infty} = \sum_{n < N} c(Y_n)S_{n+1} + f(Y_N)1_{N < \infty}$$

where N is the first time $(Y_n)_{n \geq 0}$ hits ∂D , and where we use the convention $0 \times \infty = 0$ on the right. We have

$$\mathbb{E}(c(Y_n)S_{n+1} \mid Y_n = j) = \tilde{c}_j = \begin{cases} c_j/q_j & \text{if } c_j > 0 \\ 0 & \text{if } c_j = 0, \end{cases}$$

so, by Fubini's theorem

$$\phi_i = \mathbb{E}_i \left(\sum_{n < N} \tilde{c}(Y_n) + f(Y_N)1_{N < \infty} \right).$$

By Theorem 4.2.3, ϕ is therefore the minimal non-negative solution to

$$\begin{cases} \phi = \Pi\phi + \tilde{c} & \text{in } D \\ \phi = f & \text{in } \partial D, \end{cases} \quad (4.4)$$

which equations have at most one bounded solution if $\mathbb{P}_i(N < \infty) = 1$ for all i . Since the finite solutions of (4.4) are exactly the finite solutions of (4.3), and since N is finite whenever T is finite, this proves the result. \square

It is natural in some economic applications to apply to future costs a discount factor $\alpha \in (0, 1)$ or rate $\lambda \in (0, \infty)$, corresponding to an interest rate. *Potentials with discounted costs* may also be calculated by linear equations; indeed the discounting actually makes the analysis easier.

Theorem 4.2.5. Suppose that $(c_i : i \in I)$ is bounded. Set

$$\phi_i = \mathbb{E}_i \sum_{n=0}^{\infty} \alpha^n c(X_n)$$

then $\phi = (\phi_i : i \in I)$ is the unique bounded solution to

$$\phi = \alpha P\phi + c.$$

Proof. Suppose that $|c_i| \leq C$ for all i , then

$$|\phi_i| \leq C \sum_{n=0}^{\infty} \alpha^n = C/(1 - \alpha)$$

so ϕ is bounded. By the Markov property

$$\mathbb{E} \left(\sum_{n=1}^{\infty} \alpha^{n-1} c(X_n) \middle| X_1 = j \right) = \mathbb{E}_j \sum_{n=0}^{\infty} \alpha^n c(X_n) = \phi_j.$$

Then

$$\begin{aligned} \phi_i &= \mathbb{E}_i \sum_{n=0}^{\infty} \alpha^n c(X_n) \\ &= c_i + \alpha \sum_{j \in I} p_{ij} \mathbb{E} \left(\sum_{n=1}^{\infty} \alpha^{n-1} c(X_n) \middle| X_1 = j \right) \\ &= c_i + \alpha \sum_{j \in I} p_{ij} \phi_j, \end{aligned}$$

so

$$\phi = c + \alpha P\phi.$$

On the other hand, suppose that ψ is bounded and also that $\psi = c + \alpha P\psi$. Set $M = \sup_i |\psi_i - \phi_i|$, then $M < \infty$. But

$$\psi - \phi = \alpha P(\psi - \phi)$$

so

$$|\psi_i - \phi_i| \leq \alpha \sum_{j \in I} p_{ij} |\psi_j - \phi_j| \leq \alpha M.$$

Hence $M \leq \alpha M$, which forces $M = 0$ and $\psi = \phi$. \square

We have a similar looking result for continuous time, which however lies a little deeper, because it really corresponds to a version of the discrete-time result where the discount factor may depend on the current state.

Theorem 4.2.6. Assume that $(X_t)_{t \geq 0}$ is non-explosive. Suppose that $(c_i : i \in I)$ is bounded. Set

$$\phi_i = \mathbb{E}_i \int_0^{\infty} e^{-\lambda t} c(X_t) dt,$$

then $\phi = (\phi_i : i \in I)$ is the unique bounded solution to

$$(\lambda - Q)\phi = c. \tag{4.5}$$

Proof. Assume for now that c is non-negative. Introduce a new state ∂ with $c_\partial = 0$. Let T be an independent $E(\lambda)$ random variable and define

$$\tilde{X}_t = \begin{cases} X_t & \text{for } t < T \\ \partial & \text{for } t \geq T. \end{cases}$$

Then $(\tilde{X}_t)_{t \geq 0}$ is a Markov chain on $I \cup \{\partial\}$ with modified transition rates

$$\tilde{q}_i = q_i + \lambda, \quad \tilde{q}_{i\partial} = \lambda, \quad \tilde{q}_\partial = 0.$$

Also T is the hitting time of ∂ , and is finite with probability 1. By Fubini's theorem

$$\phi_i = \mathbb{E}_i \int_0^T c(\tilde{X}_t) dt.$$

Suppose $c_i \leq C$ for all i , then

$$\phi_i \leq C \int_0^\infty e^{-\lambda t} dt \leq C/\lambda,$$

so ϕ is bounded. Hence, by Theorem 4.2.4, ϕ is the unique bounded solution to

$$-\tilde{Q}\phi = c,$$

which is the same as (4.5).

When c takes negative values we can apply the preceding argument to the potentials

$$\phi_i^\pm = \mathbb{E}_i \int_0^\infty e^{-\lambda t} c_i^\pm(X_t) dt$$

where $c_i^\pm = (\pm c) \vee 0$. Then $\phi = \phi^+ - \phi^-$ so ϕ is bounded. We have

$$(\lambda - Q)\phi^\pm = c^\pm$$

so, subtracting

$$(\lambda - Q)\phi = c.$$

Finally, if ψ is bounded and $(\lambda - Q)\psi = c$, then $(\lambda - Q)(\psi - \phi) = 0$, so $\psi - \phi$ is the unique bounded solution for the case when $c = 0$, which is 0. \square

The point of view underlying the last four theorems was that we were interested in a given potential associated to a Markov chain, and wished to calculate it. We shall now take a brief look at some structural aspects of the set of all potentials of a given Markov chain. What we describe is just the simplest case of a structure of great generality. First we shall look at the Green matrix, and then at the role of the boundary.

Let us consider potentials with non-negative costs c , and without boundary. The potential is defined by

$$\phi_i = \mathbb{E}_i \sum_{n=0}^{\infty} c(X_n)$$

in discrete time, and in continuous time

$$\phi_i = \mathbb{E}_i \int_0^{\infty} c(X_t) dt.$$

By Fubini's theorem we have

$$\phi_i = \sum_{n=0}^{\infty} \mathbb{E}_i c(X_n) = \sum_{n=0}^{\infty} (P^n c)_i = (Gc)_i$$

where $G = (g_{ij} : i, j \in I)$ is the *Green* matrix

$$G = \sum_{n=0}^{\infty} P^n.$$

Similarly, in continuous time $\phi = Gc$, with

$$G = \int_0^{\infty} P(t) dt.$$

Thus, once we know the Green matrix, we have explicit expressions for all potentials of the Markov chain. The Green matrix is also called the *fundamental solution* of the linear equations (4.1) and (4.3). The j th column $(g_{ij} : i \in I)$ is itself a potential. We have

$$g_{ij} = \mathbb{E}_i \sum_{n=0}^{\infty} 1_{X_n=j}$$

in discrete time, and in continuous time

$$g_{ij} = \mathbb{E}_i \int_0^{\infty} 1_{X_t=j} dt.$$

Thus g_{ij} is the expected total time in j starting from i . These quantities have already appeared in our discussions of transience and recurrence in Sections 1.5 and 2.11: we know that $g_{ij} = \infty$ if and only if i leads to j and j is recurrent. Indeed, in discrete time

$$g_{ij} = h_i^j / (1 - f_j)$$

where h_i^j is the probability of hitting j from i , and f_j is the return probability for j . The formula for continuous time is

$$g_{ij} = h_i^j/q_j(1 - f_j).$$

For potentials with discounted costs the situation is similar: in discrete time

$$\phi_i = \mathbb{E}_i \sum_{n=0}^{\infty} \alpha^n c(X_n) = \sum_{n=0}^{\infty} \alpha^n \mathbb{E}_i c(X_n) = (R_\alpha c)_i$$

