Chapter 3
Reversible Markov Chains

David Aldous
Department of Statistics
University of California
Berkeley, CA 94720

James Allen Fill
Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, MD 21218-2692

September 2, 1994

Chapter 2 reviewed some aspects of the elementary theory of general finite irreducible Markov chains. In this chapter we specialize to reversible chains, treating the discrete and continuous-time cases in parallel. After section 3 we shall assume that we are dealing with reversible chains without continually repeating this assumption, and shall instead explicitly say "general" to mean not-necessarily reversible.

1 Introduction

Recall $\mathbf{P}$ denotes the transition matrix and π the stationary distribution of a finite irreducible discrete-time chain X_t. Call the chain *reversible* if

$$\pi_i p_{ij} = \pi_j p_{ji} \text{ for all } i, j. \tag{1}$$

Equivalently, suppose (for given irreducible $\mathbf{P}$) that π is a probability distribution satisfying (1). Then π is the unique stationary distribution and the chain is reversible. This is true because (1), sometimes called the *detailed balance equations*, implies

$$\sum_i \pi_i p_{ij} = \pi_j \sum_i p_{ji} = \pi_j \quad \text{for all } j$$

and therefore π satisfies the *balance equations* of Chapter 2 yyy.

The name *reversible* comes from the following fact. If (X_t) is the stationary chain, that is if X_0 has distribution π, then

$$(X_0, X_1, \ldots, X_t) \stackrel{d}{=} (X_t, X_{t-1}, \ldots, X_0).$$

More vividly, given a movie of the chain run forwards and the same movie run backwards, you cannot tell which is which.

It is elementary that the same symmetry property (1) holds for the t-step transition matrix $\mathbf{P}^t$

$$\pi_i p^{(t)}_{i,j} = \pi_j p^{(t)}_{ji}$$

and thence for the matrix $\mathbf{Z}$ of Chapter 2 section yyy

$$\pi_i Z_{ij} = \pi_j Z_{ji}. \tag{2}$$

But beware that the symmetry property does not work for mean hitting times: the assertion

$$\pi_i E_i T_j = \pi_j E_j T_i$$

is definitely *false* in general. See Chapter 7 for further discussion.

The discussion above extends to continuous time with only notational changes, e.g. the detailed balance equation (1) becomes

$$\pi_i q_{ij} = \pi_j q_{ji} \text{ for all } i, j. \tag{3}$$

1.1 Time-reversals and cat-and-mouse games

For a general chain we can define the *time-reversed* chain to have transition matrix $\mathbf{P}^*$ where

$$\pi_i p_{ij} = \pi_j p^*_{ji}$$

so that the chain is reversible iff $\mathbf{P}^* = \mathbf{P}$. One can check (cf. (2))

$$\pi_i Z_{ij} = \pi_j Z^*_{ji}. \tag{4}$$

The stationary $\mathbf{P}^*$-chain is just the stationary $\mathbf{P}$-chain run backwards in time. Consider Examples yyy and yyy from Chapter 2. In Example yyy (patterns in coin tossing) the time-reversal $\mathbf{P}^*$ just "shifts left" instead of shifting right, i.e. from $HTTTT$ the possible transitions are to $HHTTT$ and $THTTT$. In Example yyy the time-reversal just reverses the direction of motion around the n-cycle:

$$p^*_{ij} = a 1_{(j=i-1)} \ + \ \frac{1-a}{n}.$$

Warning. These examples are simple because the stationary distributions are uniform. If the stationary distribution has no simple form then typically $\mathbf{P}^*$ will have no simple form.

A few facts about reversible chains are really specializations of facts about general chains which involve both $\mathbf{P}$ and $\mathbf{P}^*$. Here is a simple instance.

Lemma 1 (The cyclic tour property) *For states $i_0, i_1, \ldots, i_m$ of a reversible chain,*

$$E_{i_0}T_{i_1} + E_{i_1}T_{i_2} + \ldots E_{i_m}T_{i_0} = E_{i_0}T_{i_m} + E_{i_m}T_{i_{m-1}} + \ldots E_{i_1}T_{i_0}.$$

The explanation is that in a general chain we have

$$E_{i_0}T_{i_1} + E_{i_1}T_{i_2} + \ldots E_{i_m}T_{i_0} = E^*_{i_0}T_{i_m} + E^*_{i_m}T_{i_{m-1}} + \ldots E^*_{i_1}T_{i_0} \quad (5)$$

where E^* refers to the time-reversed chain $\mathbf{P}^*$. Equality (5) is intuitively obvious when we visualize running a movie backwards. But saying that argument precisely requires a little sophistication (see Notes). It is however straightforward to *verify* (5) using (4) and the mean hitting time formula $E_iT_j = (Z_{jj} - Z_{ij})/\pi_j$.

We shall encounter several results which have amusing interpretations as cat-and-mouse games. The common feature of these games is that the cat moves according to a transition matrix $\mathbf{P}$ and the mouse moves according to the time-reversed transition matrix $\mathbf{P}^*$.

Cat-and-mouse game 1. Both animals are placed at the same state, chosen according to the stationary distribution. The mouse makes a jump according to $\mathbf{P}^*$, and then stops. The cat starts moving according to $\mathbf{P}$ and continues until it finds the mouse, after M steps.

The notable feature of this game is the simple formula for EM:

$$EM = n - 1, \text{ where } n \text{ is the number of states.} \quad (6)$$

This is simple once you see the right picture. Consider the stationary $\mathbf{P}$-chain $(X_0, X_1, X_2, \ldots)$. We can specify the game in terms of that chain by taking the initial state to be X_1 and the mouse's jump to be to X_0, and the cat's moves to be to $X_2, X_3, \ldots$. So $M = T^+ - 1$ with

$$T^+ \equiv \min\{t \geq 1 : X_t = X_0\}.$$

And $ET^+ = \sum_i \pi_i E_i T_i^+ = \sum_i \pi_i \frac{1}{\pi_i} = n$.

Cat-and-mouse game 2. This game, and Proposition 2, are rephrasings of results of Coppersmith et al [14]. Think of the cat and the mouse as pieces in a solitaire board game. The player sees their positions and chooses which one to move: if the cat is chosen, it makes a move according to $\mathbf{P}$, and if the mouse is chosen, it makes a move according to $\mathbf{P}^*$. Let M denote the number of moves until the cat and mouse meet. Then one expects the mean $E_{(x,y)}M$ to depend on the initial positions (x, y) of (cat, mouse) and on the player's strategy. But consider the example of asymmetric random walk on a n-cycle, with (say) chance 2/3 of moving clockwise and chance 1/3 of moving counterclockwise. A moment's thought reveals that the distance (measured clockwise from cat to mouse) between the animals does not depend on the player's strategy, and hence neither does $E_{(x,y)}M$. In general EM does depend on the strategy, but the following result implies that the size of the effect of strategy changes can be bounded in terms of a measure of non-symmetry of the chain.

Proposition 2 *Regardless of strategy,*

$$\min_z E_\pi T_z \leq E_{(x,y)}M - (E_x T_y - E_\pi T_y) \leq \max_z E_\pi T_z$$

where hitting times T refer to the $\mathbf{P}$-chain.

Proof. Consider the functions

$$f(x,y) \equiv E_x T_y - E_\pi T_y$$

$$f^*(y,x) \equiv E_y T^*_x - E_\pi T^*_x.$$

The first-step recurrences for $x \to E_x T_y$ and $y \to E_y T^*_x$ give

$$f(x,y) = 1 + \sum_z p_{xz} f(z,y), \ y \neq x \quad (7)$$

$$f^*(y,x) = 1 + \sum_z p^*_{yz} f^*(z,x), \ y \neq x. \quad (8)$$

By the mean hitting time formula

$$f(x,y) = \frac{-Z_{xy}}{\pi_y} = \frac{-Z^*_{yx}}{\pi_x} = f^*(y,x)$$

so we may rewrite (8) as

$$f(x,y) = 1 + \sum_z p^*_{yz} f(x,z), \ y \neq x. \quad (9)$$

Now let $(\hat{X}_t, \hat{Y}_t)$ be the positions of (cat, mouse) after t moves according to some strategy. Consider

$$W_t \equiv t + f(\hat{X}_t, \hat{Y}_t).$$

Equalities (7,9) are exactly what is needed to verify

$$(W_t; 0 \leq t \leq M) \text{ is a martingale .}$$

So the optional stopping theorem says $E_{(x,y)}W_0 = E_{(x,y)}W_M$, that is

$$f(x,y) = E_{(x,y)}M + E_{(x,y)}f(\hat{X}_M, \hat{Y}_M). \tag{10}$$

But $\hat{X}_M = \hat{Y}_M$ and $-f(z,z) = E_\pi T_z$, so

$$\min_z E_\pi T_z \leq -f(\hat{X}_M, \hat{Y}_M) \leq \max_z E_\pi T_z$$

and the result follows from (10).

Remarks. Symmetry conditions in the reversible setting are discussed in Chapter 7. Vertex-transitivity forces $E_\pi T_z$ to be independent of z, and hence in the present setting implies $E_{(x,y)}M = E_x T_y$ regardless of strategy. For a reversible chain without this symmetry condition, consider (x_0, y_0) attaining the *min* and *max* of $E_\pi T_z$. The Proposition then implies $E_{y_0}T_{x_0} \leq E_{(x_0,y_0)}M \leq E_{x_0}T_{y_0}$ and the bounds are attained by keeping one animal fixed. But for general initial states the bounds of the Proposition are not attained. Indeed, the proof shows that to attain the bounds we need a strategy which forces the animals to meet at states attaining the extrema of $E_\pi T_z$. Finally, in the setting of random walk on a n-vertex graph we can combine Proposition 2 with mean hitting time bounds from Chapter 6 to show that EM is at worst $O(n^3)$.

2 Reversible chains and weighted graphs

Our convention is that a *graph* has finite vertex set $\mathcal{V} = \{v, x, y, \ldots\}$ and edge-set $\mathcal{E} = \{e_1, e_2, \ldots\}$, is connected, undirected, has no multiple edges and has no self-loops. A *weighted graph* also has a weight $0 < w_{v,x} = w_{x,v} < \infty$ attached to each edge (v,x), and we allow a weighted graph to have self-loops.

Given a weighted graph, there is a natural definition of a Markov chain on the vertices. This requires an arbitrary choice of convention: do we want to regard an absent edge as having weight 0 or weight $+\infty$? In terms of electrical networks (section 3) the question is whether to regard weights as conductances or as resistances of wires. Conceptually one can make good arguments for either choice, but formulas look simpler with the conductance convention (absent edges have weight 0), so we'll adopt that convention. Define discrete-time *random walk* on a weighted graph to be the Markov chain with transition matrix

$$p_{vx} = w_{vx}/w_v; \ x \neq v \tag{11}$$

where

$$w_v = \sum_x w_{vx}, \ w = \sum_v w_v.$$

Note that w is the total edge-weight, when each edge is counted *twice*, i.e. once in each direction. The fundamental fact is that this chain is automatically reversible with stationary distribution

$$\pi_v = w_v/w \tag{12}$$

because (1) is obviously satisfied by $\pi_v p_{vx} = \pi_x p_{xv} = w_{vx}/w$. Our standing convention that graphs be connected implies that the chain is irreducible. Conversely, with our standing convention that chains be irreducible, any reversible chain can be regarded as as random walk on the weighted graph with edge-weights $w_{vx} = \pi_v p_{vx}$. Note also that the "aperiodic" condition for a Markov chain (occurring in the convergence theorem Chapter 2 yyy) is just the condition that the graph be not bipartite.

An unweighted graph can be fitted into this setup by simply assigning weight 1 to each edge. Since we'll be talking a lot about this case, let's write out the specialization explicitly. The transition matrix becomes

$$\begin{aligned} p_{vx} &= 1/d_v \text{ if } (v,x) \text{ is an edge} \\ &= 0 \text{ if not} \end{aligned}$$

where d_v is the degree of vertex v. The stationary distribution becomes

$$\pi_v = \frac{d_v}{2|\mathcal{E}|} \tag{13}$$

where $|\mathcal{E}|$ is the number of edges of the graph. In particular, on an unweighted *regular* graph the stationary distribution is uniform.

In continuous time there are two *different* ways to associate a walk with a weighted or unweighted graph. One way (and we use this way unless otherwise mentioned) is just to use (11) as the definition of the transition rates q_{vx}. In the language of Chapter 2 this is the continuization of the discrete-time walk, and has the same stationary distribution and mean hitting times as the discrete-time walk. The alternative definition, which we call the *fluid model*, uses the weights directly as transition rates:

$$q_{vx} = w_{vx}, \ x \neq v. \tag{14}$$

In this model the stationary distribution is always uniform (cf. section 2.1). In the case of an unweighted *regular* graph the two models are identical up to a deterministic time rescaling, but for non-regular graphs there are typically no exact relations between numerical quantities for the two continuous-time models. Note that, given an arbitrary continuous-time reversible chain, we can define edge-weights (w_{ij}) via

$$\pi_i q_{ij} = \pi_j q_{ji} = w_{ij}, \text{ say}$$

but the weights (w_{ij}) do not completely determine the chain: we can specify the π_i independently and then solve for the q's.

Though there's no point in writing out all the specializations of the general theory of Chapter 2, let us emphasize the simple expressions for mean return times of discrete-time walk obtained from Lemma yyy and the expressions (12,13) for the stationary distribution.

Lemma 3 *For random walk on an n-vertex graph*

$$\begin{aligned} E_v T_v^+ &= \frac{w}{w_v} \quad \textit{(weighted)} \\ &= \frac{2|\mathcal{E}|}{d_v} \quad \textit{(unweighted)} \\ &= n \quad \textit{(unweighted regular).} \end{aligned}$$

The following cute variation is sometimes useful. Given the discrete-time random walk (X_t), consider the process

$$Z_t = (X_{t-1}, X_t)$$

recording the present position at time t and also the previous position. Clearly Z_t is a Markov chain whose state-space is the set $\vec{\mathcal{E}}$ of directed edges, and its stationary distribution (ρ, say) is

$$\rho(v,x) = \frac{w_{vx}}{w}$$

in the general weighted case, and hence

$$\rho(v,x) = \frac{1}{|\ \vec{\mathcal{E}}\ |}, \quad (x,v) \in \vec{\mathcal{E}}$$

in the unweighted case. Now given an edge (x, v), we can apply Lemma yyy of Chapter 2 to (Z_t) and the state (x, v) to deduce the following.

