Chapter 8
Advanced L^2 Techniques for Bounding Mixing Times

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

James Allen Fill
Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, MD 21218-2682

May 19, 1999

xxx In next revision, we should change the definition [in Chapter 4, yyy:(14)] **of $\hat{d}(t)$ so that what is now $\sqrt{\hat{d}(2t)}$ becomes $\hat{d}(t)$.**

This chapter concerns advanced L^2-based techniques, developed mainly by Persi Diaconis and Laurent Saloff-Coste [2, 3, 4, 5] for bounding mixing times for (finite, irreducible) reversible Markov chains. For convenience, we will work in continuous time throughout this chapter, unless otherwise noted. Many of the results are conveniently expressed in terms of an "L^2 threshold time" $\hat{\tau}$ (xxx use different notation?) defined by

$$\hat{\tau} := \inf\{t > 0 : \max_i \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le e^{-1}\}. \tag{1}$$

xxx For NOTES: Discussion of discrete time, esp. negative eigenvalues.

Several preliminary comments are in order here. First, the definition of the L^2 distance $\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2$ may be recalled from Chapter 2 section yyy:6.2, and Chapter 3 yyy:(55) and the spectral representation give useful reexpressions:

$$\begin{aligned} \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2^2 &= \sum_j \pi_j \left(\frac{p_{ij}(t)}{\pi_j} - 1\right)^2 \\ &= \frac{p_{ii}(2t)}{\pi_i} - 1 \qquad (2) \\ &= \pi_i^{-1} \sum_{m=2}^{n} \exp(-2\lambda_m t) u_{im}^2. \end{aligned}$$

Second, from (2) and Chapter 4 yyy:(14) we may also write the maximum L^2 distance appearing in (1) using

$$\max_i \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2^2 = \max_i \frac{p_{ii}(2t)}{\pi_i} - 1 = \max_{i,j} \frac{p_{ij}(2t)}{\pi_j} - 1 = \hat{d}(2t).$$

Third, by the application of the Cauchy–Schwarz lemma in Chapter 4 Lemma yyy:8, variation distance can be bounded by L^2 distance:

$$\begin{aligned} 4d_i^2(t) &:= 4\|P_i(X_t \in \cdot) - \pi(\cdot)\|^2 \le \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2^2, &(3)\\ 4d^2(t) &:= 4\max_i \|P_i(X_t \in \cdot) - \pi(\cdot)\|^2 \le \hat{d}(2t); &(4)\end{aligned}$$

these inequalities are the primary motivation for studying L^2 distance.

As argued in Chapter 4 yyy:just following (23),

$$\hat{d}(2t) \le \pi_*^{-1} e^{-2t/\tau_2}, \tag{5}$$

where $\tau_2 := \lambda_2^{-1}$ is the relaxation time and $\pi_* := \min_i \pi_i$. Thus if

$$t \ge \tau_2 \left(\frac{1}{2} \log \frac{1}{\pi_*} + c \right),$$

then

$$d(t) \le \tfrac{1}{2}\sqrt{\hat{d}(2t)} \le \tfrac{1}{2} e^{-c}, \tag{6}$$

which is small if c is large; in particular, (6) gives the upper bound in

$$\tau_2 \le \hat{\tau} \le \tau_2 \left(\frac{1}{2} \log \frac{1}{\pi_*} + 1 \right), \tag{7}$$

and the lower bound follows easily.

For many simple chains (see Chapter 5), τ_2 can be computed exactly. Typically, however, τ_2 can only be bounded. This can be done using the "distinguished paths" method of Chapter 4 Section yyy:3. In Section 1 we will see that that method may be regarded as a special case of a "comparison method" whereby a chain with "unknown" relaxation time is compared to a second chain with "known" relaxation time. The greater generality often leads to improved bounds on τ_2. As a bonus, the comparison method also gives bounds on the other "unknown" eigenvalues, and such bounds in turn can sometimes further decrease the time t required to guarantee that $\hat{d}(2t)$, and hence also $d(t)$, is small.

A second set of advanced techniques, encompassing the notions of Nash inequalities, moderate growth, and local Poincaré inequaltities, is described in Section 3. The development there springs from the inequality

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \leq N(s) e^{-(t-s)/\tau_2}, \tag{8}$$

established for all $0 \leq s \leq t$ in Section 2, where

$$N(t) = \max_i \|P_i(X_t \in \cdot)\|_2 = \max_i \sqrt{\frac{p_{ii}(2t)}{\pi_i}}, \quad t \geq 0. \tag{9}$$

Choosing $s = 0$ in (8) gives

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \leq \pi_i^{-1/2} e^{-t/\tau_2},$$

and maximizing over i recaptures (5). The point of Section 3, however, is that one can sometimes reduce the bound by a better choice of s and suitable estimates of the decay rate of $N(\cdot)$. Such estimates can be provided by so-called Nash inequalities, which are implied by (1) moderate growth conditions and (2) local Poincaré inequalities. Roughly speaking, for chains satisfying these two conditions, judicious choice of s shows that variation mixing time and $\hat{\tau}$ are both of order Δ^2, where Δ is the diameter of the graph underlying the chain.

xxx Might not do (1) or (2), so need to modify the above.

To outline a third direction of improvement, we begin by noting that neither of the bounds in (7) can be much improved in general. Indeed, ignoring $\Theta(1)$ factors as usual, the lower bound is equality for the n-cycle (Chapter 5, Example yyy:7) and the upper bound is equality for the M/M/1/n queue (Chapter 5, Example yyy:6) with traffic intensity $\rho \in (0,1)$.