where

$$R_\alpha = \sum_{n=0}^{\infty} \alpha^n P^n,$$

and in continuous time

$$\phi_i = \mathbb{E}_i \int_0^{\infty} e^{-\lambda t} c(X_t) dt = \int_0^{\infty} e^{-\lambda t} \mathbb{E}_i c(X_t) dt = (R_\lambda c)_i$$

where

$$R_\lambda = \int_0^{\infty} e^{-\lambda t} P(t) dt.$$

We call $(R_\alpha : \alpha \in (0, 1))$ and $(R_\lambda : \lambda \in (0, \infty))$ the *resolvent* of the Markov chain. Unlike the Green matrix the resolvent is always finite. Indeed, for finite state-space we have

$$R_\alpha = (I - \alpha P)^{-1}$$

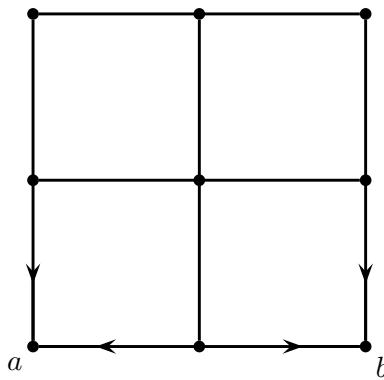
and

$$R_\lambda = (\lambda I - Q)^{-1}.$$

We return to the general case, with boundary ∂D . Any bounded function $(\phi_i : i \in I)$ for which

$$\phi = P\phi \quad \text{in } D$$

is called *harmonic* in D . Our object now is to examine the relation between non-negative functions, harmonic in D , and the boundary ∂D . Here are two examples.



Example 4.2.7

Consider the random walk $(X_n)_{n \geq 0}$ on the above graph, where each allowable transition is made with equal probability. States a and b are absorbing. We set $\partial D = \{a, b\}$. Let h_i^a denote the absorption probability for a , starting from i . By the method of Section 1.3 we find

$$h^a = \begin{pmatrix} 3/5 & 1/2 & 2/5 \\ 7/10 & 1/2 & 3/10 \\ 1 & 1/2 & 0 \end{pmatrix}$$

where we have written the vector h^a as a matrix, corresponding in an obvious way to the state-space. The linear equations for the vector h^a read

$$\begin{cases} h^a = Ph^a & \text{in } D \\ h_a^a = 1, h_b^a = 0. \end{cases}$$

Thus we can find two non-negative functions h^a and h^b , harmonic in D , but with different boundary values. In fact, the most general non-negative harmonic function ϕ in D satisfies

$$\begin{cases} \phi = P\phi & \text{in } D \\ \phi = f & \text{in } \partial D \end{cases}$$

where $f_a, f_b \geq 0$, and this implies

$$\phi = f_a h^a + f_b h^b.$$

Thus the boundary points a and b give us extremal generators h^a and h^b of the set of all non-negative harmonic functions.

Example 4.2.8

Consider the random walk $(X_n)_{n \geq 0}$ on \mathbb{Z} which jumps towards 0 with probability q and jumps away from 0 with probability $p = 1 - q$, except that at 0 it jumps to -1 or 1 with probability $1/2$. We choose $p > q$ so that the walk is transient. In fact, starting from 0, we can show that $(X_n)_{n \geq 0}$ is equally likely to end up drifting to the left or to the right, at speed $p - q$. Let us consider the problem of determining for $(X_n)_{n \geq 0}$ the set C of all non-negative harmonic functions ϕ . We must have

$$\begin{aligned}\phi_i &= p\phi_{i+1} + q\phi_{i-1} && \text{for } i = 1, 2, \dots, \\ \phi_0 &= \frac{1}{2}\phi_1 + \frac{1}{2}\phi_{-1}, \\ \phi_i &= q\phi_{i+1} + p\phi_{i-1} && \text{for } i = -1, -2, \dots.\end{aligned}$$

The first equation has general solution

$$\phi_i = A + B(1 - (q/p)^i) \quad \text{for } i = 0, 1, 2, \dots,$$

which is non-negative provided $A + B \geq 0$. Similarly, the third equation has general solution

$$\phi_i = A' + B'(1 - (q/p)^{-i}) \quad \text{for } i = 0, -1, -2, \dots,$$

non-negative provided $A' + B' \geq 0$. To obtain a general harmonic function we must match the values ϕ_0 and satisfy $\phi_0 = (\phi_1 + \phi_{-1})/2$. This forces $A = A'$ and $B + B' = 0$. It follows that all non-negative harmonic functions have the form

$$\phi = f^- h^- + f^+ h^+$$

where $f^-, f^+ \geq 0$ and where $h_i^- = h_{-i}^+$ and

$$h_i^+ = \begin{cases} \frac{1}{2} + \frac{1}{2}(1 - (q/p)^i) & \text{for } i = 0, 1, 2, \dots, \\ \frac{1}{2} - \frac{1}{2}(1 - (q/p)^{-i}) & \text{for } i = -1, -2, \dots. \end{cases}$$

In the preceding example the generators of C were in one-to-one correspondence with the points of the boundary – the possible places for the chain to end up. In this example there is no boundary, *but the generators of C still correspond to the two possibilities for the long-time behaviour of the chain.* For we have

$$h_i^+ = \mathbb{P}_i(X_n \rightarrow \infty \text{ as } n \rightarrow \infty).$$

The suggestion of this example, which is fully developed in other works, is that the set of non-negative harmonic functions may be used to identify a

generalized notion of boundary for Markov chains, which sometimes just consists of points in the state-space, but more generally corresponds to the varieties of possible limiting behaviour for X_n as $n \rightarrow \infty$. See, for example, *Markov Chains* by D. Revuz (North-Holland, Amsterdam, 1984).

We cannot begin to give the general theory corresponding to Example 4.2.8, but we can draw some general conclusions from Theorem 4.2.3 when the situation is more like Example 4.2.7. Suppose we have a Markov chain $(X_n)_{n \geq 0}$ with absorbing boundary ∂D . Set

$$h_i^\partial = \mathbb{P}_i(T < \infty)$$

where T is the hitting time of ∂D . Then by the methods of Section 1.3 we have

$$\begin{cases} h^\partial = Ph^\partial & \text{in } D \\ h^\partial = 1 & \text{in } \partial D. \end{cases} \quad (4.6)$$

Note that $h_i^\partial = 1$ for all i always gives a possible solution. Hence if (4.6) has a unique bounded solution then $h_i^\partial = \mathbb{P}_i(T < \infty) = 1$ for all i . Conversely, if $\mathbb{P}_i(T < \infty) = 1$ for all i , then, as we showed in Theorem 4.2.3, (4.6) has a unique bounded solution. Indeed, we showed more generally that this condition implies that

$$\begin{cases} \phi = P\phi + c & \text{in } D \\ \phi = f & \text{in } \partial D \end{cases}$$

has at most one bounded solution, and since

$$\phi_i = \mathbb{E}_i \left(\sum_{n < T} c(X_n) + f(X_T) 1_{T < \infty} \right) \quad (4.7)$$

is the minimal solution, any bounded solution is given by (4.7). Suppose from now on that $\mathbb{P}_i(T < \infty) = 1$ for all i . Let ϕ be a bounded non-negative function, harmonic in D , with boundary values $\phi_i = f_i$ for $i \in \partial D$. Then, by monotone convergence

$$\phi_i = \mathbb{E}_i(f(X_T)) = \sum_{j \in \partial D} f_j \mathbb{P}_i(X_T = j).$$

Hence every bounded harmonic function is determined by its boundary values and, indeed

$$\phi = \sum_{j \in \partial D} f_j h_j^i,$$

where

$$h_i^j = \mathbb{P}_i(X_T = j).$$

Just as in Example 4.2.7, the hitting probabilities for boundary states form a set of extremal generators for the set of all bounded non-negative harmonic functions.

Exercises

4.2.1 Consider a discrete-time Markov chain $(X_n)_{n \geq 0}$ and the potential ϕ with costs $(c_i : i \in D)$ and boundary values $(f_i : i \in \partial D)$. Set

$$\tilde{X}_n = \begin{cases} X_n & \text{if } n \leq T \\ \partial & \text{if } n > T, \end{cases}$$

where T is the hitting time of ∂D and ∂ is a new state. Show that $(\tilde{X}_n)_{n \geq 0}$ is a Markov chain and determine its transition matrix.