Lemma 4 *Given an edge (v, x) define*

$$U = \min\{t \geq 1 : X_t = v, X_{t-1} = x\}.$$

Then

$$\begin{aligned} E_v U &= \frac{w}{w_{vx}} \quad \textit{(weighted)} \\ &= 2|\mathcal{E}| \quad \textit{(unweighted)} . \end{aligned}$$

Corollary 5 (The edge-commute inequality) *For an edge (v, x),*

$$\begin{aligned} E_v T_x + E_x T_v &\leq \frac{w}{w_{vx}} \quad \textit{(weighted)} \\ &\leq 2|\mathcal{E}| \quad \textit{(unweighted)} \end{aligned}$$

We shall soon see (section 3.3) this inequality has a natural interpretation in terms of electrical resistance, but it is worth remembering that the result is more elementary than that.

Here is another variant of Lemma 3.

Lemma 6 *For random walk on a weighted n-vertex graph,*

$$\sum_{e=(v,x)} w_e(E_v T_x + E_x T_v) = w(n-1).$$

where the sum is over undirected edges.

Proof. Writing $\sum_v \sum_x$ for the sum over directed edges (v, x), the left side

$$\begin{aligned} &= \frac{1}{2} \sum_v \sum_x w_{vx}(E_v T_x + E_x T_v) \\ &= \sum_v \sum_x w_{vx} E_x T_v \text{ by symmetry} \\ &= w \sum_v \pi_v \sum_x p_{vx} E_x T_v \\ &= w \sum_v \pi_v (E_v T_v^+ - 1) \\ &= w \sum_v \pi_v (\frac{1}{\pi_v} - 1) \\ &= w(n-1). \end{aligned}$$

2.1 The fluid model

Imagine a finite number of identical buckets which can hold unit quantity of fluid. Some pairs of buckets are connected by tubes through their bottoms. If a tube connects buckets i and j then, when the quantities of fluid in buckets i and j are p_i and p_j, the flow rate through the tube should be proportional to the pressure difference and hence should be $w_{ij}(p_i - p_j)$ in the direction $i \to j$, where $w_{ij} = w_{ji}$ is a parameter. Neglecting the fluid in the tubes, the quantities of fluid $(p_i(t))$ at time t will evolve according to the differential equations

$$\frac{dp_j(t)}{dt} = \sum_{i \neq j} w_{ij}(p_i(t) - p_j(t)).$$

These of course are the same equations as the *forward equations* (Chapter 2 yyy) for $p_i(t)$ (the probability of being in state i at time t) for the continuous-time chain with transition rates $q_{ij} = w_{ij}, j \neq i$. Hence we call this particular way of defining a continuous-time chain in terms of a weighted graph the *fluid model.* Our main purpose in mentioning this notion is to distinguish it from the electrical network analogy in the next section. Our intuition about fluids says that as $t \to \infty$ the fluid will distribute itself uniformly amongst buckets, which corresponds to the elementary fact that the stationary distribution of the "fluid model" chain is always uniform. Our intuition also says that increasing a "specific flow rate" parameter w_{ij} will make the fluid settle faster, and this corresponds to a true fact about the "fluid model" Markov chain (in terms of the eigenvalue interpretation of asymptotic convergence rate – see Corollary 25). On the other hand the same assertion for the usual discrete-time chain or its continuization is just false.

3 Electrical networks

3.1 Flows

This is a convenient place to record some definitions. A *flow* $\mathbf{f} = (f_{ij})$ on a graph is required only to satisfy the conditions

$$\begin{aligned} f_{ij} &= -f_{ji} \text{ if } (i,j) \text{ is an edge} \\ &= 0 \text{ if not .} \end{aligned} \qquad (15)$$

So the net flow out of i is $f_{(i)} \equiv \sum_{j \neq i} f_{ij}$, and by symmetry $\sum_i f_{(i)} = 0$. We will be concerned with flows satisfying extra conditions. Given disjoint non-empty subsets A, B of vertices, a *unit flow from B to A* is a flow satisfying

$$\sum_{i \in B} f_{(i)} = 1, \qquad f_{(j)} = 0 \text{ for all } j \notin A \cup B \tag{16}$$

which implies $\sum_{i \in A} f_{(i)} = -1$. Given a probability distribution ρ on vertices, a *unit flow from i_0 to ρ* is a flow satisfying

$$f_{(i)} = 1_{(i=i_0)} - \rho_i \text{ for all } i. \tag{17}$$

Given a Markov chain X (in particular, given a weighted graph we can use the random walk) we can define two special flows as follows. Given $v_0 \notin A$, define $\mathbf{f}^{v_0 \to A}$ by

$$f_{ij} = E_{v_0} \sum_{t=1}^{T_A} \left(1_{(X_{t-1}=i, X_t=j)} - 1_{(X_{t-1}=j, X_t=i)} \right). \tag{18}$$

So f_{ij} is the mean number of transitions $i \to j$ minus the mean number of transitions $j \to i$, for the chain started at v_0 and run until hitting A. Clearly $\mathbf{f}^{v_0 \to A}$ is a unit flow from v_0 to A. In section 7.2 we will use a notion of "a unit flow from v_0 to the stationary distribution".

3.2 The analogy

Since I cannot improve on existing textbook treatments (see Notes) of the basic analogy between random walks on weighted graphs and electrical networks, I shall be brief. Take a weighted graph and fix a vertex v_0 and a subset A of vertices not containing v_0. Consider random walk on the weighted graph. In terms of the random walk we can define the flow $\mathbf{f}^{v_0 \to A}$ as at (18), and we can define a function g on vertices as follows.

$$g(v) = P_v(T_{v_0} < T_A) \tag{19}$$

so that

$$g(v_0) = 1; \; g = 0 \text{ on } A; \; 0 \leq g \leq 1. \tag{20}$$

Proposition 7 *Now consider the graph as an electrical network, where a wire linking v and x has conductance w_{vx}, i.e. resistance $1/w_{vx}$. Apply voltage* 1 *at v_0 and ground (i.e. set at voltage* 0*) the set A of vertices. Then* *the voltage at each vertex v equals $g(v)$ and the current along each wire (v,x) is f_{vx}/r, where $\mathbf{f} = \mathbf{f}^{v_0 \to A}$ and where*

$$r = \frac{1}{w_{v_0} P_{v_0}(T_A < T_{v_0}^+)} \tag{21}$$

is the effective resistance between v_0 and A.

What we will prove is, for quantities defined in terms of random walk,

$$\frac{f_{vx}}{r} = (g(v) - g(x)) w_{vx}. \tag{22}$$

Without going into details about electrical networks, (22) is *Ohm's law*

$$\text{Current} = \frac{\text{Potential difference}}{\text{Resistance}}$$

for each wire, and the voltages and current flow (normalized to 1) must satisfy the boundary conditions in (16,20). So the issue in relating the random walk story to the electrical network story is checking that (16,20,22) have a *unique* solution. But (16,22) imply that

$$\sum_x p_{vx} g(x) = g(v), \ v \notin \{v_0\} \cup A.$$

And Chapter 2 Lemma yyy shows that this equation, together with the boundary conditions (20), does indeed have a unique solution.

Proof of (22). Here is a "brute force" proof by writing everything in terms of mean hitting times. First, there is no less of generality in assuming that A is a singleton a, by the collapsing principle (Chapter 2 section yyy). Now by Chapter 2 Lemma yyy

$$f_{vx} = E_{v_0}(\text{number of visits to } v \text{ before time } T_a) \ p_{vx}$$

$$- E_{v_0}(\text{number of visits to } x \text{ before time } T_a) \ p_{xv}.$$

Chapter 2 Lemma yyy gave a formula for the expectations above, and using $\pi_v p_{vx} = \pi_x p_{xv} = w_{vx}/w$ we get

$$\frac{w}{w_{vx}} f_{vx} = E_a T_v - E_{v_0} T_v - E_a T_x + E_{v_0} T_x. \tag{23}$$

And Chapter 2 Corollaries yyy and yyy give a formula for g:

$$g(v) = (E_v T_a + E_a T_{v_0} - E_v T_{v_0}) P_{v_0}(T_a < T_{v_0}^+) \pi_{v_0}$$

which leads to

$$\frac{g(v) - g(x)}{P_{v_0}(T_a < T_{v_0}^+)\pi_{v_0}} = E_v T_a - E_x T_a - E_v T_{v_0} + E_x T_{v_0}. \tag{24}$$

But the right sides of (24) and (23) are equal, by the cyclic tour property (Lemma 1) applied to the tour v_0, x, a, v, v_0, and the result (22) follows after rearrangement, using $\pi_{v_0} = w_{v_0}/w$.

Remark. Note that, when identifying a reversible chain with an electrical network, the procedure of collapsing the set A of states of the chain to a singleton corresponds to the procedure of shorting together the vertices A of the electrical network.

3.3 Mean commute times

The classical use of the electrical network analogy in the mathematical literature is in the study of the recurrence or transience of infinite-state reversible chains by comparison arguments (Chapter yyy). As discussed in Doyle and Snell [17], the comparisons involve "cutting or shorting". Cutting an edge, or more generally decreasing an edge's conductance, can only increase an effective resistance. Shorting two vertices together (i.e. linking them with an edge of infinite conductance), or more generally increasing an edge's conductance, can only decrease an effective resistance. These ideas can be formalized via the extremal characterizations of section 7 without explicitly relying on the electrical analogy.

In our context of finite-state chains the key observation is the following. For not-necessarily-reversible chains we have (Chapter 2 Corollary yyy)

$$\frac{1}{\pi_v P_v(T_a < T_v^+)} = E_v T_a + E_a T_v \quad v \neq a \tag{25}$$

where we may call the right side the *mean commute time* between v and a. Comparing with (21) and using $\pi_v = w_v/w$ gives

Corollary 8 (commute interpretation of resistance) *Given two vertices v, a in a weighted graph, the effective resistance r_{va} between v and a is related to the mean commute time of the associated random walk by*

$$E_v T_a + E_a T_v = w r_{va}.$$

Note that the Corollary takes a simple form in the case of unweighted graphs:

$$E_v T_a + E_a T_v = 2|\mathcal{E}| r_{va}. \tag{26}$$

Note also that the Corollary does not hold so simply if a and v are both replaced by subsets – see Corollary 33.

Corollary 8 was apparently not stated explicitly or exploited until a 1989 paper of Chandra et al [10], but then rapidly became popular in the "randomized algorithms" community. The point is that "cutting or shorting" arguments can be used to bound mean commute times. As the simplest example, it is obvious that the effective resistance r_{vx} across an edge (v, x) is at most the resistance $1/w_{vx}$ of the edge itself, and so Corollary 8 implies the edge-commute inequality (Corollary 5). For other uses of "cutting" see yyy. For a use of "shorting" see yyy. Finally, we can use Corollary 8 to get simple exact expressions for mean commute times in some special cases, in particular for birth-and-death processes (i.e. weighted linear graphs) discussed in Chapter 5.

As with the infinite-space results, the electrical analogy provides a vivid language for comparison arguments, but the arguments themselves can be justified via the extremal characterizations of section 7 without explicit use of the analogy.

3.4 Foster's theorem

The commute interpretation of resistance allows us to rephrase Lemma 6 as the following result about electrical networks, due to Foster [19].

Corollary 9 (Foster's Theorem) *In a weighted n-vertex graph, let r_e be the effective resistance between the ends (a, b) of an edge e. Then*

$$\sum_e r_e w_e = n - 1.$$

* *

CONVENTION.

For the rest of the chapter we make the convention that we are dealing with a finite state, irreducible <u>reversible</u> chain, and we will not repeat the "reversible" hypothesis in each result. Instead we will say "general chain" to mean not-necessarily-reversible chain.

* *

4 The spectral representation

Use the transition matrix $\mathbf{P}$ to define

$$s_{ij} = \pi_i^{1/2} p_{ij} \pi_j^{-1/2}.$$

From definition (1), $\mathbf{S}$ is a *symmetric* matrix. So we can apply the elementary diagonalization theorem. The author finds it helpful to distinguish between the state space $I = \{i, j, \ldots\}$, of size n say, and the index set of integers $[n] = \{1, 2, \ldots, n\}$. The point being that the state space may have a lot of extra structure, whereas the index set has no obvious structure. The spectral theorem ([23] Theorem 4.1.5) gives a representation

$$\mathbf{S} = \mathbf{U} \Lambda \mathbf{U}^T$$

where $\mathbf{U} = (u_{im})_{i \in I, m \in [n]}$ is an orthonormal matrix, and $\Lambda = (\lambda_{m,m'})_{m,m' \in [n]}$ is a diagonal real matrix. We can write the diagonal entries of Λ as (λ_m), and arrange them in decreasing order. Then

$$1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n \geq -1. \quad (27)$$

The classical fact that $|\lambda_i| \leq 1$ follows easily from the fact that the entries of $\mathbf{S}^{(t)}$ are bounded as $t \to \infty$ by (28) below. These λ's are the eigenvalues of $\mathbf{P}$, as well as $\mathbf{S}$. That is, the solutions $(\lambda; x)$ with $x_i \not\equiv 0$ of

$$\sum_i x_i p_{ij} = \lambda x_j \text{ for all } j$$

are exactly the pairs

$$(\lambda = \lambda_m; x_i = c_m \pi_i^{1/2} u_{im}, i = 1, \ldots, n)$$

for $m = 1, \ldots, n$, where $c_m \neq 0$ is arbitrary. And the solutions of

$$\sum_j p_{ij} y_j = \lambda y_i \text{ for all } i$$

are exactly the pairs

$$(\lambda = \lambda_m; y_i = c_m \pi_i^{-1/2} u_{im}, i = 1, \ldots, n).$$

Note that an eigenvector (u_{i1}) corresponding to the eigenvector $\lambda_1 = 1$ is

$$u_{i1} = \pi_i^{1/2}.$$

Uniqueness of the stationary distribution now implies $\lambda_2 < 1$.

Now consider matrix powers. We have

$$\mathbf{S}^{(t)} = \mathbf{U}\Lambda^{(t)}\mathbf{U}^T$$

and

$$p_{ij}^{(t)} = \pi_i^{-1/2} s_{ij}^{(t)} \pi_j^{1/2}, \tag{28}$$

so

$$P_i(X_t = j) = \pi_i^{-1/2} \pi_j^{1/2} \sum_{m=1}^{n} \lambda_m^t u_{im} u_{jm}. \tag{29}$$

This is the *spectral representation formula.* In continuous time, the analogous formula is

$$P_i(X_t = j) = \pi_i^{-1/2} \pi_j^{1/2} \sum_{m=1}^{n} \exp(-\lambda_m t) u_{im} u_{jm}. \tag{30}$$

As before, $\mathbf{U}$ is an orthonormal matrix and $u_{i1} = \pi_i^{1/2}$, and now the λ's are the eigenvalues of $-\mathbf{Q}$. In the continuous-time setting, the eigenvalues satisfy

$$0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n. \tag{31}$$

Rather than give the general proof, let us consider the effect of continuizing the discrete-time chain (29). The continuized chain Y_t can be represented as $Y_t = X_{N(t)}$ where $N(t)$ has Poisson(t) distribution, so by conditioning on $N(t) = \nu$,

$$\begin{aligned} P_i(Y_t = j) &= \pi_i^{-1/2} \pi_j^{-1/2} \sum_{m=1}^{n} u_{im} u_{jm} \sum_{\nu=0}^{\infty} \lambda_m^{\nu} \frac{e^{-t} t^{\nu}}{\nu!} \\ &= \pi_i^{-1/2} \pi_j^{-1/2} \sum_{m=1}^{n} u_{im} u_{jm} \exp(-(1 - \lambda_m)t). \end{aligned}$$

So when we compare the spectral representations (29,30) for a discrete-time chain and its continuization, the orthonormal matrices are identical, and the eigenvalues are related by

$$\lambda_m^{(c)} = 1 - \lambda_m^{(d)} \tag{32}$$

superscripts (c) and (d) indicating continuous or discrete time. In particular, this relation holds for the basic discrete and continuous time random walks on a graph.

Let us point out some interesting simple consequences of the spectral representation. For these purposes continuous time is simpler. First,

$$P_i(X_t = j) - \pi_j \sim c_{ij} e^{-\lambda_2 t} \text{ as } t \to \infty \tag{33}$$

where $c_{ij} = \sum_{m:\lambda_m=\lambda_2} u_{im} u_{jm}$ and where "typically" $c_{ij} \neq 0$. (A precise statement is: there exists i such that

$$P_i(X_t = i) - \pi_i \sim c_{ii} e^{-\lambda_2 t}, \quad c_{ii} > 0 \tag{34}$$

by considering i such that $u_{i2} \neq 0$.) Thus λ_2 has the interpretation of "asymptotic rate of convergence to the stationary distribution". The author finds it simpler to interpret parameters measuring "time" rather than "1/ time", and so prefers to work with the *relaxation time* τ_2 defined by

$$\begin{aligned} \tau_2 &\equiv 1/\lambda_2 \text{ for a continuous-time chain} && (35) \\ &\equiv 1/(1-\lambda_2) \text{ for a discrete-time chain.} && (36) \end{aligned}$$

Note that by (32) the value of τ_2 is unchanged by continuizing a discrete-time chain.

Still in continuous time, the spectral representation gives

$$P_i(X_t = i) = \pi_i + \sum_{m \geq 2} u_{im}^2 \exp(-\lambda_m t) \tag{37}$$

so the right side is decreasing with t, and in fact is *completely monotone*, a subject pursued in section 5. Thus Z_{ii} defined in Chapter 2 section yyy satisfies

$$\begin{aligned} Z_{ii} &= \int_0^\infty (P_i(X_t = i) - \pi_i) dt \\ &= \sum_{m \geq 2} u_{im}^2 \lambda_m^{-1} \text{ by (37).} && (38) \end{aligned}$$

Using the orthonormal property of $\mathbf{U}$,

$$\sum_i Z_{ii} = \sum_{m \geq 2} \lambda_m^{-1}.$$

Applying Corollary yyy of Chapter 2, we obtain a fundamental result relating average hitting times to eigenvalues.

Proposition 10 (The eigentime identity) *For each i,*

$$\begin{aligned} \sum_j \pi_j E_i T_j &= \sum_{m \geq 2} \lambda_m^{-1} \text{ (continuous-time)} \\ &= \sum_{m \geq 2} (1 - \lambda_m)^{-1} \text{ (discrete-time)}. \end{aligned}$$

(The discrete-time version follows from (32).) Proposition 10 expands upon the *random target lemma*, which said that (even for non-reversible chains) $\sum_j \pi_j E_i T_j$ does not depend on i.

4.1 Mean hitting times and reversible chains

In Chapter 2 section yyy we listed identities for general chains such as such as the mean hitting time formulas

$$E_i T_j = (Z_{jj} - Z_{ij})/\pi_j; \quad E_\pi T_j = Z_{jj}/\pi_j.$$

There are a number of more complicated identities for general chains in which one side becomes zero for any reversible chain (by the symmetry property $\pi_i Z_{ij} = \pi_j Z_{ji}$) and which therefore simplify to give identities for reversible chains. We have already seen one example, the cyclic tour lemma, and the following result may be considered an extension of that Lemma.

Corollary 11 $E_\pi T_j - E_\pi T_i = E_i T_j - E_j T_i$.

This identity follows immediately from the mean hitting time formulas and the symmetry property. Note the following interpretation of the corollary. Define an ordering $i \ll j$ on the states by

$$i \ll j \text{ iff } E_\pi T_i \leq E_\pi T_j.$$

Then Corollary 11 implies

$$E_i T_j \geq E_j T_i \text{ iff } i \ll j.$$

Warning. Corollary 11 does not imply

$$\max_{i,j} E_i T_j \text{ is attained by some pair } (i_*, j_*) \text{ such that}$$

$$i_* \text{ attains } \min_i E_\pi T_i \text{ and } \ j_* \text{ attains } \max_j E_\pi T_j.$$

To outline a counter-example, consider random walk with holds on the n-cycle:

$$p_{i,i} = p_{i,i+1} = p_{i,i-1} = 1/3$$

with ± 1 taken modulo n (see Chapter 5 yyy). Here $E_\pi T_i$ is constant and $E_i T_j$ is maximized when $|i-j|$ is maximized. Now choose sufficiently small $\varepsilon > 0$ and states i_*, j_* and redefine

$$p_{i_*,i_*} = (1-2\varepsilon)/3, \ p_{i_*,i_*+1} = p_{i_*,i_*-1} = (1+\varepsilon)/3$$

$$p_{j_*,j_*} = (1+2\varepsilon)/3, \ p_{j_*,j_*+1} = p_{j_*,j_*-1} = (1-\varepsilon)/3.$$

Then i_* and j_* will attain the minimum and maximum of $E_\pi T_i$, but the maximum of $E_i T_j$ will still be attained by maximally distant (i,j). □

As a second instance of what reversibility implies, note that from (30) and the definition of Z_{ij},

$$Z_{ij} = \pi_i^{-1/2} \pi_j^{1/2} \sum_{m \geq 2} \lambda_m^{-1} u_{im} u_{jm}.$$

This implies

$$\text{the symmetrized matrix } \pi_i^{1/2} Z_{ij} \pi_j^{-1/2} \text{ is non-negative definite.} \tag{39}$$

Potential sophisticated applications of this fact will be touched upon in Chapter yyy. For now, note that a symmetric non-negative definite matrix M_{ij} has the property $M_{ij}^2 \leq M_{ii} M_{jj}$. This gives

$$Z_{ij}^2 \leq Z_{ii} Z_{jj} \pi_j / \pi_i. \tag{40}$$

This enables us to upper bound mean hitting times from arbitrary starts in terms of mean hitting times from stationary starts.

Lemma 12 $\max_{ij} E_i T_j \leq 2 \max_k E_\pi T_k$.

Proof. Using (40),

$$(Z_{ij}/\pi_j)^2 \leq (Z_{ii}/\pi_i)\ (Z_{jj}/\pi_j)$$

and so

$$-Z_{ij}/\pi_j \leq \max_k Z_{kk}/\pi_k.$$

So the mean hitting time formula gives the two equalities in

$$E_i T_j = \frac{Z_{jj}}{\pi_j} - \frac{Z_{ij}}{\pi_j} \leq 2 \max_k \frac{Z_{kk}}{\pi_k} = 2 \max_k E_\pi T_k.$$

5 Complete monotonicity

One advantage of working in continuous time is to exploit complete monotonicity properties. Abstractly, call $f : [0,\infty) \to [0,\infty)$ *completely monotone* (CM) if there is a non-negative measure μ on $[0,\infty)$ such that

$$f(t) = \int_0^\infty e^{-\theta t} \mu(d\theta),\ \ 0 \leq t < \infty. \tag{41}$$

Our applications will use only the special case of a finite sum

$$f(t) = \sum_m a_m e^{-\theta_m t},\ \text{ for some } a_m, \theta_m \geq 0. \tag{42}$$

but finiteness plays no essential role. If f is CM then (provided they exist) so are

$$-f'(t)$$

$$\bar{F}(t) \equiv \int_t^\infty f(s)ds \tag{43}$$

A *probability distribution* ν on $[0,\infty)$ is called CM if its tail distribution function $\bar{F}(t) = \nu(t,\infty)$ is CM; equivalently, if its density function f is CM (except here we must in the general case allow the possibility $f(0) = \infty$). In more probabilistic language, ν is CM iff it can be expressed as the distribution of ξ/Λ, where ξ and Λ are independent random variables such that

$$\xi \text{ has exponential(1) distribution; } \Lambda > 0. \tag{44}$$

Given a CM function or distribution, the *spectral gap* $\lambda \geq 0$ can be defined consistently by

$$\begin{aligned} \lambda &= \inf\{t > 0 : \mu[0,t] > 0\} \text{ in setting (41)} \\ &= \min\{\theta_m\} \text{ in setting (42)} \\ &= \operatorname{ess\,inf} \Lambda \text{ in setting (44).} \end{aligned}$$

This λ controls the behavior of $f(t)$ as $t \to \infty$. A key property of CM functions is that their value at a general time t can be bounded in terms of their behavior at 0 and at ∞, as follows.

Lemma 13 *Let f be CM with $0 < f(0) < \infty$. Then*

$$\exp\left(\frac{f'(0)t}{f(0)}\right) \leq \frac{f(t)}{f(0)} \leq \frac{\bar{F}(t)}{\bar{F}(0)} \leq \exp(-\lambda t), \ \ 0 \leq t < \infty$$

where λ is the spectral gap.

We might have $\bar{F}(0) = \infty$, but then $\bar{F}(t) = \infty$ and $\lambda = 0$ so the convention $\infty/\infty = 1$ works.

Proof. By scaling we may suppose $f(0) = 1$. So we can rewrite (41) as

$$f(t) = Ee^{-\Theta t} \tag{45}$$

where Θ has distribution μ. Then $f'(t) = -E\Theta e^{-\Theta t}$. Because $\theta \to e^{-\theta t}$ is decreasing, the random variables Θ and $e^{-\Theta t}$ are negatively correlated and so $E\Theta e^{-\Theta t} \leq (E\Theta)(Ee^{-\Theta t})$. This says $-f'(t) \leq -f'(0)\, f(t)$, in other words $\frac{d}{dt} \log f(t) \geq f'(0)$. Integrating gives $\log f(t) \geq t f'(0)$, which is the leftmost inequality. (Recall we rescaled to make $f(0) = 1$). For the second inequality,

$$\begin{aligned} \bar{F}(t) &= E\Theta^{-1}e^{-\Theta t}, \text{ integrating (45)} \\ &\geq (E\Theta^{-1})(Ee^{-\Theta t}) \text{ by positive correlation} \\ &= \bar{F}(0)\, f(t). \end{aligned}$$

Finally, from the definition of the spectral gap λ it is clear that $f(t)/f(0) \leq e^{-\lambda t}$. But $\bar{F}$ has the same spectral gap as f. □

Returning to the study of continuous-time reversible chains, the spectral representation (37) says that $P_i(X_t = i)$ is a CM function. It is often convenient to subtract the limit and say

$$P_i(X_t = i) - \pi_i \text{ is a CM function.} \tag{46}$$

More generally, given any function $g : I \to R$ the function

$$\rho(t) \equiv E g(X_t) g(X_0)$$

is CM for the *stationary* chain, because by (30)

$$\rho(t) = \sum_{m=1}^{n} \left(\sum_i \pi_i^{1/2} g(i) u_{im} \right) \left(\sum_j \pi_j^{1/2} g(j) u_{jm} \right) \exp(-\lambda_m t)$$

$$= \sum_{m=1}^{n} \left(\sum_i \pi_i^{1/2} g(i) u_{im} \right)^2 \exp(-\lambda_m t).$$

Specializing to the case $g = 1_A$ and conditioning,

$$P(X_t \in A | X_0 \in A) \text{ is a CM function} \qquad (47)$$

again assuming the stationary chain. When A is a singleton, this is (46).

Remark. To study directly discrete-time reversible chains, one would replace CM functions by sequences (f_n) of the form

$$f_n = \int_{-1}^{1} \theta^n \ \mu(d\theta).$$

But analogs of Lemma 13 and subsequent results (e.g. Proposition 19) become messier – so we prefer to derive discrete-time results by continuization.

5.1 Lower bounds on mean hitting times

As a quick application, we give bounds on mean hitting times to a single state from a stationary start. Recall $q_i = \sum_{j \neq i} q_{ij}$ is the exit rate from i, and τ_2 is the relaxation time of the chain.

Lemma 14 *For any state i in a continuous-time chain,*

$$\frac{(1-\pi_i)^2}{q_i \pi_i} \le E_\pi T_i \le \frac{\tau_2 (1-\pi_i)}{\pi_i}.$$

By continuization, the Lemma holds in discrete time, replacing q_i by $1 - p_{ii}$.

Proof. The mean hitting time formula is

$$\pi_i E_\pi T_i = Z_{ii} = \int_0^\infty (P_i(X_t = i) - \pi_i) dt.$$

Write $f(t)$ for the integrand. We know f is CM, and here $\lambda \geq \lambda_2$ by (37), and $f'(0) = -q_i$, so the extreme bounds of Lemma 13 become, after multiplying by $f(0) = 1 - \pi_i$,

$$(1-\pi_i)\exp(-q_i t/(1-\pi_i)) \leq f(t) \leq (1-\pi_i)e^{-\lambda_2 t}.$$

Integrating these bounds gives the result. □

We can now give general lower bounds on some basic parameters we will study in Chapter 4.

Proposition 15 *For a discrete-time chain on* n *states,*

$$\sum_j \pi_j E_\pi T_j \geq \frac{(n-1)^2}{n} \tag{48}$$

$$\max_{i,j}(E_i T_j + E_j T_i) \geq 2(n-1) \tag{49}$$

$$\max_{i,j} E_i T_j \geq n-1 \tag{50}$$

$$\tau_2 \geq \frac{n-1}{n}. \tag{51}$$

Remark. These inequalities become equalities for random walk on the complete graph (Chapter 5 Example yyy). By examining the proof, it can be shown that this is the only chain where an equality holds.