Check that

$$\phi_i = \mathbb{E}_i \sum_{n < \tilde{T}} c(\tilde{X}_n) = \mathbb{E}_i \sum_{n=0}^{\infty} c(\tilde{X}_n)$$

where $\tilde{T} = T + 1$ and where we set $c_i = f_i$ on ∂D and $c_{\partial} = 0$. This shows that a general potential may always be considered as a potential with boundary value zero or, indeed, without boundary at all.

Can you find a similar reduction for continuous-time chains?

4.2.2 Prove the fact claimed in Example 4.2.8 that

$$h_i^+ = \mathbb{P}_i(X_n \rightarrow \infty \text{ as } n \rightarrow \infty).$$

4.2.3 Let $(c_i : i \in I)$ be a non-negative function. Partition I as $D \cup \partial D$ and suppose that the linear equations

$$\begin{cases} \phi = P\phi + c & \text{in } D \\ \phi = 0 & \text{in } \partial D \end{cases}$$

have a unique bounded solution. Show that the Markov chain $(X_n)_{n \geq 0}$ with transition matrix P is certain to hit ∂D .

Consider now a new partition $\tilde{D} \cup \partial \tilde{D}$, where $\tilde{D} \subseteq D$. Show that the linear equations

$$\begin{cases} \psi = P\psi + c & \text{in } \tilde{D} \\ \psi = 0 & \text{in } \partial \tilde{D} \end{cases}$$

also have a unique bounded solution, and that

$$\psi_i \leq \phi_i \quad \text{for all } i \in I.$$

4.3 Electrical networks

An electrical network has a countable set I of *nodes*, each node i having a *capacity* $\pi_i > 0$. Some nodes are joined by *wires*, the wire between i and j having *conductivity* $a_{ij} = a_{ji} \geq 0$. Where no wire joins i to j we take $a_{ij} = 0$. In practice, each ‘wire’ contains a resistor, which determines the conductivity as the reciprocal of its resistance. Each node i holds a certain *charge* χ_i , which determines its *potential* ϕ_i by

$$\chi_i = \phi_i \pi_i.$$

A *current* or *flow of charge* is any matrix $(\gamma_{ij} : i, j \in I)$ with $\gamma_{ij} = -\gamma_{ji}$. Physically it is found that the current γ_{ij} from i to j obeys *Ohm’s law*:

$$\gamma_{ij} = a_{ij}(\phi_i - \phi_j).$$

Thus charge flows from nodes of high potential to nodes of low potential.

The first problem in electrical networks is to determine equilibrium flows and potentials, subject to given external conditions. The nodes are partitioned into two sets D and ∂D . External connections are made at the nodes in ∂D and possibly at some of the nodes in D . These have the effect that each node $i \in \partial D$ is held at a given potential f_i , and that a given current g_i enters the network at each node $i \in D$. The case where $g_i = 0$ corresponds to a node with no external connection. In equilibrium, current may also enter or leave the network through ∂D , but here it is not the current but the potential which is determined externally.

Given a flow $(\gamma_{ij} : i, j \in I)$ we shall write γ_i for the *total flow from i to the network*:

$$\gamma_i = \sum_{j \in I} \gamma_{ij}.$$

In equilibrium the charge at each node is constant, so

$$\gamma_i = g_i \quad \text{for } i \in D.$$

Therefore, by Ohm’s law, any equilibrium potential $\phi = (\phi_i : i \in I)$ must satisfy

$$\begin{cases} \sum_{j \in I} a_{ij}(\phi_i - \phi_j) = g_i, & \text{for } i \in D \\ \phi_i = f_i, & \text{for } i \in \partial D. \end{cases} \quad (4.8)$$

There is a simple correspondence between electrical networks and reversible Markov chains in continuous-time, given by

$$a_{ij} = \pi_i q_{ij} \quad \text{for } i \neq j.$$

We shall assume that the total conductivity at each node is finite:

$$a_i = \sum_{j \neq i} a_{ij} < \infty.$$

Then $a_i = \pi_i q_i = -\pi_i q_{ii}$. The capacities π_i are the components of an invariant measure, and the symmetry of a_{ij} corresponds to the detailed balance equations. The equations for an equilibrium potential may now be written in a form familiar from the preceding section:

$$\begin{cases} -Q\phi = c & \text{in } D \\ \phi = f & \text{in } \partial D, \end{cases} \quad (4.9)$$

where $c_i = g_i/\pi_i$. It is natural that c appears here and not g , because ct and f have the same physical dimensions. We know that these equations may fail to have a unique solution, indicating the interesting possibility that there may be more than one equilibrium potential. However, to keep matters simple here, we shall assume that I is finite, that the network is connected, and that ∂D is non-empty. This is enough to ensure uniqueness of potentials. Then, by Theorem 4.2.4, the equilibrium potential is given by

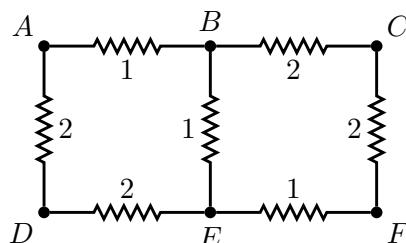
$$\phi_i = \mathbb{E}_i \left(\int_0^T c(X_t) dt + f(X_T) \right) \quad (4.10)$$

where T is the hitting time of ∂D .

In fact, the case where ∂D is empty may be dealt with as follows: we must have

$$\sum_{i \in I} g_i = 0$$

or there is no possibility of equilibrium; pick one node k , set $\partial D = \{k\}$, and replace the condition $\gamma_k = g_k$ by $\phi_k = 0$. The new problem is equivalent to the old, but now ∂D is non-empty.



Example 4.3.1

Determine the equilibrium current in the network shown on the preceding page when unit current enters at A and leaves at F . The conductivities are shown on the diagram. Let us set $\phi_A = 1$ and $\phi_F = 0$. This will result in some flow from A to F , which we can scale to get a unit flow. By symmetry, $\phi_E = 1 - \phi_B$ and $\phi_D = 1 - \phi_C$. Then, by Ohm's law, since the total current leaving B and C must vanish

$$\begin{aligned} (\phi_B - \phi_A) + (\phi_B - \phi_E) + 2(\phi_B - \phi_C) &= 0, \\ 2(\phi_C - \phi_F) + 2(\phi_C - \phi_B) &= 0. \end{aligned}$$

Hence, $\phi_B = 1/2$ and $\phi_C = 1/4$, and the associated flow is given by $\gamma_{AB} = 1/2$, $\gamma_{BC} = 1/2$, $\gamma_{CF} = 1/2$, $\gamma_{BE} = 0$. In fact, we were lucky – no scaling was necessary.

Note that the node capacities do not affect the problem we considered. Let us arbitrarily assign to each node a capacity 1. Then there is an associated Markov chain and, according to (4.10), the equilibrium potential is given by

$$\phi_i = \mathbb{E}_i(1_{X_T=A}) = \mathbb{P}_i(X_T = A)$$

where T is the hitting time of $\{A, F\}$. Different node capacities result in different Markov chains, but the same jump chain and hence the same hitting probabilities.

Here is a general result expressing equilibrium potentials, flows and charges in terms of the associated Markov chain.

Theorem 4.3.2. Consider a finite network with external connections at two nodes A and B , and the associated Markov chain $(X_t)_{t \geq 0}$.

(a) The unique equilibrium potential ϕ with $\phi_A = 1$ and $\phi_B = 0$ is given by

$$\phi_i = \mathbb{P}_i(T_A < T_B)$$

where T_A and T_B are the hitting times of A and B .

(b) The unique equilibrium flow γ with $\gamma_A = 1$ and $\gamma_B = -1$ is given by

$$\gamma_{ij} = \mathbb{E}_A(\Gamma_{ij} - \Gamma_{ji})$$

where Γ_{ij} is the number of times that $(X_t)_{t \geq 0}$ jumps from i to j before hitting B .

(c) The charge χ associated with γ , subject to $\chi_B = 0$, is given by

$$\chi_i = \mathbb{E}_A \int_0^{T_B} 1_{\{X_t=i\}} dt.$$

Proof. The formula for ϕ is a special case of (4.10), where $c = 0$ and $f = 1_{\{A\}}$. We shall prove (b) and (c) together. Observe that if $X_0 = A$ then

$$\sum_{j \neq i} (\Gamma_{ij} - \Gamma_{ji}) = \begin{cases} 1 & \text{if } i = A \\ 0 & \text{if } i \notin \{A, B\} \\ -1 & \text{if } i = B \end{cases}$$

so if $\gamma_{ij} = \mathbb{E}_A(\Gamma_{ij} - \Gamma_{ji})$ then γ is a unit flow from A to B . We have

$$\Gamma_{ij} = \sum_{n=0}^{\infty} 1_{\{Y_n=i, Y_{n+1}=j, n < N_B\}}$$

where N_B is the hitting time of B for the jump chain $(Y_n)_{n \geq 0}$. So, by the Markov property of the jump chain

$$\begin{aligned} \mathbb{E}_A(\Gamma_{ij}) &= \sum_{n=0}^{\infty} \mathbb{P}_A(Y_n = i, Y_{n+1} = j, n < N_B) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_A(Y_n = i, n < N_B) \pi_{ij}. \end{aligned}$$

Set

$$\chi_i = \mathbb{E}_A \int_0^{T_B} 1_{\{X_t=i\}} dt$$

and consider the associated potential $\psi_i = \chi_i / \pi_i$. Then

$$\chi_i q_{ij} = \chi_i q_i \pi_{ij} = \sum_{n=0}^{\infty} \mathbb{P}_A(Y_n = i, n < N_B) \pi_{ij} = \mathbb{E}_A(\Gamma_{ij})$$

so

$$(\psi_i - \psi_j) a_{ij} = \chi_i q_{ij} - \chi_j q_{ij} = \gamma_{ij}.$$

Hence $\psi = \phi$, γ is the equilibrium unit flow and χ the associated charge, as required. \square

The interpretation of potential theory in terms of electrical networks makes it natural to consider notions of *energy*. We define for a potential $\phi = (\phi_i : i \in I)$ and a flow $\gamma = (\gamma_{ij} : i, j \in I)$

$$\begin{aligned} E(\phi) &= \frac{1}{2} \sum_{i,j \in I} (\phi_i - \phi_j)^2 a_{ij}, \\ I(\gamma) &= \frac{1}{2} \sum_{i,j \in I} \gamma_{ij}^2 a_{ij}^{-1}. \end{aligned}$$

The $1/2$ means that each wire is counted once. When ϕ and γ are related by Ohm's law we have

$$E(\phi) = \frac{1}{2} \sum_{i,j \in I} (\phi_i - \phi_j) \gamma_{ij} = I(\gamma)$$

and $E(\phi)$ is found physically to give the rate of dissipation of energy, as heat, by the network. Moreover, we shall see that certain equilibrium potentials and flows determined by Ohm's law minimize these energy functions. This characteristic of energy minimization can indeed replace Ohm's law as the fundamental physical principle.

Theorem 4.3.3. *The equilibrium potential and flow may be determined as follows.*

- (a) *The equilibrium potential $\phi = (\phi_i : i \in I)$ with boundary values $\phi_i = f_i$ for $i \in \partial D$ and no current sources in D is the unique solution to*

$$\begin{aligned} & \text{minimize} && E(\phi) \\ & \text{subject to} && \phi_i = f_i \quad \text{for } i \in \partial D. \end{aligned}$$

- (b) *The equilibrium flow $\gamma = (\gamma_{ij} : i, j \in I)$ with current sources $\gamma_i = g_i$ for $i \in D$ and boundary potential zero is the unique solution to*

$$\begin{aligned} & \text{minimize} && I(\gamma) \\ & \text{subject to} && \gamma_i = g_i \quad \text{for } i \in D. \end{aligned}$$

Proof. For any potential $\phi = (\phi_i : i \in I)$ and any flow $\gamma = (\gamma_{ij} : i, j \in I)$ we have

$$\sum_{i,j \in I} (\phi_i - \phi_j) \gamma_{ij} = 2 \sum_{i \in I} \phi_i \gamma_i.$$

- (a) Denote by $\phi = (\phi_i : i \in I)$ and by $\gamma = (\gamma_{ij} : i, j \in I)$ the equilibrium potential and flow. We have $\gamma_i = 0$ for $i \in D$. We can write any potential in the minimization problem in the form $\phi + \varepsilon$, where $\varepsilon = (\varepsilon_i : i \in I)$ with $\varepsilon_i = 0$ for $i \in \partial D$. Then

$$\sum_{i,j \in I} (\varepsilon_i - \varepsilon_j)(\phi_i - \phi_j) a_{ij} = \sum_{i,j \in I} (\varepsilon_i - \varepsilon_j) \gamma_{ij} = 2 \sum_{i \in I} \varepsilon_i \gamma_i = 0$$

so

$$E(\phi + \varepsilon) = E(\phi) + E(\varepsilon) \geq E(\phi)$$

with equality only if $\varepsilon = 0$.

(b) Denote by $\phi = (\phi_i : i \in I)$ and by $\gamma = (\gamma_{ij} : i, j \in I)$ the equilibrium potential and flow. We have $\phi_i = 0$ for $i \in \partial D$. We can write any flow in the minimization problem in the form $\gamma + \delta$, where $\delta = (\delta_{ij} : i, j \in I)$ is a flow with $\delta_i = 0$ for $i \in D$. Then

$$\sum_{i,j \in I} \gamma_{ij} \delta_{ij} a_{ij}^{-1} = \sum_{i,j \in I} (\phi_i - \phi_j) \delta_{ij} = 2 \sum_{i \in I} \phi_i \delta_i = 0,$$

so

$$I(\gamma + \delta) = I(\gamma) + I(\delta) \geq I(\delta)$$

with equality only if $\delta = 0$. \square

The following reformulation of part (a) of the preceding result states that harmonic functions minimize energy.

Corollary 4.3.4. Suppose that $\phi = (\phi_i : i \in I)$ satisfies

$$\begin{cases} Q\phi = 0 & \text{in } D \\ \phi = f & \text{in } \partial D. \end{cases}$$

Then ϕ is the unique solution to

$$\begin{aligned} &\text{minimize} && E(\phi) \\ &\text{subject to} && \phi = f \quad \text{in } \partial D. \end{aligned}$$

An important feature of electrical networks is that networks with a small number of external connections look like networks with a small number of nodes altogether. In fact, given any network, there is always another network of wires joining the externally connected nodes alone, equivalent in its response to external flows and potentials.

Let $J \subseteq I$. We say that $\bar{a} = (\bar{a}_{ij} : i, j \in J)$ is an *effective conductivity* on J if, for all potentials $f = (f_i : i \in J)$, the external currents into J when J is held at potential f are the same for (J, \bar{a}) as for (I, a) . We know that f determines an equilibrium potential $\phi = (\phi_i : i \in I)$ by

$$\begin{cases} \sum_{j \in I} (\phi_i - \phi_j) a_{ij} = 0 & \text{for } i \notin J \\ \phi_i = f_i & \text{for } i \in J. \end{cases}$$

Then \bar{a} is an effective conductivity if, for all f , for $i \in J$ we have

$$\sum_{j \in I} (\phi_i - \phi_j) a_{ij} = \sum_{j \in J} (f_i - f_j) a_{ij}.$$

For a conductivity matrix \bar{a} on J , for a potential $f = (f_i : i \in J)$ and a flow $\delta = (\delta_{ij} : i, j \in J)$ we set

$$\bar{E}(f) = \frac{1}{2} \sum_{i,j \in J} (f_i - f_j)^2 \bar{a}_{ij},$$

$$\bar{I}(\delta) = \frac{1}{2} \sum_{i,j \in J} \delta_{ij}^2 \bar{a}_{ij}^{-1}.$$

Theorem 4.3.5. *There is a unique effective conductivity \bar{a} given by*

$$\bar{a}_{ij} = a_{ij} + \sum_{k \notin J} a_{ik} \phi_k^j$$

where for each $j \in J$, $\phi^j = (\phi_i^j : i \in I)$ is the potential defined by

$$\begin{cases} \sum_{k \in I} (\phi_i^j - \phi_k^j) a_{ik} = 0 & \text{for } i \notin J \\ \phi_i^j = \delta_{ij} & \text{for } i \in J. \end{cases} \quad (4.11)$$