Proof. We go to the continuized chain, which has $q_i = 1 - p_{ii} \leq 1$. Then

$$\begin{aligned}
\sum_j \pi_j E_\pi T_j &\geq \sum_j (1-\pi_j)^2 \text{ by Lemma 14} \\
&= n - 2 + \sum_j \pi_j^2 \\
&\geq n - 2 + 1/n \\
&= (n-1)^2/n,
\end{aligned}$$

giving (48). By the eigentime identity,

$$\sum_j \pi_j E_\pi T_j = \sum_{m\geq 2} \lambda_m^{-1} \leq (m-1)\tau_2$$

and so (51) follows from (48).

Now fix i and write $\tau_0 = \sum_j \pi_j E_k T_j$, which (by the random target lemma) doesn't depend on k. Then

$$\sum_{j\neq i} \frac{\pi_j}{1-\pi_i}(E_i T_j + E_j T_i) = \frac{\tau_0 + E_\pi T_i}{1-\pi_i}. \tag{52}$$

If the right side were strictly less than $2(n-1)$ for all i, then $\sum_i \pi_i(\tau_0 + E_\pi T_i) < 2(n-1)\sum_i \pi_i(1-\pi_i)$, which implies

$$2\tau_0 < 2(n-1)(1-\sum_i \pi_i^2) \leq 2(n-1)(1-\frac{1}{n}) = \frac{2(n-1)^2}{n}$$

contradicting (48). Therefore there exists an i such that

$$\sum_{j\neq i} \frac{\pi_j}{1-\pi_i}(E_iT_j + E_jT_i) \geq 2(n-1)$$

and so there exists $j \neq i$ such that $E_iT_j + E_jT_i \geq 2(n-1)$. This is (49), and (50) follows immediately. □

There are several other results in the spirit of Lemma 14 and Proposition 15. For instance, Chapter 2 yyy says that for a general discrete-time chain,

$$\text{var }_iT_i^+ = \frac{2E_\pi T_i + 1}{\pi_i} - \frac{1}{\pi_i^2}.$$

Appealing to Lemma 14 gives, after a little algebra,

Corollary 16 *For any state i in a discrete-time chain,*

$$\text{var}_iT_i^+ \geq \frac{(1-\pi_i)(1-2\pi_i)}{\pi_i^2}.$$

Again. equality holds for random walk on the complete graph.

5.2 Smoothness of convergence

We're going to build some vague discussion around the following simple result.

Lemma 17

$$\sum_j \frac{p_{ij}^2(t)}{\pi_j} = \frac{p_{ii}(2t)}{\pi_i} \tag{53}$$

$$\left|\frac{p_{ik}(t+s)}{\pi_k} - 1\right| \leq \sqrt{\left(\frac{p_{ii}(2t)}{\pi_i} - 1\right)\left(\frac{p_{kk}(2s)}{\pi_k} - 1\right)}. \tag{54}$$

Proof.

$$\frac{p_{ik}(t+s)}{\pi_k} = \sum_j p_{ij}(t)\frac{p_{jk}(s)}{\pi_k} = \sum_j p_{ij}(t)\frac{p_{kj}(s)}{\pi_j}$$

by reversibility. Putting $k = i, \ s = t$ gives (53). Rewriting the equality as

$$\frac{p_{ik}(t+s)}{\pi_k} - 1 = \sum_j \pi_j \frac{p_{ij}(t) - \pi_j}{\pi_j}\frac{p_{kj}(s) - \pi_j}{\pi_j}$$

and applying the Cauchy-Schwartz inequality, we get the bound $\sqrt{a_i(t)a_k(s)}$, where

$$a_i(t) = \sum_j \frac{(p_{ij}(t) - \pi_j)^2}{\pi_j} = \sum_j \frac{p_{ij}^2(t)}{\pi_j} - 1 = \frac{p_{ii}(2t)}{\pi_i} - 1.$$

Remark. It is sometimes easier to use a cruder version of (54):

$$\frac{p_{ik}(t+s)}{\pi_k} \leq \sqrt{\frac{p_{ii}(2t)}{\pi_i}\frac{p_{kk}(2s)}{\pi_k}}$$

Discussion. Recalling from Chapter 2 section yyy the definition of L^2 distance between distributions, (53) says

$$||P_i(X_t \in \cdot) - \pi||_2^2 = \frac{p_{ii}(2t)}{\pi_i} - 1. \qquad (55)$$

In continuous time, we may regard the assertion "$||P_i(X_t \in \cdot) - \pi||_2$ is decreasing in t" as a consequence of the equality in (55) and the CM property of $p_{ii}(t)$. This assertion in fact holds for general chains, as pointed out in Chapter 2 Lemma yyy. Loosely, the general result of Chapter 2 Lemma yyy says that in a general chain the probabilities $(P_\rho(X_t = j)/\pi_j, j \in I)$ considered as an *unordered* set tend to smooth out as t increases. For a reversible chain, much more seems to be true. There is some "intrinsic geometry" on the state space such that, for the chain started at i, the probability distribution as time increases from 0 "spreads out smoothly" with respect to the geometry. It's hard to formalize that idea convincingly. On the other hand, (55) does say convincingly that the rate of convergence of the single probability $p_{ii}(t)$ to $\pi(i)$ is connected to a rate of convergence of the entire distribution $P_i(X_t \in \cdot)$ to $\pi(\cdot)$. This intimate connection between the local and the global behavior of reversible chains underlies many of the technical inequalities concerning mixing times in Chapter 4 and subsequent Chapters.

5.3 Inequalities for hitting time distributions on subsets

We mentioned in Chapter 2 section yyy that most of the simple identities there for mean hitting times $E_i T_j$ on singletons had no simple analogs for hitting times T_A on subsets. One exception is Kac's formula (Chapter 2 yyy), which says that for a general discrete-time chain

$$E_{\pi_A} T_A^+ = 1/\pi(A). \tag{56}$$

It turns out that for reversible chains there are useful inequalities relating the distributions of T_A under different initial distributions. These are simplest in continuous time as consequences of CM: as always, interesting consequences may be applied to discrete-time chains via continuization.

Recall π_A is the stationary distribution conditioned to A:

$$\pi_A(i) \equiv \pi(i)/\pi(A), \ i \in A.$$

Trivially

$$\begin{aligned} P_\pi(T_A > t) &= \pi(A^c) P_{\pi_{A^c}}(T_A > t) && (57) \\ E_\pi T_A &= \pi(A^c) E_{\pi_{A^c}} T_A. && (58) \end{aligned}$$

Define the *ergodic exit* distribution ρ_A from A by

$$\rho_A(j) \equiv \frac{\sum_{i \in A} \pi_i q_{ij}}{Q(A, A^c)}, \quad j \in A^c, \tag{59}$$

where $Q(A, A^c)$ is the *ergodic flow rate* out of A:

$$Q(A, A^c) \equiv \sum_{i \in A} \sum_{k \in A^c} \pi_i q_{ik}. \tag{60}$$

By stationarity, $Q(A, A^c) = Q(A^c, A)$.

Proposition 18 *Fix a subset A in a continuous-time chain.*

(i) T_A has CM distribution, when the initial distribution of the chain is any of the three distributions π or π_{A^c} or ρ_A.

(ii) The three hitting time distributions determine each other via (57) and

$$P_{\pi_{A^c}}(T_A \in (t, t+dt)) = \frac{P_{\rho_A}(T_A > t)}{E_{\rho_A} T_A} \ dt. \tag{61}$$

(iii) Write λ_A for the spectral gap associated with T_A (which is the same for each of the three initial distributions). Then

$$P_{\rho_A}(T_A > t) \le P_{\pi_{A^c}}(T_A > t) = \frac{P_\pi(T_A > t)}{\pi(A^c)} \le \exp(-\lambda_A t), \ t > 0 \qquad (62)$$

and in particular

$$\frac{\pi(A^c)}{Q(A, A^c)} = E_{\rho_A} T_A \le E_{\pi_{A^c}} T_A = \frac{E_\pi T_A}{\pi(A^c)} \le 1/\lambda_A. \qquad (63)$$

(iv)

$$E_\pi T_A \le \frac{\tau_2 \pi(A^c)}{\pi(A)}. \qquad (64)$$

Remarks. (a) In discrete time we can define ρ_A and $Q(A, A^c)$ by replacing q_{ij} by p_{ij} in (59,60), and then (63) holds in discrete time. The left equality is then a reformulation of Kac's formula (56), because

$$\begin{aligned} E_{\pi_A} T_A^+ &= 1 + P_{\pi_A}(X_1 \in A^c) E_{\pi_A}(T_A^+ | X_1 \in A^c) \\ &= 1 + \frac{Q(A, A^c)}{\pi(A)} E_{\rho_A} T_A. \end{aligned}$$

(b) Corollary 31 later shows that $1/\lambda_A \le \tau_2/\pi(A)$. So (64) can be regarded as a consequence of (63). Reverse inequalities will be studied in Chapter 4.

Proof of Proposition 18. First consider the case where A is a singleton a. Then (64) is an immediate consequence of Lemma 14. The equalities in (63) and in (61) are general identities for stationary processes (Chapter 2 yyy). We shall prove below that T_A is CM under $P_{\pi_{I \setminus \{a\}}}$. Then by (57) and (61,43) T_A is also CM under the other two initial distributions. Then the second inequality of (62) is the upper bound in Lemma 13, and the first is a consequence of (61) and Lemma 13. And (63) follows from (62) by integrating over t.

To prove that T_A is CM under $P_{\pi_{I \setminus \{a\}}}$, introduce a parameter $0 < \varepsilon < 1$ and consider the modified chain X_t^ε with transition rates

$$q_{ij}^\varepsilon = q_{ij}, \ i \ne a$$

$$q_{aj}^\varepsilon = \varepsilon q_{aj}.$$

The modified chain remains reversible, and its stationary distribution is of the form

$$\pi_i^\varepsilon = b_1 \pi_i, i \ne a; \quad \pi_a^\varepsilon = b_2$$

where the weights b_1, b_2 depend only on ε and π_a. Now as $\varepsilon \to 0$ with t fixed,

$$P_{\pi_{I\setminus\{a\}}}(X_t^\varepsilon \in I \setminus \{a\}) \to P_{\pi_{I\setminus\{a\}}}(T_a > t) \tag{65}$$

because the chain gets "stuck" upon hitting a. But the left side is CM by (47), so the right side (which does not depend on ε) is CM, because the class of CM distributions is closed under pointwise limits. (The last assertion is in general the continuity theorem for Laplace transforms [18] p. 83, though for our purposes we need only the simpler fact that the set of functions of the form (42) with at most n summands is closed).

This completes the proof when A is a singleton. We now claim that the case of general A follows from the collapsing principle (Chapter 2 section yyy), i.e. by applying the special case to the chain in which the subset A is collapsed into a single state. This is clear for all the assertions of the Lemma except for (64), for which we need the fact that the relaxation time τ_2^A of the collapsed chain is at most τ_2. This fact is proved as Corollary 24 below.

Remark. Note that the CM property implies a *super*multiplicitivity property for hitting times from stationarity in a continuous-time reversible chain:

$$P_\pi(T_A > s + t) \ge P_\pi(T_A > s)P_\pi(T_A > t).$$

Contrast with the general *sub*multiplicitivity property (Chapter 2 section yyy) which holds when P_π is replaced by $\max_i P_i$.

5.4 Approximate exponentiality of hitting times

In many circumstances, the distribution of the first hitting time T_A on a subset A of states with $\pi(A)$ small (equivalently, with ET_A large) can be approximated by the exponential distribution with the same mean. As with the issue of convergence to the stationary distribution, such approximations can be proved for general chains (see Notes), but it is easier to get explicit bounds in the reversible setting. If T has a CM distribution, then (as at (44), but replacing $1/\Lambda$ by Θ) we may suppose $T \stackrel{d}{=} \Theta\xi$. We calculate

$$ET = (E\Theta)(E\xi) = E\Theta; \quad ET^2 = (E\Theta^2)(E\xi^2) = 2E\Theta^2$$

and so

$$\frac{ET^2}{2(ET)^2} = \frac{E\Theta^2}{(E\Theta)^2} \ge 1$$

with equality iff Θ is constant, i.e. iff T has exponential distribution. This suggests that the difference $\frac{ET^2}{2(ET)^2} - 1$ can be used as a measure of "deviation from exponentiality". Let us quote a result of Mark Brown ([8] Theorem 4.1(iii)) which quantifies this idea in a very simple way.

Proposition 19 *Let T have CM distribution. Then*

$$\sup_t |P(T > t) - e^{-t/ET}| \le \frac{ET^2}{2(ET)^2} - 1.$$

So we can use this bound for hitting times T_A in a stationary reversible chain. At first sight the bound seems useful only if we can estimate $E_\pi T_A^2$ and $E_\pi T_A$ accurately. But the following remarkable variation shows that for the hitting time distribution to be approximately exponential it is sufficient that the mean hitting time be large compared to the relaxation time τ_2.

Proposition 20 *For a subset A of a continuous-time chain,*

$$\sup_t |P_\pi(T_A > t) - \exp(-t/E_\pi T_A)| \le \tau_2/E_\pi T_A.$$

Proof. By the collapsing principle (Chapter 2 section yyy) we may suppose A is a singleton j, because (Corollary 24 below) collapsing cannot increase the relaxation time. Combining the mean hitting time formula with the expression (38) for Z_{jj} in terms of the spectral representation (30),

$$E_\pi T_j = \pi_j^{-1} Z_{jj} = \pi_j^{-1} \sum_{m \ge 2} u_{jm}^2 \lambda_m^{-1}. \tag{66}$$

A similar calculation, exhibited below, shows

$$\frac{E_\pi T_j^2 - 2(E_\pi T_j)^2}{2} = \pi_j^{-1} \sum_{m \ge 2} u_{jm}^2 \lambda_m^{-2}. \tag{67}$$

But $\lambda_m^{-2} \le \lambda_2^{-1} \lambda_m^{-1} = \tau_2 \lambda_m^{-1}$ for $m \ge 2$, so the right side of (67) is bounded by $\pi_j^{-1} \tau_2 \sum_{m \ge 2} u_{jm}^2 \lambda_m^{-1}$, which by (66) equals $\tau_2 E_\pi T_j$. Applying Proposition 19 gives Proposition 20.

We give a straightforward but tedious verification of (67) (see also Notes). The identity $x^2/2 = \int_0^\infty (x-t)^+ dt,\ x \ge 0$ starts the calculation

$$\begin{aligned} \frac{1}{2} E_\pi T_j^2 &= \int_0^\infty E_\pi (T_j - t)^+ dt \\ &= \int_0^\infty \sum_i P_\pi(X_t = i, T_j > t) E_i T_j dt \end{aligned}$$

$$
\begin{aligned}
&= \sum_i E_iT_j \; E_\pi(\text{time spent at } i \text{ before } T_j) \\
&= \sum_i \frac{Z_{jj}-Z_{ij}}{\pi_j}\frac{Z_{jj}\pi_i - Z_{ji}\pi_j}{\pi_j} \\
&\quad \text{by Chapter 2 Lemmas yyy and yyy} \\
&= \sum_i \pi_j^{-2}\pi_i(Z_{jj}-Z_{ij})^2.
\end{aligned}
$$

Expanding the square, the cross-term vanishes and the first term becomes $(Z_{jj}/\pi_j)^2 = (E_\pi T_j)^2$, so

$$\frac{1}{2}E_\pi T_j^2 - (E_\pi T_j)^2 = \pi_j^{-2}\sum_i \pi_i Z_{ij}^2.$$

To finish the calculation,

$$
\begin{aligned}
\pi_j^{-1}\sum_i \pi_i Z_{ij}^2 &= \\
&= \pi_j^{-1}\sum_i \pi_i\Big(\int (p_{ij}(s)-\pi_j)ds\Big)\Big(\int (p_{ij}(t)-\pi_j)dt\Big) \\
&= \sum_i\Big(\int (p_{ji}(s)-\pi_i)ds\Big)\Big(\int (p_{ij}(t)-\pi_j)dt\Big) \\
&= \int\int (p_{jj}(s+t)-\pi_j)dsdt \\
&= \int t(p_{jj}(t)-\pi_j)dt \\
&= \int \sum_{m\geq 2} u_{jm}^2 t e^{-\lambda_m t}dt \\
&= \sum_{m\geq 2} u_{jm}^2\lambda_m^{-2}. \qquad \Box
\end{aligned}
$$

See the Notes for a related result, Theorem 39.

6 Extremal characterizations of eigenvalues

6.1 The Dirichlet formalism

A reversible chain has an associated *Dirichlet form* $\mathcal{E}$, defined as follows. For functions $g : I \to R$ write

$$\mathcal{E}(g,g) \equiv \frac{1}{2}\sum_i\sum_{j\neq i}\pi_i p_{ij}(g(j)-g(i))^2 \qquad (68)$$

in discrete time, and substitute q_{ij} for p_{ij} in continuous time. One can immediately check the following equivalent definitions. In discrete time

$$\mathcal{E}(g,g) = \frac{1}{2} E_\pi (g(X_1) - g(X_0))^2 = E_\pi g(X_0)(g(X_0) - g(X_1)). \tag{69}$$

In continuous time

$$\begin{aligned}\mathcal{E}(g,g) = \frac{1}{2} \lim_{t\to 0} t^{-1} E_\pi (g(X_t) - g(X_0))^2 &= \lim_{t\to 0} t^{-1} E_\pi g(X_0)(g(X_0) - g(X_t)) \\ &= - \sum_i \sum_j \pi_i g(i) q_{ij} g(j) \qquad (70)\end{aligned}$$

where the sum includes $j = i$. Note also that for random walk on a weighted graph, (68) becomes

$$\mathcal{E}(g,g) \equiv \frac{1}{2} \sum_i \sum_{j \neq i} \frac{w_{ij}}{w} (g(j) - g(i))^2. \tag{71}$$

Recall from Chapter 2 section yyy the discussion of L^2 norms for functions and measures. In particular

$$||g||_2^2 = \sum_i \pi_i g^2(i) = E_\pi g^2(X_0).$$

$$||\mu - \pi|_2^2 = \sum_i \frac{\mu_i^2}{\pi_i} - 1 \text{ for a probability distribution } \mu.$$

The relevance of $\mathcal{E}$ can be seen in the following lemma.

Lemma 21 *Write $\rho(t) = (\rho_j(t))$ for the distribution at time t of a continuous-time chain, with arbitrary initial distribution. Write $f_j(t) = \rho_j(t)/\pi_j$. Then*

$$\frac{d}{dt} ||\rho(t) - \pi||_2^2 = -2\mathcal{E}(f(t), f(t)).$$

Proof. $||\rho(t) - \pi||_2^2 = \sum_j \pi_j^{-1} \rho_j^2(t) - 1$, so using the forward equations

$$\frac{d}{dt} \rho_j(t) = \sum_i \rho_i(t) q_{ij}$$

we get

$$\begin{aligned}\frac{d}{dt} ||\rho(t) - \pi||_2^2 &= \sum_j \sum_i 2\pi_j^{-1} \rho_j(t) \rho_i(t) q_{ij} \\ &= 2 \sum_j \sum_i f_j(t) f_i(t) \pi_i q_{ij}\end{aligned}$$

and the result follows from (70).

6.2 Summary of extremal characterizations

For ease of comparison we state below three results which will be proved in subsequent sections. These results are commonly presented "the other way up" as *infs* rather than *sups*, but our presentation is forced by our convention of consistently defining parameters to have dimensions of "time" rather than "1/time". The *sups* are over functions $g : I \to R$ satisfying specified constraints, and excluding $g \equiv 0$. The results below are the same in continuous and discrete time – that is, continuization doesn't change the numerical values of the quantities we consider. We shall give the proofs in discrete time.

Extremal characterization of relaxation time. *The relaxation time τ_2 satisfies*

$$\tau_2 = \sup\{||g||_2^2/\mathcal{E}(g,g) : \sum_i \pi_i g(i) = 0\}.$$

Extremal characterization of quasistationary mean hitting time. *Given a subset A, let α_A be the quasistationary distribution on A^c defined at (75). Then the quasistationary mean exit time is*

$$E_{\alpha_A} T_A = \sup\{||g||_2^2/\mathcal{E}(g,g) : g \geq 0, g = 0 \text{ on } A\}.$$

Extremal characterization of mean commute times. *For distinct states i, j the mean commute time satisfies*

$$E_i T_j + E_j T_i = \sup\{1/\mathcal{E}(g,g) : 0 \leq g \leq 1, g(i) = 1, g(j) = 0\}.$$

Because the state space is finite, the *sups* are attained, and there are theoretical descriptions of the g attaining the extrema in each case. An immediate practical use of thse characterizations in concrete examples is to obtain lower bounds on the parameters by inspired guesswork, that is by choosing some simple explicit "test function" g which seems qualitatively right and computing the right-hand quantity. See yyy for examples. Of course we cannot obtain *upper* bounds this way, but extremal characterizations can be used as a starting point for further theoretical work (see in particular the bounds on τ_2 in Chapter 4 yyy).

6.3 The extremal characterization of relaxation time

The first two extremal characterizations are in fact just reformulations of the classical *Rayleigh-Ritz* extremal characterization of eigenvalues, which goes as follows ([23] Theorem 4.2.2 and eq. 4.2.7). Let S be a symmetric matrix with eigenvalues $\mu_1 \geq \mu_2 \geq \ldots$. Then

$$\mu_1 = \sup_{\mathbf{x}} \frac{\sum_i \sum_j x_i s_{ij} x_j}{\sum_i x_i^2} \tag{72}$$

and an $\mathbf{x}$ attaining the *sup* is an eigenvalue corresponding to μ_1 (of course *sups* are over $\mathbf{x} \not\equiv 0$). And

$$\mu_2 = \sup_{\mathbf{y}: \sum_i y_i x_i = 0} \frac{\sum_i \sum_j y_i s_{ij} y_j}{\sum_i y_i^2} \tag{73}$$

and a $\mathbf{y}$ attaining the *sup* is an eigenvalue corresponding to μ_2.

As observed in section 4, given a discrete-time chain with transition matrix P, the symmetric matrix $s_{ij} = \pi_i^{1/2} p_{ij} \pi_j^{-1/2}$ has maximal eigenvalue 1 with corresponding eigenvector $\pi_i^{1/2}$. So applying (73) and writing $y_i = \pi_i^{1/2} g(i)$, the second-largest eigenvalue (of S and hence of P) is given by

$$\lambda_2 = \sup_{g: \sum_i \pi_i g(i) = 0} \frac{\sum_i \sum_j \pi_i g(i) p_{ij} g(j)}{\sum_i \pi_i g^2(i)}.$$

In probabilistic notation the fraction is

$$\frac{E_\pi g(X_0) g(X_1)}{E_\pi g^2(X_0)} = 1 - \frac{E_\pi g(X_0)(g(X_1) - g(X_0))}{E_\pi g^2(X_0)} = 1 - \frac{\mathcal{E}(g,g)}{||g||_2^2}.$$

Since $\tau_2 = 1/(1-\lambda_2)$ in discrete time we have proved the first of our extremal characterizations.

Theorem 22 (Extremal characterization of relaxation time) *The relaxation time τ_2 satisfies*

$$\tau_2 = \sup\{||g||_2^2 / \mathcal{E}(g,g) : \sum_i \pi_i g(i) = 0\}.$$

xxx improve/delete discussion below???

A function g_0, say, attaining the *sup* in the extremal characterization is, by examining the argument above, a right eigenvector associated with λ_2:

$$\sum_j p_{ij} g_0(j) = \lambda_2 g_0(i).$$

The corresponding left eigenvector θ :

$$\sum_i \theta_i p_{ij} = \lambda_2 \theta_j \text{ for all } j$$

is the signed measure θ such that $\theta_i = \pi_i g_0(i)$. To continue a somewhat informal discussion of the interpretation of g_0, it is convenient to switch to continuous time (to avoid issues of negative eigenvalues) and to assume λ_2 has multiplicity 1. The equation which relates distribution at time t to initial distribution,

$$\rho_j(t) = \sum_i \rho_i(0) p_{ij}(t)$$

can also be used to define signed measures evolving from an initial signed measure. For the initial measure θ we have

$$\theta(t) = e^{-t/\tau_2}\theta.$$

For any signed measure with $\sum_i \nu_i(0) = 0$ we have

$$\nu(t) \sim ce^{-t/\tau_2}\theta; \quad c = \sum_i \nu_i(0)\theta_i/\pi_i = \sum_i \nu_i(0) g_0(i).$$

So θ can be regarded as "the signed measure which relaxes to 0 most slowly". For a probability measure $\rho(0)$, considering $\rho(0) - \pi$ gives

$$\rho(t) - \pi \sim ce^{-t/\tau_2}\theta, \quad c = \sum_i (\rho_i(0) - \pi_i) g_0(i) = \sum_i \rho_i(0) g_0(i). \tag{74}$$

So θ has the interpretation of "the asymptotic normalized difference between the true distribution at time t and the stationary distribution". Finally, from (74) with $\rho(0)$ concentrated at i (or from the spectral representation)

$$P_i(X_t \in \cdot) - \pi \sim g_0(i) e^{-t/\tau_2}\theta.$$

So g_0 has the interpretation "the asymptotic normalized size of deviation from stationarity, as a function of the starting state".

xxx In using the extremal characterization to obtain a lower bound from a test function g, guess "smooth" function g, cf. the cosine function on the n-cycle and Chapter 14 Example yyy.

6.4 Simple applications

Here is a fundamental "finite-time" result.

Lemma 23 (xxx needs good name) *Write $\rho(t) = (\rho_j(t))$ for the distribution at time t of a continuous-time chain, with arbitrary initial distribution. Then*

$$||\rho(t) - \pi||_2 \leq e^{-t/\tau_2}||\rho(0) - \pi||_2.$$

Proof. Write $f_j(t) = \rho_j(t)/\pi_j$. Then

$$\begin{aligned}
\frac{d}{dt}||\rho(t) - \pi||_2^2 &= -2\mathcal{E}(f(t), f(t)) \text{ by Lemma 21} \\
&= -2\mathcal{E}(f(t) - 1, f(t) - 1) \\
&\leq -2\,\frac{||f(t) - 1||_2^2}{\tau_2} \text{ by the extremal characterization} \\
&= \frac{-2}{\tau_2}\,||\rho(t) - \pi||_2^2.
\end{aligned}$$

Integrating, $||\rho(t) - \pi||_2^2 \leq e^{-2t/\tau_2}||\rho(0) - \pi||_2^2$, and the result follows. □.

Our main use of the extremal characterization is to compare relaxation times of different chains on the same (or essentially the same) state space. Here are three instances. The first is a result we have already exploited in section 5.

Corollary 24 *Given a chain with relaxation time τ_2, let τ_2^A be the relaxation time of the chain with subset A collapsed to a singleton $\{a\}$ (Chapter 2 section yyy). Then $\tau_2^A \leq \tau_2$.*

Proof. Any function g on the states of the collapsed chain can be extended to the original state space by setting $g = g(a)$ on A, and $\mathcal{E}(g, g)$ and $\sum_i \pi_i g(i)$ and $||g||_2^2$ are unchanged. So consider a g attaining the *sup* in the extremal characterization of τ_2^A and use this as a test function in the extremal characterization of τ_2.

Remark. An extension of Corollary 24 will be provided by the contraction principle (Chapter 4 Proposition yyy).

Corollary 25 *Let τ_2 be the relaxation time for a "fluid model" continuous-time chain associated with a graph with weights (w_e) (recall (14)) and let τ_2^* be the relaxation time when the weights are (w_e^*). If $w_e^* \geq w_e$ for all edges e then $\tau_2^* \leq \tau_2$.*

Proof. Each stationary distribution is uniform, so $||g||_2^2 = ||g||_2^{*2}$ while $\mathcal{E}^*(g,g) \geq \mathcal{E}(g,g)$. So the result is immediate from the extremal characterization. $\Box$.

The next result is a prototype for more complicated "indirect comparison" arguments later (Chapters yyy). It is convenient to state it in terms of random walk on a weighted graph. Recall (section 2) that a reversible chain specifies a weighted graph with edge-weights $w_{ij} = \pi_i p_{ij}$, vertex-weights $w_i = \pi_i$ and total weight $w = 1$.