Moreover, \bar{a} is characterized by the Dirichlet variational principle

$$\overline{E}(f) = \inf_{\phi_i = f_i \text{ on } J} E(\phi),$$

and also by the Thompson variational principle

$$\inf_{\delta_i = g_i \text{ on } J} \overline{I}(\delta) = \inf_{\gamma_i = \begin{cases} g_i \text{ on } J \\ 0 \text{ off } J \end{cases}} I(\gamma).$$

Proof. Given $f = (f_i : i \in J)$, define $\phi = (\phi_i : i \in I)$ by

$$\phi_i = \sum_{j \in J} f_j \phi_i^j$$

then ϕ is the equilibrium potential given by

$$\begin{cases} \sum_{j \in I} a_{ij} (\phi_i - \phi_j) = 0 & \text{for } i \notin J \\ \phi_i = f_i & \text{for } i \in J, \end{cases}$$

and, by Corollary 4.3.4, ϕ solves

$$\begin{aligned} & \text{minimize} && E(\phi) \\ & \text{subject to} && \phi_i = f_i \quad \text{for } i \in J. \end{aligned}$$

We have, for $i \in J$

$$\sum_{j \in I} a_{ij} \phi_j = \sum_{j \in J} a_{ij} f_j + \sum_{k \notin J} \sum_{j \in J} a_{ik} \phi_k^j f_j = \sum_{j \in J} \bar{a}_{ij} f_j.$$

In particular, taking $f \equiv 1$ we obtain

$$\sum_{j \in I} a_{ij} = \sum_{j \in J} \bar{a}_{ij}.$$

Hence we have equality of external currents:

$$\sum_{j \in I} (\phi_i - \phi_j) a_{ij} = \sum_{j \in J} (f_i - f_j) \bar{a}_{ij}.$$

Moreover, we also have equality of energies:

$$\begin{aligned} \sum_{i,j \in I} (\phi_i - \phi_j)^2 a_{ij} &= 2 \sum_{i \in I} \phi_i \sum_{j \in I} (\phi_i - \phi_j) a_{ij} \\ &= 2 \sum_{i \in J} f_i \sum_{j \in J} (f_i - f_j) \bar{a}_{ij} = \sum_{i,j \in J} (f_i - f_j)^2 \bar{a}_{ij}. \end{aligned}$$

Finally, if $g_{ij} = (f_i - f_j) \bar{a}_{ij}$ and $\gamma_{ij} = (\phi_i - \phi_j) a_{ij}$, then

$$\begin{aligned} \sum_{i,j \in I} \gamma_{ij}^2 a_{ij}^{-1} &= \sum_{i,j \in I} (\phi_i - \phi_j)^2 a_{ij} \\ &= \sum_{i,j \in J} (f_i - f_j)^2 \bar{a}_{ij} = \sum_{i,j \in J} g_{ij}^2 \bar{a}_{ij}^{-1}, \end{aligned}$$

so, by Theorem 4.3.3, for any flow $\delta = (\delta_{ij} : i, j \in I)$ with $\delta_i = g_i$ for $i \in J$ and $\delta_i = 0$ for $i \notin J$, we have

$$\sum_{i,j \in I} \delta_{ij}^2 a_{ij}^{-1} \geq \sum_{i,j \in J} g_{ij}^2 \bar{a}_{ij}^{-1}. \quad \square$$

Effective conductivity is also related to the associated Markov chain $(X_t)_{t \geq 0}$ in an interesting way. Define the *time spent in J*

$$A_t = \int_0^t 1_{\{X_s \in J\}} ds$$

and a time-changed process $(\bar{X}_t)_{t \geq 0}$ by

$$\bar{X}_t = X_{\tau(t)}$$

where

$$\tau(t) = \inf\{s \geq 0 : A_s > t\}.$$

We obtain $(\bar{X}_t)_{t \geq 0}$ by observing $(X_t)_{t \geq 0}$ whilst in J , and stopping the clock whilst $(X_t)_{t \geq 0}$ makes excursions outside J . This is really a transformation of the jump chain. By applying the strong Markov property to the jump chain we find that $(\bar{X}_t)_{t \geq 0}$ is itself a Markov chain, with jump matrix $\bar{\Pi}$ given by

$$\bar{\pi}_{ij} = \pi_{ij} + \sum_{k \notin J} \pi_{ik} \phi_k^j \quad \text{for } i, j \in J,$$

where

$$\phi_k^j = \mathbb{P}_k(X_T = j)$$

and T denotes the hitting time of J . See Example 1.4.4. Hence $(\bar{X}_t)_{t \geq 0}$ has Q -matrix given by

$$\bar{q}_{ij} = q_{ij} + \sum_{k \notin J} q_{ik} \phi_k^j.$$

Since $\phi^j = (\phi_k^j : k \in I)$ is the unique solution to (4.11), this shows that

$$\pi_i \bar{q}_{ij} = \bar{a}_{ij}$$

so $(\bar{X}_t)_{t \geq 0}$ is the Markov chain on J associated with the effective conductivity \bar{a} .

There is much more that one can say, for example in tying up the non-equilibrium behaviour of Markov chains and electrical networks. Moreover, methods coming from one theory one provide insights into the other. For an entertaining and illuminating account of the subject, you should see *Random Walks and Electrical Networks* by P. G. Doyle and J. L. Snell (Carus Mathematical Monographs 22, Mathematical Association of America, 1984).

4.4 Brownian motion

Imagine a symmetric random walk in Euclidean space which takes infinitesimal jumps with infinite frequency and you will have some idea of Brownian motion. It is named after a botanist who observed such a motion when looking at pollen grains under a microscope. The mathematical object now called Brownian motion was actually discovered by Wiener, and is also called the Wiener process.

A discrete approximation to Euclidean space \mathbb{R}^d is provided by

$$c^{-1/2} \mathbb{Z}^d = \{c^{-1/2} z : z \in \mathbb{Z}^d\}$$

where c is a large positive number. The simple symmetric random walk $(X_n)_{n \geq 0}$ on \mathbb{Z}^d is a Markov chain which is by now quite familiar. We shall show that the scaled-down and speeded-up process

$$X_t^{(c)} = c^{-1/2} X_{ct}$$

is a good approximation to Brownian motion. This provides an elementary way of thinking about Brownian motion. Also, it makes it reasonable to

suppose that some properties of the random walk carry over to Brownian motion. At the end of this section we state some results which confirm that this is true to a remarkable extent.

Why is space rescaled by the square-root of the time-scaling? Well, if we hope that $X_t^{(c)}$ converges in some sense as $c \rightarrow \infty$ to a non-degenerate limit, we will at least want $\mathbb{E}[|X_t^{(c)}|^2]$ to converge to a non-degenerate limit. For $ct \in \mathbb{Z}^+$ we have

$$\mathbb{E}[|X_{ct}|^2] = ct\mathbb{E}[|X_1|^2]$$

so the square-root scaling gives

$$\mathbb{E}[|X_t^{(c)}|^2] = t\mathbb{E}[|X_1|^2]$$

which is independent of c .

We begin by defining Brownian motion, and then show that this is not an empty definition; that is to say, Brownian motions exist.

A real-valued random variable is said to have *Gaussian distribution with mean 0 and variance t* if it has density function

$$\phi_t(x) = (2\pi t)^{-1/2} \exp\{-x^2/2t\}.$$

The fundamental role of Gaussian distributions in probability derives from the following result.

Theorem 4.4.1 (Central limit theorem). *Let X_1, X_2, \dots be a sequence of independent and identically distributed real-valued random variables with mean 0 and variance $t \in (0, \infty)$. Then, for all bounded continuous functions f , as $n \rightarrow \infty$ we have*

$$\mathbb{E}[f((X_1 + \dots + X_n)/\sqrt{n})] \rightarrow \int_{\mathbb{R}} f(x)\phi_t(x)dx.$$

We shall take this result and a few other standard properties of the Gaussian distribution for granted in this section. There are many introductory texts on probability which give the full details.