Lemma 26 (the direct comparison lemma) *Let (w_e) and (w_e^*) be edge-weights on a graph, let (w_i) and (w_i^*) be the vertex-weights, and let τ_2 and τ_2^* be the relaxation times for the associated random walks. Then*

$$\frac{\min_e w_e/w_e^*}{\max_i w_i/w_i^*} \leq \frac{\tau_2}{\tau_2^*} \leq \frac{\max_i w_i/w_i^*}{\min_e w_e/w_e^*}$$

where in $\min_e$ *we don't count loops* $e = (v,v)$.

Proof. For any g, by (71)

$$w^*\mathcal{E}^*(g,g) \geq w\mathcal{E}(g,g) \min_e w_e^*/w_e.$$

And since $w||g||_2^2 = \sum_i w_i g^2(i)$,

$$w^*||g||_2^{*2} \leq w||g|_2^2 \ \max_i w_i^*/w_i.$$

So if g has π^*-mean 0 and π-mean b then

$$\frac{||g||_2^{*2}}{\mathcal{E}^*(g,g)} \leq \frac{||g-b||_2^{*2}}{\mathcal{E}^*(g-b,g-b)} \leq \frac{||g-b||_2^2}{\mathcal{E}(g-b,g-b)} \ \frac{\max_i w_i^*/w_i}{\min_e w_e^*/w_e}.$$

By considering the g attaining the extremal characterization of τ_2^*,

$$\tau_2^* \leq \tau_2 \ \frac{\max_i w_i^*/w_i}{\min_e w_e^*/w_e}.$$

This is the lower bound in the lemma, and the upper bound by reversing the roles of w_e and w_e^*.

Remarks. Sometimes τ_2 is very sensitive to apparently-small changes in the chain. Consider random walk on an unweighted graph. If we add extra edges, but keeping the total number of added edges small relative to the number of original edges, then we might guess that τ_2 could not increase or decrease much. But the examples outlined below show that τ_2 may in fact change substantially in either direction.

Example 27 Take two complete graphs on n vertices and join with a single edge. Then $w = 2n(n-1)+2$ and $\tau_2 \sim n^2/2$. But if we extend the single join-edge to an n-edge matching of the vertices in the original two complete graphs, then $w^* = 2n(n-1)+2n \sim w$ but $\tau_2^* \sim n/2$.

Example 28 Take a complete graph on n vertices. Take $k = o(n^{1/2})$ new vertices and attach each to distinct vertices of the original complete graph. Then $w = n(n-1) + 2k$ and τ_2 is bounded. But if we now add all edges within the new k vertices, $w^* = n(n-1)+2k+k(k-1) \sim w$ but $\tau_2^* \sim k$ provided $k \to \infty$.

As these examples suggest, comparison arguments are most effective when the stationary distributions coincide. Specializing Lemma 26 to this case, and rephrasing in terms of (reversible) chains, gives

Lemma 29 (the direct comparison lemma) *For transition matrices* $\mathbf{P}$ *and* $\mathbf{P}^*$ *with the same stationary distribution* π, *if*

$$p_{ij} \geq \delta p^*_{ij} \text{ for all } j \neq i$$

then $\tau_2 \leq \delta^{-1}\tau_2^*$.

Remarks. The hypothesis can be rephrased as $\mathbf{P} = \delta\mathbf{P}^* + (1-\delta)\mathbf{Q}$, where $\mathbf{Q}$ is a (maybe not irreducible) reversible transition matrix with stationary distribution π. When $\mathbf{Q} = \mathbf{I}$ we have $\tau_2^* = \delta^{-1}\tau_2$, so an interpretation of the lemma is that "combining transitions of $\mathbf{P}$ with noise can't increase mixing time any more than combining transitions with holds".

6.5 Quasistationarity

Given a subset A of states, let $\mathbf{P}^A$ be $\mathbf{P}$ restricted to A^c. Then $\mathbf{P}^A$ will be a substochastic matrix, i.e. the row-sums are at most 1, and some row-sum is strictly less than 1. Suppose $\mathbf{P}^A$ is irreducible. As a consequence of the Perron-Frobenius theorem (e.g. [23] Theorem 8.4.4) for the positive matrix $\mathbf{P}^A$, there is a unique $0 < \lambda < 1$ (specifically, the largest eigenvalue of $\mathbf{P}^A$) such that there is a probability distribution α satisfying

$$\alpha = 0 \text{ on } A, \quad \sum_i \alpha_i p_{ij} = \lambda\alpha_j, \ j \in A^c. \tag{75}$$

Writing α_A and λ_A to emphasize dependence on A, (75) can be rephrased as saying that under P_{α_A} the hitting time T_A has geometric distribution

$$P_{\alpha_A}(T_A \geq m) = (1-\lambda_A)^m, \ m \geq 0$$

$$E_{\alpha_A} T_A = \frac{1}{1-\lambda_A}.$$

Call α_A the *quasistationary distribution* and $E_{\alpha_A} T_A$ the *quasistationary mean exit time*. The facts above do not depend on reversibility, but invoking now our standing assumption that chains are reversible we can prove our second extremal characterization.

Theorem 30 (Extremal characterization of quasistationary mean hitting time) *The quasistationary mean exit time satisfies*

$$E_{\alpha_A} T_A = \sup\{||g||_2^2/\mathcal{E}(g,g) : g \geq 0, g = 0 \text{ on } A\}. \tag{76}$$

Proof. $s_{ij}^A = \pi_i^{1/2} p_{ij}^A \pi_j^{-1/2}$ is symmetric with largest eigenvalue λ_A. Putting $x_i = \pi_i^{1/2} g(i)$ in the characterization (72) gives

$$\lambda_A = \sup_g \frac{\sum_i \sum_j \pi_i g(i) p_{ij}^A g(j)}{\sum_i \pi_i g^2(i)}.$$

Clearly the *sup* is attained by non-negative g, and though the sums above are technically over A^c we can sum over all I by setting $g = 0$ on A. So

$$\lambda_A = \sup \left\{ \frac{\sum_i \sum_j \pi_i g(i) p_{ij}^A g(j)}{\sum_i \pi_i g^2(i)} : g \geq 0, g = 0 \text{ on } A \right\}.$$

As in the proof of Theorem 22 this rearranges to

$$\frac{1}{1-\lambda_A} = \sup\{||g||_2^2/\mathcal{E}(g,g) : g \geq 0, g = 0 \text{ on } A\}$$

establishing Theorem 30.

Remarks. These remarks closely parallel the remarks at the end of section 6.3. The *sup* in Theorem 30 is attained by the function g_0 which is the right eigenvector associated with λ_A, and by reversibility this is

$$g_0(i) = \alpha_A(i)/\pi_i. \tag{77}$$

It easily follows from (75) that

$$P_{\alpha_A}(X_t = j | T_A > t) = \alpha_A(j) \text{ for all } j \text{ and } t$$

which explains the name *quasistationary distribution* for α_A. A related interpretation of α_A is as the distribution of the Markov chain conditioned on having been in A^c for the infinite past. More precisely, one can use the Perron-Frobenius theorem to prove that

$$P(X_t = j | T_A > t) \to \alpha_A(j) \text{ as } t \to \infty$$

provided P^A is aperiodic as well as irreducible.

Our fundamental use of quasistationarity is the following.

Corollary 31 *For any subset A, the quasistationary mean hitting time satisfies*

$$E_{\alpha_A} T_A \le \tau_2 / \pi(A).$$

Proof. As at (77) set $g(i) = \alpha_A(i)/\pi_i$, so

$$E_{\alpha_A} T_A = ||g||_2^2 / \mathcal{E}(g,g). \tag{78}$$

Now $E_\pi g(X_0) = 1$, so applying the extremal characterization of relaxation time to $g - 1$,

$$\tau_2 \ge \frac{||g||_2^2}{\mathcal{E}(g-1, g-1)} = \frac{||g||_2^2 - 1}{\mathcal{E}(g,g)} = (E_{\alpha_A} T_A)\left(1 - \frac{1}{||g||_2^2}\right) \tag{79}$$

the last equality using (78). Since α_A is a probability distribution on A^c we have

$$1 = E_\pi 1_{A^c}(X_0) g(X_0)$$

and so by Cauchy-Schwarz

$$1^2 \le (E_\pi 1_{A^c}(X_0)) \times ||g||_2^2 = (1 - \pi(A))||g||_2^2.$$

Rearranging,

$$1 - \frac{1}{||g||_2^2} \ge \pi(A)$$

and substituting into (79) gives the desired bound.

7 Extremal characterizations and mean hitting times

Theorem 32 (Extremal characterization of mean commute times) *For distinct states i, a the mean commute time satisfies*

$$E_i T_a + E_a T_i = \sup\{1/\mathcal{E}(g,g) : 0 \le g \le 1, g(i) = 1, g(a) = 0\} \tag{80}$$

and the sup *is attained by* $g(j) = P_j(T_i < T_a)$. *For a subset* A *and a state* $i \notin A$,

$$\pi_i P_i(T_A < T_i^+) = \inf\{\mathcal{E}(g,g) : 0 \leq g \leq 1, g(i) = 1, g(\cdot) = 0 \textit{ on } A\} \qquad (81)$$

and the inf *is attained by* $g(j) = P_j(T_i < T_A)$.

Proof. As noted at (25), form (80) follows from form (81) with $A = \{a\}$. To prove (81), consider g satisfying the specified boundary conditions. Inspecting (68), the contribution to $\mathcal{E}(g,g)$ involving a fixed state j is

$$\sum_{k \neq j} \pi_j p_{jk} (g(k) - g(j))^2. \qquad (82)$$

As a function of $g(j)$ this is minimized by

$$g(j) = \sum_k p_{jk} g(k). \qquad (83)$$

Thus the g which minimizes $\mathcal{E}$ subject to the prescribed boundary conditions on $A \cup \{i\}$ must satisfy (83) for all $j \notin A \cup \{i\}$, and by Chapter 2 Lemma yyy the unique solution of these equations is $g(j) = P_j(T_i < T_A)$. Now apply to this g the general expression (69)

$$\mathcal{E}(g,g) = \sum_j \pi_j g(j) \left(g(j) - \sum_k p_{jk} g(k) \right).$$

For $j \notin A \cup \{i\}$ the term $(g(j) - \sum_k p_{jk} g(k))$ equals zero, and for $j \in A$ we have $g(j) = 0$, so only the $j = i$ term contributes. So

$$\begin{aligned} \mathcal{E}(g,g) &= \pi_i (1 - \sum_k p_{ik} g(k)) \\ &= \pi_i (1 - P_i(T_i^+ < T_A)) \\ &= \pi_i P_i(T_A < T_i^+) \end{aligned} \qquad (84)$$

giving (81). □

The analogous result for two disjoint subsets A, B is a little complicated to state. The argument above shows that

$$\inf\{\mathcal{E}(g,g) : g(\cdot) = 0 \text{ on } A, g(\cdot) = 1 \text{ on } B\}.$$

is attained by $g_0(j) = P_j(T_B < T_A)$ and that this g_0 satisfies

$$\mathcal{E}(g_0, g_0) = \sum_{i \in B} \pi_i P_i(T_A < T_B^+). \tag{85}$$

We want to interpret this inverse of this quantity as a mean time for the chain to commute from A to B and back. Consider the stationary chain $(X_t; -\infty < t < \infty)$. We can define what is technically called a "marked point process" which records the times at which A is first entered after a visit to B and vice versa. Precisely, define Z_t taking values in $\{\alpha, \beta, \delta\}$ by

$$Z_t = \beta \text{ if } \exists s < t \text{ such that } X_s \in A, X_t \in B, X_u \notin A \cup B \ \forall s < u < t.$$

$$Z_t = \alpha \text{ if } \exists s < t \text{ such that } X_s \in B, X_t \in A, X_u \notin B \cup A \ \forall s < u < t.$$

$$Z_t = \delta \text{ otherwise .}$$

So the times t when $Z_t = \beta$ are the times of first return to B after visiting A, and the times t when $Z_t = \alpha$ are the times of first return to A after visiting B. Now (Z_t) is a stationary process. By considering the time-reversal of X, we see that for $i \in B$

$$P(X_0 = i, Z_0 = \beta) = P(X_0 = i, T_A < T_B^+) = \pi_i P_i(T_A < T_B^+).$$

So (85) shows $P(Z_0 = \beta) = \mathcal{E}(g_0, g_0)$. If we define T_{BAB}, "the typical time to go from B to A and back to B", to have the conditional distribution of $\min\{t \geq 1 : Z_t = \beta\}$ given $Z_0 = \beta$, then Kac's formula for the (non-Markov) stationary process Z (see e.g. [18] Theorem 6.3.3) says that $ET_{BAB} = 1/P(Z_0 = \beta)$. So we have proved

Corollary 33

$$ET_{BAB} = \sup\{1/\mathcal{E}(g,g) : 0 \leq g \leq 1, g(\cdot) = 0 \textit{ on } A, g(\cdot) = 1 \textit{ on } B\}$$

and the sup *is attained by* $g(i) = P_i(T_B < T_A)$.

As another interpretation of this quantity, define

$$\rho_B(\cdot) = P(X_0 \in \cdot | Z_0 = \beta), \quad \rho_A(\cdot) = P(X_0 \in \cdot | Z_0 = \alpha).$$

Interpret ρ_B and ρ_A as the distribution of hitting places on B and on A in the commute process. It is intuitively clear, and not hard to verify, that

$$P_{\rho_A}(X(T_B) \in \cdot) = \rho_B(\cdot), \ P_{\rho_B}(X(T_A) \in \cdot) = \rho_A(\cdot)$$

$$ET_{BAB} = E_{\rho_B} T_A + E_{\rho_A} T_B.$$

In particular

$$\min_{i \in B} E_i T_A + \min_{i \in A} E_i T_B \leq ET_{BAB} \leq \max_{i \in B} E_i T_A + \max_{i \in A} E_i T_B.$$

7.1 Thompson's principle and leveling networks.

Theorem 32 was stated in terms of (reversible) Markov chains. Rephrasing in terms of discrete-time random walk on a weighted graph gives the usual "electrical network" formulation of the Dirichlet principle stated below, using (71,81) and (84). Recall from Proposition 7 that the effective resistance r between v_0 and A is, in terms of the random walk,

$$r = \frac{1}{w_{v_0} P_{v_0}(T_A < T_{v_0}^+)}. \quad (86)$$

Proposition 34 (The Dirichlet principle) *Take a weighted graph and fix a vertex v_0 and a subset A of vertices not containing v_0. Then $\frac{1}{2}\sum_i \sum_j w_{ij}(g(j) - g(i))^2$ is minimized, over all functions $g : I \to [0,1]$ with $g(v_0) = 1, g(\cdot) = 0$ on A, by the function $g(i) \equiv P_i(T_{v_0} < T_A)$ (where probabilities refer to random walk on the weighted graph), and the minimum value equals $1/r$, where r is the effective resistance (86).*

There is a dual form of the Dirichlet principle, which following Doyle and Snell [17] we call

Proposition 35 (Thompson's principle) *Take a weighted graph and fix a vertex v_0 and a subset A of vertices not containing v_0. Let $\mathbf{f} = f_{ij}$ denote a unit flow from v_0 to A. Then $\frac{1}{2}\sum_i \sum_j f_{ij}^2/w_{ij}$ is minimized, over all such flows, by the flow $\mathbf{f}^{v_0 \to A}$ (defined at (18)) associated with the random walk from v_0 to A, and the minimum value equals the effective resistance r appearing in (86).*

Recall that a flow is required to have $f_{ij} = 0$ whenever $w_{ij} = 0$, and interpret sums $\sum_i \sum_j$ as sums over ordered pairs (i,j) with $w_{ij} > 0$.

Proof. Write $\psi(\mathbf{f}) \equiv \frac{1}{2}\sum_i \sum_j f_{ij}^2/w_{ij}$. By formula (22) relating the random walk notions of "flow" and "potential", the fact that $\psi(\mathbf{f}^{v_0 \to A}) = r$ is immediate from the corresponding equality in the Dirichlet principle. So the issue is to prove that for a unit flow $\mathbf{f}^*$, say, attaining the minimum of $\psi(\mathbf{f})$, we have $\psi(\mathbf{f}^*) = \psi(\mathbf{f}^{v_0 \to A})$. To prove this, consider two arbitrary paths (y_i) and (z_j) from v_0 to A, and let $\mathbf{f}^\varepsilon$ denote the flow $\mathbf{f}^*$ modified by adding flow rates $+\varepsilon$ along the edges (y_i, y_{i+1}) and by adding flow rates $-\varepsilon$ along the edges (z_i, z_{i+1}). Then $\mathbf{f}^\varepsilon$ is still a unit flow from v_0 to A. So the function $\varepsilon \to \psi(\mathbf{f}^\varepsilon)$ must have derivative zero at $\varepsilon = 0$, and this becomes the condition that

$$\sum_i f^*_{y_i,y_{i+1}}/w_{y_i,y_{i+1}} = \sum_i f^*_{z_i,z_{i+1}}/w_{z_i,z_{i+1}}.$$

So the sum is the same for all paths from v_0 to A. Fixing x, the sum must be the same for all paths from x to A, because two paths from x to A could be extended to paths from v_0 to A by appending a common path from v_0 to x. It follows that we can define $g^*(x)$ as the sum $\sum_i f^*_{x_i,x_{i+1}}/w_{x_i,x_{i+1}}$ over some path (x_i) from x to A, and the sum does not depend on the path chosen. So

$$g^*(x) - g^*(z) = \frac{f^*_{xz}}{w_{xz}} \text{ for each edge } (x,z) \text{ not contained within } A. \qquad (87)$$

The fact that $\mathbf{f}^*$ is a flow means that, for $x \notin A \cup \{v_0\}$,

$$0 = \sum_{z:w_{xz}>0} f^*_{xz} = \sum_z w_{xz}(g^*(x) - g^*(z)).$$

So g^* is a harmonic function outside $A \cup \{v_0\}$, and $g^* = 0$ on A. So by the uniqueness result (Chapter 2 Lemma yyy) we have that g^* must be proportional to g, the minimizing function in Proposition 34. So $\mathbf{f}^*$ is proportional to $\mathbf{f}^{v_0 \to A}$, because the relationship (87) holds for both, and then $\mathbf{f}^* = \mathbf{f}^{v_0 \to A}$ because both are *unit* flows. □

A remarkable statistical interpretation was discussed in a monograph of Borre and Meissl [5]. Imagine a finite set of locations such as hilltops. For each pair of locations (i,j) with a clear line-of-sight, measure the elevation difference D_{ij} = height of j minus height of i. Consider the associated graph (whose edges are such pairs (i,j)), and suppose it is connected. Take one location v_0 as a benchmark "height 0". If our measurements were exact we could determine the height of location x by adding the D's along a path from v_0 to x, and the sum would not depend on the path chosen. But suppose our measurements contain random errors. Precisely, suppose $D_{i,j}$ equals the true height difference $h(j) - h(i)$ plus an error Y_{ij} which has mean 0, variance $1/w_{ij}$ and is independent for different measurements. Then it seems natural to estimate the height of x by taking some average $\hat{h}(x)$ over paths from v_0 to x, and it turns out that the "best" way to average is to use the random walk from v_0 to x and average (over realizations of the walk) the net height climbed by the walk.

In mathematical terms, the issue is to choose weights f_{ij}, not depending on the function h, such that

$$\hat{h}(x) \equiv \frac{1}{2} \sum_i \sum_j f_{ij} D_{ij}$$

has $E\hat{h}(x) = h(x)$ and minimal variance. It is not hard to see that the former "unbiased" property holds iff f is a unit flow from v_0 to x. Then

$$\text{var } \hat{h}(x) = \frac{1}{4}\sum_i \sum_j f_{ij}^2 \text{var}(D_{ij}) = \frac{1}{4}\sum_i \sum_j \frac{f_{ij}^2}{w_{ij}}$$

and Proposition 35 says this is minimized when we use the flow obtained from the random walk on the weighted graph from v_0 to x. But then

$$\hat{h}(x) = E_{v_0} \sum_{t=1}^{T_x} D_{X_{t-1}X_t}$$

the expectation referring to the random walk.

7.2 Hitting times and Thompson's principle

Using the commute interpretation of resistance (Corollary 8) to translate Thompson's principle into an assertion about mean commute times gives the following.

Corollary 36 *For random walk on a weighted graph and distinct vertices* v, a

$$E_v T_a + E_a T_v = w \ \inf\{\frac{1}{2}\sum_i \sum_j f_{ij}^2 / w_{ij} : \mathbf{f} \textit{ unit flow } a \textit{ to } v\}$$

and the min *is attained by the flow* $\mathbf{f}^{a\to v}$ *associated with the random walk.*

Comparing with Theorem 32 we have two different extremal characterizations of mean commute times, as a *sup* over potential functions and as an *inf* over flows. In practice this "flow" form is less easy to use than the "potential" form, because writing down a flow $\mathbf{f}$ is harder than writing down a function g. But, when we can write down and calculate with some plausible flow, it gives upper bounds on mean commute times.

One-sided mean hitting times $E_i T_j$ don't have simple extremal characterizations of the same kind, with the exception of hitting times from stationarity. To state the result, fix a state a and define the flow $\mathbf{f}^{a\to\pi}$ by

$$f_{ij} = \lim_{t_0\to\infty} E_a \sum_{t=1}^{t_0} \left(1_{(X_{t-1}=i, X_t=j)} - 1_{(X_{t-1}=j, X_t=i)}\right) \qquad (88)$$

with the usual convention in the aperiodic case. So f_{ij} is the mean excess of transitions $i \to j$ compared to transitions $j \to i$, for the chain started at v_0 and run for ever. This must be a unit flow from a to π, in the sense of section 3.1. Chapter 2 Lemma yyy and reversibility give the first equality in

$$\begin{aligned} f_{ij} &= Z_{ai}p_{ij} - Z_{aj}p_{ji} \\ &= \frac{Z_{ia}\pi_i p_{ij}}{\pi_a} - \frac{Z_{ja}\pi_j p_{ji}}{\pi_a} \\ &= \frac{(Z_{ia} - Z_{ja})\pi_i p_{ij}}{\pi_a} \\ &= \frac{(E_jT_a - E_iT_a)w_{ij}}{w_a} \end{aligned} \tag{89}$$

switching to "weighted graphs" notation. Note also the the first-step recurrence for the function $i \to Z_{ia}$ is

$$Z_{ia} = \left(1_{(i=a)} - \pi_a\right) + \sum_j p_{ij} Z_{ja}. \tag{90}$$

Proposition 37 *For random walk on a weighted graph and a subset A of vertices,*

$$E_\pi T_a = w \ \inf\{\frac{1}{2}\sum_i \sum_j f_{ij}^2/w_{ij} : \mathbf{f} \text{ a unit flow from } A \text{ to } \pi\}$$

$$= \sup\{1/\mathcal{E}(g,g) : -\infty \le g \le \infty, g(\cdot) = 1 \text{ on } A, \sum_i \pi_i g(i) = 0\}.$$

When A is a singleton $\{a\}$, the minimizing flow is the flow $\mathbf{f}^{a\to\pi}$ defined above, and the maximizing function g is $g(i) = Z_{ia}/Z_{aa}$.

Proof. Suppose $A = \{a\}$. We start by showing that the extremizing flow, $\mathbf{f}^*$ say, is of the form stated. By considering adding a flow of size ε along a directed cycle, and copying the argument for (87) in the proof of Proposition 35, there must exist a function g^* such that

$$g^*(x) - g^*(z) = \frac{f^*_{xz}}{w_{xz}} \text{ for each edge } (x,z). \tag{91}$$

The fact that $\mathbf{f}^*$ is a unit flow from a to π says that

$$1_{(x=a)} - \pi_x = \sum_z f^*_{xz} = \sum_z w_{xz}(g^*(x) - g^*(z))$$

which implies

$$\frac{1_{(x=a)} - \pi_x}{w_x} = \sum_z p_{xz}(g^*(x) - g^*(z)) = g^*(x) - \sum_z p_{xz} g^*(z).$$

Since $w_x = w\pi_x$ and $1/\pi_a = E_a T_a^+$, these become

$$g^*(x) = \sum_z p_{xz} g^*(z) - w^{-1}(1 - (E_a T_a^+) 1_{(x=a)}).$$

Now these equations have a unique solution g^*, up to an additive constant, because the difference between two solutions is a harmonic function. On the other hand, a solution is $g^*(x) = -\frac{E_x T_a}{w}$, by considering the first-step recurrence for $E_x T_a$. So by (91) $f^*_{xz} = (E_z T_a - E_x T_a) w_{xz}/w$, and so $\mathbf{f}^* = \mathbf{f}^{a\to\pi}$ by (89).

Now consider the function g which minimizes $\mathcal{E}(g,g)$ under the constraints $\sum_i \pi_i g(i) = 0$ and $g(a) = 1$. By introducing a Lagrange multiplier γ we may consider g as minimizing $\mathcal{E}(g,g) + \gamma \sum_i \pi_i g(i)$. Repeating the argument at (82), the minimizing g satisfies

$$-2\sum_{k\neq j} \pi_j p_{jk}(g(k) - g(j)) \ + \gamma\pi_j = 0, \ j \neq a.$$

Rearranging, and introducing a term $\beta 1_{(j=a)}$ to cover the case $j = a$, we have

$$g(j) = \sum_k p_{jk} g(k) \ + \gamma/2 + \beta 1_{(j=a)} \text{ for all } j,$$

for some γ, β. Because $\sum_j \pi_j g(j) = 0$ we have

$$0 = 0 + \gamma/2 + \beta\pi_a,$$

allowing us to rewrite the equation as

$$g(j) = \sum_k p_{jk} g(k) \ + \beta(1_{(j=a)} - \pi_a).$$

By the familiar "harmonic function" argument this has a unique solution, and (90) shows the solution is $g(j) = \beta Z_{ja}$. Then the constraint $g(a) = 1$ gives $g(j) = Z_{ja}/Z_{aa}$.

Next consider the relationship betrween the flow $\mathbf{f} = \mathbf{f}^{a\to\pi}$ and the function $g(i) = Z_{ia}/Z_{aa}$. We have

$$w\,\frac{f_{ij}^2}{w_{ij}} \ = \ (E_j T_a - E_i T_a)^2 \frac{w_{ij}}{w} \text{ by (89}$$

$$
\begin{aligned}
&= (E_\pi T_a)^2 \frac{w_{ij}}{w} \left(\frac{E_j T_a - E_i T - a}{E_\pi T_a} \right)^2 \\
&= (E_\pi T_a)^2 \frac{w_{ij}}{w} \left(\frac{Z_{ia} - Z_{ja}}{Z_{aa}} \right)^2 \\
&= (E_\pi T_a)^2 \frac{w_{ij}}{w} \, (g(i) - g(j))^2.
\end{aligned}
$$

Thus it is enough to prove

$$
w \sum_i \sum_j f_{ij}^2 / w_{ij} = E_\pi T_a \tag{92}
$$

and it will then follow that

$$
1/\mathcal{E}(g,g) = E_\pi T_a.
$$

To prove (92), introduce a parameter ε (which will later go to 0) and a new vertex z and edge-weights $w_{iz} = \varepsilon w_i$. Writing superscripts $^\varepsilon$ to refer to this new graph and its random walk, Corollary 36 says

$$
E_a T_z^\varepsilon + E_z T_a^\varepsilon = w^\varepsilon \frac{1}{2} \sum_i \sum_j (f_{ij}^\varepsilon)^2 / w_{ij} \tag{93}
$$

where $\mathbf{f}^\varepsilon$ is the unit flow from a to z associated with the new graph. We want to interpret these quantities in terms of the original graph. Clearly $w^\varepsilon = w(1+2\varepsilon)$. The new walk has chance $\frac{\varepsilon}{1+\varepsilon}$ to jump to z from each other vertex, so $E_a T_z^\varepsilon = \frac{1+\varepsilon}{\varepsilon}$. Starting from z, after one step the new walk has the stationary distribution π on the original graph, and it easily follows that $E_z T_a^\varepsilon = 1 + E_\pi T_a (1 + O(\varepsilon))$. We can regard the new walk up to time T_z as the old walk sent to z at a random time U^ε with geometric$(\varepsilon/(1+\varepsilon))$ distribution, so for $i, j \neq z$ the flow f_{ij}^ε is the expected net number of transitions $i \to j$ up to time U^ε, which by the convergence theorem converges to f_{ij} exponentially fast (in $1/\varepsilon$). And $f_{iz}^\varepsilon = P_a(X(U^\varepsilon - 1) = i)$, which converges to π_i exponentially fast. So up to $+o(1)$ terms, (93) becomes

$$
\frac{1+\varepsilon}{\varepsilon} + 1 + E_\pi T_a = w(1+2\varepsilon) \left(\frac{1}{2} \sum_i \sum_j f_{ij}^2 / w_{ij} \; + \frac{1}{w\varepsilon} \right)
$$

where we snuck in the calculation $\sum_i \frac{\pi_i^2}{\varepsilon w_i} = \frac{1}{w\varepsilon}$. Letting $\varepsilon \to 0$ gives the result.

xxx say above proof better.

xxx case of general A.