A real-valued process $(X_t)_{t \geq 0}$ is said to be continuous if

$$\mathbb{P}(\{\omega : t \mapsto X_t(\omega) \text{ is continuous}\}) = 1.$$

A continuous real-valued process $(B_t)_{t \geq 0}$ is called a *Brownian motion* if $B_0 = 0$ and for all $0 = t_0 < t_1 < \dots < t_n$ the increments

$$B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}}$$

are independent Gaussian random variables of mean 0 and variance

$$t_1 - t_0, \dots, t_n - t_{n-1}.$$

The conditions made on $(B_t)_{t \geq 0}$ are enough to determine all the probabilities associated with the process. To put it properly, the law of Brownian motion, which is a measure on the set of continuous paths, is uniquely determined. However, it is not obvious that there is any such process. We need the following result.

Theorem 4.4.2 (Wiener's theorem). *Brownian motion exists.*

Proof. For $N = 0, 1, 2, \dots$, denote by D_N the set of integer multiples of 2^{-N} in $[0, \infty)$, and denote by D the union of these sets. Let us say that $(B_t : t \in D_N)$ is a *Brownian motion indexed by D_N* if $B_0 = 0$ and for all $0 = t_0 < t_1 < \dots < t_n$ in D_N the increments

$$B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}}$$

are independent Gaussian random variables of mean 0 and variance

$$t_1 - t_0, \dots, t_n - t_{n-1}.$$

We suppose given, for each $t \in D$, an independent Gaussian random variable Y_t of mean 0 and variance 1. For $t \in D_0 = \mathbb{Z}^+$ set

$$B_t = Y_1 + Y_2 + \dots + Y_t$$

then $(B_t : t \in D_0)$ is a Brownian motion indexed by D_0 . We shall show how to extend this process successively to Brownian motions $(B_t : t \in D_N)$ indexed by D_N . Then $(B_t : t \in D)$ is a Brownian motion indexed by D . Next we shall show that $(B_t : t \in D)$ extends continuously to $t \in [0, \infty)$, and finally check that the extension is a Brownian motion.

Suppose we have constructed $(B_t : t \in D_{N-1})$, a Brownian motion indexed by D_{N-1} . For $t \in D_N \setminus D_{N-1}$ set $r = t - 2^{-N}$ and $s = t + 2^{-N}$ so that $r, s \in D_{N-1}$ and define

$$\begin{aligned} Z_t &= 2^{-(N+1)/2} Y_t, \\ B_t &= \frac{1}{2}(B_r + B_s) + Z_t. \end{aligned}$$

We obtain two new increments:

$$\begin{aligned} B_t - B_r &= \frac{1}{2}(B_s - B_r) + Z_t, \\ B_s - B_t &= \frac{1}{2}(B_s - B_r) - Z_t. \end{aligned}$$

We compute

$$\begin{aligned}\mathbb{E}[(B_t - B_r)^2] &= \mathbb{E}[(B_s - B_t)^2] = \frac{1}{4}2^{-(N-1)} + 2^{-(N+1)} = 2^{-N}, \\ \mathbb{E}[(B_t - B_r)(B_s - B_t)] &= \frac{1}{4}2^{-(N-1)} - 2^{-(N+1)} = 0.\end{aligned}$$

The two new increments, being Gaussian, are therefore independent and of the required variance. Moreover, being constructed from $B_s - B_r$ and Y_t , they are certainly independent of increments over intervals disjoint from (r, s) . Hence $(B_t : t \in D_N)$ is a Brownian motion indexed by D_N , as required. Hence, by induction, we obtain a Brownian motion $(B_t : t \in D)$.

For each N denote by $(B_t^{(N)})_{t \geq 0}$ the continuous process obtained by linear interpolation from $(B_t : t \in D_N)$. Also, set $Z_t^{(N)} = B_t^{(N)} - B_t^{(N-1)}$. For $t \in D_{N-1}$ we have $Z_t^{(N)} = 0$. For $t \in D_N \setminus D_{N-1}$, by our construction we have

$$Z_t^{(N)} = B_t - \frac{1}{2}(B_{t-2^{-N}} + B_{t+2^{-N}}) = Z_t = 2^{-(N+1)/2}Y_t$$

with Y_t Gaussian of mean 0 and variance 1. Set

$$M_N = \sup_{t \in [0, 1]} |Z_t^{(N)}|.$$

Then, since $(Z_t^{(N)})_{t \geq 0}$ interpolates linearly between its values on D_N , we obtain

$$M_N = \sup_{t \in (D_N \setminus D_{N-1}) \cap [0, 1]} 2^{-(N+1)/2}|Y_t|.$$

There are 2^{N-1} points in $(D_N \setminus D_{N-1}) \cap [0, 1]$. So for $\lambda > 0$ we have

$$\mathbb{P}(M_N > \lambda 2^{-(N+1)/2}) \leq 2^{N-1} \mathbb{P}(|Y_1| > \lambda).$$

For a random variable $X \geq 0$ and $p > 0$ we have the formula

$$\mathbb{E}(X^p) = \mathbb{E} \int_0^\infty 1_{\{X>\lambda\}} p\lambda^{p-1} d\lambda = \int_0^\infty p\lambda^{p-1} \mathbb{P}(X > \lambda) d\lambda.$$

Hence

$$\begin{aligned}2^{p(N+1)/2} \mathbb{E}(M_N^p) &= \int_0^\infty p\lambda^{p-1} \mathbb{P}(2^{(N+1)/2} M_N > \lambda) d\lambda \\ &\leq 2^{N-1} \int_0^\infty p\lambda^{p-1} \mathbb{P}(|Y_1| > \lambda) d\lambda = 2^{N-1} \mathbb{E}(|Y_1|^p)\end{aligned}$$

and hence, for any $p > 2$

$$\begin{aligned}\mathbb{E} \sum_{N=0}^{\infty} M_N &= \sum_{N=0}^{\infty} \mathbb{E}(M_N) \\ &\leq \sum_{N=0}^{\infty} \mathbb{E}(M_N^p)^{1/p} \leq \mathbb{E}(|Y_1|^p)^{1/p} \sum_{N=0}^{\infty} (2^{(p-2)/2p})^{-N} < \infty.\end{aligned}$$

It follows that, with probability 1, as $N \rightarrow \infty$

$$B_t^{(N)} = B_t^{(0)} + Z_t^{(1)} + \dots + Z_t^{(N)}$$

converges uniformly in $t \in [0, 1]$, and by a similar argument uniformly for t in any bounded interval. Now $B_t^{(N)}$ eventually equals B_t for any $t \in D$ and the uniform limit of continuous functions is continuous. Therefore, $(B_t : t \in D)$ has a continuous extension $(B_t)_{t \geq 0}$, as claimed.

It remains to show that the increments of $(B_t)_{t \geq 0}$ have the required joint distribution. But given $0 < t_1 < \dots < t_n$ we can find sequences $(t_k^m)_{m \in \mathbb{N}}$ in D such that $0 < t_1^m < \dots < t_n^m$ for all m and $t_k^m \rightarrow t_k$ for all k . Set $t_0 = t_0^m = 0$. We know that the increments

$$B_{t_1^m} - B_{t_0^m}, \dots, B_{t_n^m} - B_{t_{n-1}^m}$$

are Gaussian of mean 0 and variance

$$t_1^m - t_0^m, \dots, t_n^m - t_{n-1}^m.$$

Hence, using continuity of $(B_t)_{t \geq 0}$ we can let $m \rightarrow \infty$ to obtain the desired distribution for the increments

$$B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}}.$$

□

Having shown that Brownian motion exists, we now want to show how it appears as a universal scaling limit of random walks, very much as the Gaussian distribution does for sums of independent random variables.

Theorem 4.4.3. *Let $(X_n)_{n \geq 0}$ be a discrete-time real-valued random walk with steps of mean 0 and variance $\sigma^2 \in (0, \infty)$. For $c > 0$ consider the rescaled process*

$$X_t^{(c)} = c^{-1/2} X_{ct}$$

where the value of X_{ct} when ct is not an integer is found by linear interpolation. Then, for all m , for all bounded continuous functions $f : \mathbb{R}^m \rightarrow \mathbb{R}$ and all $0 \leq t_1 < \dots < t_m$, we have

$$\mathbb{E}[f(X_{t_1}^{(c)}, \dots, X_{t_m}^{(c)})] \rightarrow \mathbb{E}[f(\sigma B_{t_1}, \dots, \sigma B_{t_m})]$$

as $c \rightarrow \infty$, where $(B_t)_{t \geq 0}$ is a Brownian motion.