Corollary 38 *For chains with transition matrices* $\mathbf{P}, \tilde{\mathbf{P}}$ *and the same stationary distribution* π,

$$\min_{i \neq j} \frac{p_{ij}}{\tilde{p}_{ij}} \leq \frac{E_\pi T_a}{E_\pi \tilde{T}_a} \leq \max_{i \neq j} \frac{p_{ij}}{\tilde{p}_{ij}}.$$

Proof. Plug the minimizing flow $\mathbf{f}^{a \to \pi}$ for the $\mathbf{P}$-chain into Proposition 37 for the $\tilde{\mathbf{P}}$-chain.

8 Notes on Chapter 3

Textbooks. Almost all the results have long been known to (different groups of) experts, but it has not been easy to find accessible textbook treatments. Of the three books on reversible chains at roughly the same level of sophistication as ours, Kelly [26] emphasizes stationary distributions of stochastic networks (cf. our Chapter yyy); Keilson [25] emphasizes mathematical properties such as complete monotonicity; and Chen [11] discusses those aspects useful in the study of interacting particle systems.

Section 1. In abstract settings reversible chains are called *symmetrizable*, but that's a much less evocative term. Elementary textbooks often give *Kolmogorov's criterion* ([26] Thm 1.7) for reversibility, but I've never found it to be useful.

Section 1.1. Though probabilists would regard the "cyclic tour" Lemma 1 as obvious, Laszlo Lovasz pointed out a complication, that with a careful definition of starts and ends of tours these times are not invariant under time-reversal. The sophisticated fix is to use doubly-infinite stationary chains and observe that tours in reversed time just interleave tours in forward time, so by ergodicity their asymptotic rates are equal. Tetali [31] shows that the cyclic tour property implies reversibility.

Cat-and-mouse game 1 is treated more opaquely in Coppersmith et al. [13], whose deeper results are discussed at yyy. Underlying the use of the optional sampling theorem in game 2 is a general result about optimal stopping, but it's much easier to prove what we want here than to appeal to general theory. Several algorithmic variations on Proposition 2 are discussed in Coppersmith et al [14] and Tetali and Winkler [32].

Section 2. Many textbooks on Markov chains note the simple explicit form of the stationary distribution for random walks on graphs. A historical note (taken from [15]) is that the first explicit treatment of random walk on a general finite graph was apparently given in 1935 by Bottema [6], who proved the convergence theorem. Amongst subsequent papers specializing Markov theory to random walks on graphs let us mention Gobel and Jagers [21], which contains a variety of the more elementary facts given in this book, for instance the unweighted version of Lemma 6. Another observation from [21] is that for a reversible chain the quantity

$$\beta_{ijl} \equiv \pi_j^{-1} E_i(\text{ number of visits to } j \text{ before time } T_l \text{ })$$

satisfies $\beta_{ijl} = \beta_{jil}$. Indeed, by Chapter 2 Lemma yyy we have

$$\beta_{ijl} = (E_i T_l + E_l T_j + E_j T_i) - (E_j T_i + E_i T_j)$$

and so the result follows from the cyclic tour property.

Just as random walks on undirected graphs are as general as reversible Markov chains, so random walks on *directed* graphs are as general as general Markov chains. In particular, one usually has no simple expression like (13) for the stationary distribution. The one tractable case is a *balanced* directed graph, where the in-degree d_v of each vertex v equals its out-degree. See Chapter yyy for further discussion of this case.

Section 2.1. Yet another way to associate a continuous-time reversible chain with a weighted graph is to set $q_{ij} = w_{ij}/\sqrt{w_i w_j}$. This construction was used by Chung and Yau [12] as the simplest way to set up discrete analogs of certain results from differential geometry.

Section 3. Doyle and Snell [17] gave a detailed elementary textbook exposition of Proposition 7 and the whole random walk – electrical network connection. Previous brief textbook accounts were given by Kemeny et al [27] and Kelly [26]. Our proof of Proposition 7 is not intuitively convincing – the reader can decide whether the different proofs in the references above are more convincing. As mentioned in the text, the first explicit use (known to me) of the mean commute interpretation was given by Chandra et al [10]. One can combine the commute formula with the general identities of Chapter 2 to obtain numerous identities relating mean hitting times and resistances, some of which are given (using bare-hands proofs instead) in Tetali [30]. The connection between Foster's theorem and Lemma 6 was noted in [13].

Section 4. The spectral theory is of course classical. In devising a symmetric matrix one could use $\pi_i p_{ij}$ or $p_{ij}\pi_j^{-1}$ instead of $\pi_i^{1/2} p_{ij} \pi_j^{-1/2}$ – there doesn't seem any systematic advantage to a particular choice. I learned the eigentime identity from Andrei Broder who used it in [7], and Lemma 12 from David Zuckerman who used it in [33]. I don't know whether Lemma

12 holds for general chains. Mark Brown (personal communication) has noted several variations on the theme of Lemma 12, for example that the unweighted average of $(E_i T_j; i, j \in A)$ is bounded by the unweighted average of $(E_\pi T_j; j \in A)$. Corollary 11 was exploited in xxx. The name *eigentime identity* is my own coinage: once we call $1/\lambda_2$ the relaxation time it is natural to start thinking of the other $1/\lambda_m$ as "eigentimes".

Section 5. We regard complete monotonicity as a name for "mixtures of exponentials", and have not used the analytic characterization via derivatives of alternating signs. Of course the CM property is implicit in much analysis of reversible Markov processes, but I find it helpful to exhibit explicitly its use in obtaining inequalities. This idea in general, and in particular the "stochastic ordering of exit times" result (Proposition 18), were first emphasized by Keilson [25] in the context of reliability and queueing models. Brown [9] gives other interesting consequences of monotonicity.

Section 5.1. Parts of Proposition 15 have been given by several authors, e.g. Broder and Karlin [7] Corollary 18 give (48). One can invent many variations. Consider for instance $\min_i \max_j E_i T_j$. On the complete graph this equals $n-1$, but this is not the minimum value, as observed by Erik Ordentlich in a homework exercise. If we take the complete graph, distinguish a vertex i_0, let the edges involving i_0 have weight ε and the other edges have weight 1, then as $\varepsilon \to 0$ we have (for $j \neq i_0$)

$$E_{i_0} T_j \to \frac{1}{n-1} + \frac{n-2}{n-1}(1 + (n-2)) = n - 2 + \frac{1}{n-1}.$$

By the random target lemma and (48), the quantity under consideration is at least $\tau_0 \geq n - 2 + \frac{1}{n}$, so the example is close to optimal.

Section 5.4. The simple result quoted as Proposition 19 is actually weaker than the result proved in Brown [8]. The ideas in the proof of Proposition 20 are in Aldous [2] and in Brown [9], the latter containing a shorter Laplace transform argument for (67). Aldous and Brown [3] give a more detailed account of the exponential approximation, including the following result which is useful in precisely the situation where Proposition 20 is applicable, that is when $E_\pi T_A$ is large compared to τ_2.

Theorem 39 *Let α_A be the quasistationary distribution on A^c defined at (75). Then*

$$P_\pi(T_A > t) \geq \left(1 - \frac{\tau_2}{E_{\alpha_A} T_A}\right) \exp\left(\frac{-t}{E_{\alpha_A} T_A}\right), \ t > 0.$$

$$E_\pi T_A \geq E_{\alpha_A} T_A - \tau_2.$$

Using this requires only a *lower* bound on $E_\alpha T_A$, which can often be obtained using the extremal characterization (76).

For general chains, explicit bounds on exponential approximation are much messier: see Aldous [1] for a bound based upon total variation mixing and Iscoe and McDonald [24] for a bound involving spectral gaps.

Section 6.1. Dirichlet forms were developed for use with continuous-space continuous-time Markov processes, where existence and uniqueness questions can be technically difficult – see e.g. Fukushima [20]. Their use subsequently trickled down to the discrete world, influenced e.g. by the paper of Diaconis and Stroock [16]. Chen [11] is the most accessible introduction.

Since mean commute times have two dual extremal characterizations, as *sups* over potential functions and as *infs* over flows, it is natural to ask

Open Problem 40 Does there exist a characterization of the relaxation time as exactly an *inf* over flows?

We will see in Chapter 4 yyy an inequality giving an upper bound on the relaxation time in terms of an *inf* over flows, but it would be more elegant to derive such inequalities from some exact characterization.

Lemma 29 is sometimes used to show, by comparison with the i.i.d. chain,

$$\text{if } \max_{j \neq i} p_{ij}/\pi_j = \delta > 0 \text{ then } \tau_2 \leq \delta^{-1}.$$

But this is inefficient: direct use of submultiplicitivity of variation distance gives a stronger conclusion.

Quasistationary distributions for general chains have long been studied in applied probability, but the topic lacks a good survey article. Corollary 31 is a good example of a repeatedly-rediscovered simple-yet-useful result which defies attempts at attribution.

Section 7. Use of Thompson's principle and the Dirichlet principle to study transience / recurrence of countable infinite state space chains is given an elementary treatment in Doyle and Snell [17] and more technical treatments in papers of Nash-Williams [29], Griffeath and Liggett [22] and Lyons [28]. Some reformulations of Thompson's principle are discussed by Berman and Konsowa [4]. I learned about the work of Borre-Meissl [5] on leveling networks from Persi Diaconis.

I've never seen in the literature an explicit statement of the extremal characterizations for mean hitting times from a stationary start (Proposition 37), but these are undoubtedly folklore. Steve Evans once showed me an argument for Corollary 38 based on the usual Dirichlet principle, and that motivated me to present the "natural explanation" given by Proposition 37.

References

[1] D.J. Aldous. Markov chains with almost exponential hitting times. *Stochastic Process. Appl.*, 13:305–310, 1982.

[2] D.J. Aldous. Hitting times for random walks on vertex-transitive graphs. *Math. Proc. Cambridge Phil. Soc.*, 106:179–191, 1989.

[3] D.J. Aldous and M. Brown. Inequalities for rare events in time-reversible Markov chains I. In M. Shaked and Y.L. Tong, editors, *Stochastic Inequalities*, volume 22 of *Lecture Notes*, pages 1–16. Institute of Mathematical Statistics, 1992.

[4] K.A. Berman and M. Konsowa. Random paths, electrical networks and reversible Markov chains. *SIAM J. Discrete Math.*, 3:311–319, 1990.

[5] K. Borre and P. Meissl. *Strength Analysis of Leveling-type Networks.* Geodaetisk Institut, Copenhagen, 1974. Volume 50.

[6] O. Bottema. Uber die Irrfahrt in einem Strassennetz. *Math. Z.*, 39:137–145, 1935.

[7] A. Broder and A.R. Karlin. Bounds on the cover time. *J. Theoretical Probab.*, 2:101–120, 1989.

[8] M. Brown. Approximating IMRL distributions by exponential distributions, with applications to first passage times. *Ann. Probab.*, 11:419–427, 1983.

[9] M. Brown. Consequences of monotonicity for Markov transition functions. Technical report, City College, CUNY, 1990.

[10] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky, and P. Tiwari. The electrical resistance of a graph captures its commute and cover times. In *Proc. 21st ACM Symp. Theory of Computing*, pages 574–586, 1989.

[11] M. F. Chen. *From Markov Chains to Non-Equilibrium Particle Systems.* World Scientific, Singapore, 1992.

[12] F.R.K. Chung and S.-T. Yau. Eigenvalues of graphs and Sobolev inequalities. Technical report, BELLCORE, 1993.

[13] D. Coppersmith, P. Doyle, P. Raghavan, and M. Snir. Random walks on weighted graphs and applications to on-line algorithms. In *Proc. 22nd ACM Symp. Theory of Computing*, pages 369–378, 1990.

[14] D. Coppersmith, P. Tetali, and P. Winkler. Collisions among random walks on a graph. *SIAM J. Disc. Math.*, 6:363–374, 1993.

[15] D.M. Cvetkovic, M. Doob, and H. Sachs. *Spectra of Graphs.* Academic Press, 1980.

[16] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, 1:36–61, 1991.

[17] P.G. Doyle and J.L. Snell. *Random Walks and Electrical Networks.* Mathematical Association of America, Washington DC, 1984.

[18] R. Durrett. *Probability: Theory and Examples.* Wadsworth, Pacific Grove CA, 1991.

[19] R.M. Foster. The average impedance of an electrical network. In *Contributions to Applied Mechanics*, pages 333–340, Ann Arbor, MI, 1949. Edwards Brothers, Inc.

[20] M. Fukushima. *Dirichlet Forms and Markov Processes.* North-Holland, 1980.

[21] F. Gobel and A.A. Jagers. Random walks on graphs. *Stochastic Process. Appl.*, 2:311–336, 1974.

[22] D. Griffeath and T. M. Liggett. Critical phenomena for Spitzer's reversible nearest particle systems. *Ann. Probab.*, 10:881–895, 1982.

[23] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[24] I. Iscoe and D. McDonald. Asymptotics of exit times for Markov jump processes I. *Ann. Probab.*, 22:372–397, 1994.

[25] J. Keilson. *Markov Chain Models - Rarity and Exponentiality.* Springer-Verlag, 1979.

[26] F.P. Kelly. *Reversibility and Stochastic Networks.* Wiley, 1979.

[27] J.G. Kemeny, J.L Snell, and A.W. Knapp. *Denumerable Markov Chains.* Springer-Verlag, 2nd edition, 1976.

[28] T.J. Lyons. A simple criterion for transience of a reversible Markov chain. *Ann. Probab.*, 11:393–402, 1983.

[29] C. St. J. A. Nash-Williams. Random walks and electric currents in networks. *Proc. Cambridge Philos. Soc.*, 55:181–194, 1959.

[30] P. Tetali. Random walks and the effective resistance of networks. *J. Theoretical Probab.*, 4:101–109, 1991.

[31] P. Tetali. An extension of Foster's network theorem. *Combin. Probab. Comput.*, 3:421–427, 1994.

[32] P. Tetali and P. Winkler. Simultaneous reversible Markov chains. In *Combinatorics: Paul Erdos is Eighty*, volume 1, pages 433–451. Bolyai Society Mathematical Studies, 1993.

[33] D. Zuckerman. Covering times of random walks on bounded degree trees and other graphs. *J. Theoretical Probab.*, 2:147–157, 1989.