Proof. The claim is that $(X_{t_1}^{(c)}, \dots, X_{t_m}^{(c)})$ converges weakly to $(\sigma B_{t_1}, \dots, \sigma B_{t_m})$ as $c \rightarrow \infty$. In the proof we shall take for granted some basic properties of weak convergence. First define

$$\tilde{X}_t^{(c)} = c^{-1/2} X_{[ct]}$$

where $[ct]$ denotes the integer part of ct . Then

$$|(X_{t_1}^{(c)}, \dots, X_{t_m}^{(c)}) - (\tilde{X}_{t_1}^{(c)}, \dots, \tilde{X}_{t_m}^{(c)})| \leq c^{-1/2} |(Y_{[ct_1]+1}, \dots, Y_{[ct_n]+1})|$$

where Y_n denotes the n th step of $(X_n)_{n \geq 0}$. The right side converges weakly to 0, so it suffices to prove the claim with $\tilde{X}_t^{(c)}$ replacing $X_t^{(c)}$.

Consider now the increments

$$U_k^{(c)} = \tilde{X}_{t_k}^{(c)} - \tilde{X}_{t_{k-1}}^{(c)}, \quad Z_k = \sigma(B_{t_k} - B_{t_{k-1}})$$

for $k = 1, \dots, m$. Since $\tilde{X}_0^{(c)} = B_0 = 0$ it suffices to show that $(U_1^{(c)}, \dots, U_m^{(c)})$ converges weakly to (Z_1, \dots, Z_m) . Then since both sets of increments are independent, it suffices to show that $U_k^{(c)}$ converges weakly to Z_k for each k . But

$$U_k^{(c)} = c^{-1/2} \sum_{n=[ct_{k-1}]+1}^{[ct_k]} Y_n \sim (c^{-1/2} N_k(c)^{1/2}) N_k(c)^{-1/2} (Y_1 + \dots + Y_{N(c)})$$

where \sim denotes identity of distribution and $N_k(c) = [ct_k] - [ct_{k-1}]$. By the central limit theorem $N_k(c)^{-1/2} (Y_1 + \dots + Y_{N(c)})$ converges weakly to $(t_k - t_{k-1})^{-1/2} Z_k$, and $(c^{-1/2} N_k(c)^{1/2}) \rightarrow (t_k - t_{k-1})^{1/2}$. Hence $U_k^{(c)}$ converges weakly to Z_k , as required. \square

To summarize the last two results, we have shown, using special properties of the Gaussian distribution, that there is a continuous process $(B_t)_{t \geq 0}$ with stationary independent increments and such that B_t is Gaussian of mean 0 and variance t , for each $t \geq 0$. That was Wiener's theorem. Then, using the central limit theorem applied to the increments of a rescaled random walk, we established a sort of convergence to Brownian motion. There now follows a series of related remarks.

Note the similarity to the definition of a *Poisson process* as a right-continuous integer-valued process $(X_t)_{t \geq 0}$ starting from 0, having stationary independent increments and such that X_t is Poisson of parameter λt for each $t \geq 0$.

Given d independent Brownian motions $(B_t^1)_{t \geq 0}, \dots, (B_t^d)_{t \geq 0}$, let us consider the \mathbb{R}^d -valued process $B_t = (B_t^1, \dots, B_t^d)$. We call $(B_t)_{t \geq 0}$ a *Brownian*

motion in \mathbb{R}^d . There is a multidimensional version of the central limit theorem which leads to a multidimensional version of Theorem 4.4.3, with no essential change in the proof. Thus if $(X_n)_{n \geq 0}$ is a random walk in \mathbb{R}^d with steps of mean 0 and covariance matrix

$$V = \mathbb{E}(X_1 X_1^T)$$

and if V is finite, then for all bounded continuous functions $f : (\mathbb{R}^d)^m \rightarrow \mathbb{R}$, as $c \rightarrow \infty$ we have

$$\mathbb{E}[f(X_{t_1}^{(c)}, \dots, X_{t_m}^{(c)})] \rightarrow \mathbb{E}[f(\sqrt{V}B_{t_1}, \dots, \sqrt{V}B_{t_m})].$$

Here are two examples. We might take $(X_n)_{n \geq 0}$ to be the simple symmetric random walk in \mathbb{Z}^3 , then $V = \frac{1}{3}I$. Alternatively, we might take the components of $(X_n)_{n \geq 0}$ to be three independent simple symmetric random walks in \mathbb{Z} , in which case $V = I$. Although these are different random walks, once the difference in variance is taken out, the result shows that in the scaling limit they behave asymptotically the same. More generally, given a random walk with a complicated step distribution, it is useful to know that on large scales all one needs to calculate is the variance (or covariance matrix). All other aspects of the step distribution become irrelevant as $c \rightarrow \infty$.

The scaling used in Theorem 4.4.3 suggests the following *scaling invariance property* of Brownian motion $(B_t)_{t \geq 0}$, which is also easy to check from the definition. For any $c > 0$ the process $(B_t^{(c)})_{t \geq 0}$ defined by

$$B_t^{(c)} = c^{-1/2}B_{ct}$$

is a Brownian motion. Thus Brownian motion appears as a fixed point of the scaling transformation, which attracts all other finite variance symmetric random walks as $c \rightarrow \infty$.

The sense in which we have shown that $(X_t^{(c)})_{t \geq 0}$ converges to Brownian motion is very weak, and one can with effort prove stronger forms of convergence. However, what we have proved is strong enough to ensure that $(X_t^{(c)})_{t \geq 0}$ does not converge, in the same sense, to anything else.

The discussion to this point has not really been about the Markov property, but rather about processes with independent increments. To remedy this we must first define *Brownian motion starting from x* : this is simply any process $(B_t)_{t \geq 0}$ such that $B_0 = x$ and $(B_t - B_0)_{t \geq 0}$ is a Brownian motion (starting from 0). As a limit of Markov chains it is natural to look in Brownian motion for the structure of a Markov process. By analogy with continuous-time Markov chains we look for a *transition semigroup* $(P_t)_{t \geq 0}$

and a *generator* G . For any bounded measurable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ we have

$$\begin{aligned}\mathbb{E}_x[f(B_t)] &= \mathbb{E}_0[f(x + B_t)] = \int_{\mathbb{R}^d} f(x + y) \phi_t(y_1) \dots \phi_t(y_d) dy_1 \dots dy_d \\ &= \int_{\mathbb{R}^d} p(t, x, y) f(y) dy\end{aligned}$$

where

$$p(t, x, y) = (2\pi t)^{-d/2} \exp\{-|y - x|^2/2t\}.$$

This is the *transition density* for Brownian motion and the *transition semigroup* is given by

$$(P_t f)(x) = \int_{\mathbb{R}^d} p(t, x, y) f(y) dy = \mathbb{E}_x[f(B_t)].$$

To check the semigroup property $P_s P_t = P_{s+t}$ we note that

$$\begin{aligned}\mathbb{E}_x[f(B_{s+t})] &= \mathbb{E}_x[f(B_s + (B_{s+t} - B_s))] \\ &= \mathbb{E}_x[P_t f(B_s)] = (P_s P_t f)(x)\end{aligned}$$

where we first took the expectation over the independent increment $B_{s+t} - B_s$. For $t > 0$ it is easy to check that

$$\frac{\partial}{\partial t} p(t, x, y) = \frac{1}{2} \Delta_x p(t, x, y)$$

where

$$\Delta_x = \frac{\partial^2}{\partial x_1^2} + \dots + \frac{\partial^2}{\partial x_d^2}.$$

Hence, if f has two bounded derivatives, we have

$$\begin{aligned}\frac{\partial}{\partial t} (P_t f)(x) &= \int_{\mathbb{R}^d} \frac{1}{2} \Delta_x p(t, x, y) f(y) dy \\ &= \int_{\mathbb{R}^d} \frac{1}{2} \Delta_y p(t, x, y) f(y) dy \\ &= \int_{\mathbb{R}^d} p(t, x, y) (\frac{1}{2} \Delta f)(y) dy \\ &= \mathbb{E}_x[(\frac{1}{2} \Delta f)(B_t)] \rightarrow \frac{1}{2} \Delta f(x)\end{aligned}$$

as $t \downarrow 0$. This suggests, by analogy with continuous-time chains, that the generator, a term we have not defined precisely, should be given by

$$G = \frac{1}{2} \Delta.$$

Where formerly we considered vectors $(f_i : i \in I)$, now there are functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$, required to have various degrees of local regularity, such as measurability and differentiability. Where formerly we considered matrices P_t and Q , now we have linear operators on functions: P_t is an integral operator, G is a differential operator.

We would like to explain the appearance of the Laplacian Δ by reference to the random walk approximation. Denote by $(X_n)_{n \geq 0}$ the simple symmetric random walk in \mathbb{Z}^d and consider for $N = 1, 2, \dots$ the rescaled process

$$X_t^{(N)} = N^{-1/2} X_{Nt}, \quad t = 0, 1/N, 2/N, \dots$$

For a bounded continuous function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, set

$$(P_t^{(N)} f)(x) = \mathbb{E}_x[f(X_t^{(N)})], \quad x \in N^{-1/2} \mathbb{Z}^d.$$

The closest thing we have to a derivative in t at 0 for $(P_t^{(N)})_{t=0,1/N,2/N,\dots}$ is

$$\begin{aligned} N(P_{1/N}^{(N)} f - f)(x) &= N\mathbb{E}_x[f(X_{1/N}^{(N)}) - f(X_0^{(N)})] \\ &= N\mathbb{E}_{N^{1/2}x}[f(N^{-1/2}X_1) - f(N^{-1/2}X_0)] \\ &= (N/2)\{f(x - N^{-1/2}) - 2f(x) + f(x + N^{-1/2})\}. \end{aligned}$$

If we assume that f has two bounded derivatives then, by Taylor's theorem, as $N \rightarrow \infty$,

$$f(x - N^{-1/2}) - 2f(x) + f(x + N^{-1/2}) = N^{-1}(\Delta f(x) + o(N)),$$

so

$$N(P_{1/N}^{(N)} f - f)(x) \rightarrow \frac{1}{2}\Delta f(x).$$

We finish by stating some results about Brownian motion which emphasise how much of the structure of Markov chains carries over. You will notice some weasel words creeping in, such as measurable, continuous and differentiable. These are various sorts of local regularity for functions defined on the state-space \mathbb{R}^d . They did not appear for Markov chains because a discrete state-space has no local structure. You might correctly guess that the proofs would require additional real analysis, relative to the corresponding results for chains, and a proper measure-theoretic basis for the probability. But, this aside, the main ideas are very similar. For further details see, for example, *Probability Theory – an analytic view* by D. W. Stroock (Cambridge University Press, 1993), or *Diffusions, Markov Processes and Martingales, Volume 1: Foundations* by L. C. G. Rogers and David Williams (Wiley, Chichester, 2nd edition 1994).

First, here is a result on recurrence and transience.

Theorem 4.4.4. Let $(B_t)_{t \geq 0}$ be a Brownian motion in \mathbb{R}^d .

(i) If $d = 1$, then

$$\mathbb{P}(\{t \geq 0 : B_t = 0\} \text{ is unbounded}) = 1.$$

(ii) If $d = 2$, then

$$\mathbb{P}(B_t = 0 \text{ for some } t > 0) = 0$$

but, for any $\varepsilon > 0$

$$\mathbb{P}(\{t \geq 0 : |B_t| < \varepsilon\} \text{ is unbounded}) = 1.$$

(iii) If $d = 3$, then

$$\mathbb{P}(|B_t| \rightarrow \infty \text{ as } t \rightarrow \infty) = 1.$$

It is natural to compare this result with the facts proved in Section 1.6, that in \mathbb{Z} and \mathbb{Z}^2 the simple symmetric random walk is recurrent, whereas in \mathbb{Z}^3 it is transient. The results correspond exactly in dimensions one and three. In dimension two we see the fact that for continuous state-space it makes a difference to demand returns to a point or to arbitrarily small neighbourhoods of a point. If we accept this latter notion of recurrence the correspondence extends to dimension two.

The invariant measure for Brownian motion is Lebesgue measure dx . This has infinite total mass so in dimensions one and two Brownian motion is only null recurrent. So that we can state some results for the positive recurrent case, we shall consider Brownian motion in \mathbb{R}^d projected onto the torus $T^d = \mathbb{R}^d / \mathbb{Z}^d$. In dimension one this just means wrapping the line round a circle of circumference 1. The invariant measure remains Lebesgue measure but this now has total mass 1. So the projected process is positive recurrent and we can expect convergence to equilibrium and ergodic results corresponding to Theorems 1.8.3 and 1.10.2.

Theorem 4.4.5. Let $(B_t)_{t \geq 0}$ be a Brownian motion in \mathbb{R}^d and let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a continuous periodic function, so that

$$f(x + z) = f(x) \text{ for all } z \in \mathbb{Z}^d.$$

Then for all $x \in \mathbb{R}^d$, as $t \rightarrow \infty$, we have

$$\mathbb{E}_x[f(B_t)] \rightarrow \bar{f} = \int_{[0,1]^d} f(z) dz$$

and, moreover

$$\mathbb{P}_x \left(\frac{1}{t} \int_0^t f(B_s) ds \rightarrow \bar{f} \text{ as } t \rightarrow \infty \right) = 1.$$

The generator $\frac{1}{2}\Delta$ of Brownian motion in \mathbb{R}^d reappears as it should in the following martingale characterization of Brownian motion.

Theorem 4.4.6. Let $(X_t)_{t \geq 0}$ be a continuous \mathbb{R}^d -valued random process. Write $(\mathcal{F}_t)_{t \geq 0}$ for the filtration of $(X_t)_{t \geq 0}$. Then the following are equivalent:

- (i) $(X_t)_{t \geq 0}$ is a Brownian motion;
- (ii) for all bounded functions f which are twice differentiable with bounded second derivative, the following process is a martingale:

$$M_t^f = f(X_t) - f(X_0) - \frac{1}{2} \int_0^t \Delta f(X_s) ds.$$

This result obviously corresponds to Theorem 4.1.2. In case you are unsure, a continuous time process $(M_t)_{t \geq 0}$ is a martingale if it is adapted to the given filtration $(\mathcal{F}_t)_{t \geq 0}$, if $\mathbb{E}|M_t| < \infty$ for all t , and

$$\mathbb{E}[(M_t - M_s)1_A] = 0$$

whenever $s \leq t$ and $A \in \mathcal{F}_s$.

We end with a result on the potentials associated with Brownian motion, corresponding very closely to Theorem 4.2.3 for Markov chains. These potentials are identical to those appearing in Newton's theory of gravity, as we remarked in Section 4.2.

Theorem 4.4.7. Let D be an open set in \mathbb{R}^d with smooth boundary ∂D . Let $c : D \rightarrow [0, \infty)$ be measurable and let $f : \partial D \rightarrow [0, \infty)$ be continuous. Set

$$\phi(x) = \mathbb{E}_x \left[\int_0^T c(B_t) dt + f(X_T) 1_{T < \infty} \right]$$

where T is the hitting time of ∂D . Then

- (i) ϕ if finite belongs to $C^2(D) \cap C(\overline{D})$ and satisfies

$$\begin{cases} -\frac{1}{2}\Delta\phi = c & \text{in } D \\ \phi = f & \text{in } \partial D; \end{cases} \quad (4.12)$$

- (ii) if $\psi \in C^2(D) \cap C(\overline{D})$ and satisfies

$$\begin{cases} -\frac{1}{2}\Delta\psi \geq c & \text{in } D \\ \psi \geq f & \text{in } \partial D \end{cases}$$

and $\psi \geq 0$, then $\psi \geq \phi$;

- (iii) if $\phi(x) = \infty$ for some x , then (4.12) has no finite solution;
- (iv) if $\mathbb{P}_x(T < \infty) = 1$ for all x , then (4.12) has at most one bounded solution in $C^2(D) \cap C(\overline{D})$.