In Section 4 we introduce the *log-Sobolev time* τ_l defined by

$$\tau_l := \sup\{\mathcal{L}(g)/\mathcal{E}(g,g) : g \not\equiv \text{constant}\} \tag{10}$$

where $\mathcal{L}(g)$ is the entropy-like quantity

$$\mathcal{L}(g) := \sum_i \pi_i g^2(i) \log(|g(i)|/\|g\|_2),$$

recalling $\|g\|_2^2 = \sum_i \pi_i g^2(i)$. Notice the similarity between (10) and the extremal characterization of τ_2 (Chapter 3 Theorem yyy:22):

$$\tau_2 = \sup\{\|g\|_2^2/\mathcal{E}(g,g) : \sum_i \pi_i g(i) = 0, \quad g \not\equiv 0\}.$$

We will see that

$$\tau_2 \le \tau_l \le \tau_2 \frac{\log\left(\frac{1}{\pi_*} - 1\right)}{2(1-2\pi_*)}$$

and that $\hat{\tau}$ is more closely related to τ_l than to τ_2, in the sense that

$$\tau_l \le \hat{\tau} \le \tau_l \left(\frac{1}{2}\log\log\frac{1}{\pi_*} + 2\right). \tag{11}$$

To illustrate the improvement over (7), from the knowledge for the d-cube (Chapter 5, Example yyy:15) that $\tau_2 = d/2$, one can deduce from (7) that

$$\tfrac{1}{2}d \le \hat{\tau} \le \tfrac{1}{4}(\log 2)d^2 + \tfrac{1}{2}d. \tag{12}$$

In Section 4.4 (Example 27) we will see that $\tau_l = d/2$; then from (11) we can deduce the substantial improvement

$$\frac{1}{2}d \le \hat{\tau} \le \frac{1}{4}d\log d + \left(1 - \frac{1}{4}\log\frac{1}{\log 2}\right)d \tag{13}$$

upon (12).

ZZZ!: Recall also the corrections in my notes on pages 8.2.11–12 (and 8.4.27). Continue same paragraph:

The upper bound here is remarkably tight: from Chapter 5 yyy:(65),

$$\hat{\tau} = \frac{1}{4}d\log d + \left(\frac{1}{4}\log\frac{1}{\log(1+e^{-2})}\right)d + o(d) \quad \text{as } d \to \infty.$$

ZZZ!: In fact, the remainder term is $O(1)$. Continue same paragraph:

Thus log-Sobolev techniques provide another means of improving mixing time bounds, both in L^2 and, because of (3)–(4), in variation. As will be seen, these techniques can also be combined usefully with comparison methods and Nash inequalities.

1 The comparison method for eigenvalues

xxx Revise Chapter 7, Sections 1.9 and 4, in light of this section?

The comparison method, introduced by Diaconis and Saloff-Coste [2, 3], generalizes the distinguished path method of Chapter 4, Section yyy:3 for bounding the relaxation time of a reversible Markov chain. As before, we first (xxx: delete word?) work in the setting of random walks on weighted graphs. We will proceed for given state space (vertex set) I by comparing a collection (w_{ij}) of weights of interest to another collection $(\tilde{w}_{ij})$; the idea will be to use known results for the random walk with weights $(\tilde{w}_{ij})$ to derive corresponding results for the walk of interest. We assume that the graph is connected under each set of weights. As in Chapter 4, Section yyy:4.3, we choose ("distinguish") paths γ_{xy} from x to y. Now, however, this need be done only for those (x, y) with $x \neq y$ and $\tilde{w}_{xy} > 0$, but we impose the additional constraint $w_e > 0$ for each edge e in the path. (Here and below, e denotes a directed edge in the graph of interest.) In other words, roughly put, we need to construct a (w_{ij})-path to effect each given $(\tilde{w}_{xy})$-edge. Recall from Chapter 3 yyy:(71) the definition of Dirichlet form:

$$\begin{aligned} \mathcal{E}(g,g) &= \frac{1}{2}\sum_i \sum_{j \neq i} \frac{w_{ij}}{w}(g(j)-g(i))^2, \qquad (14) \\ \tilde{\mathcal{E}}(g,g) &= \frac{1}{2}\sum_i \sum_{j \neq i} \frac{\tilde{w}_{ij}}{\tilde{w}}(g(j)-g(i))^2. \end{aligned}$$

Theorem 1 (comparison of Dirichlet forms) *For each ordered pair (x, y) of distinct vertices with $\tilde{w}_{xy} > 0$, let γ_{xy} be a path from x to y with $w_e > 0$ for every $e \in \gamma_{xy}$. Then the Dirichlet forms (14) satisfy*

$$\tilde{\mathcal{E}}(g,g) \leq A\mathcal{E}(g,g) = \mathcal{E}(g,g)\frac{w}{\tilde{w}}\max_e \frac{1}{w_e}\sum_x \sum_{y \neq x} \tilde{w}_{xy}|\gamma_{xy}| 1_{(e \in \gamma_{xy})}$$

for every g.

Proof. For an edge $e = (i, j)$ write $\Delta g(e) = g(j) - g(i)$. Then

$$\begin{aligned} 2\tilde{w}\tilde{\mathcal{E}}(g,g) &= \sum_x \sum_{y \neq x} \tilde{w}_{xy}(g(y)-g(x))^2 \\ &= \sum_x \sum_{y \neq x} \tilde{w}_{xy}\left(\sum_{e \in \gamma_{xy}} \Delta g(e)\right)^2 \\ &\leq \sum_x \sum_{y \neq x} \tilde{w}_{xy}|\gamma_{xy}| \sum_{e \in \gamma_{xy}} (\Delta g(e))^2 \quad \text{by Cauchy–Schwarz} \\ &\leq A\sum_e w_e(\Delta g(e))^2 = A \cdot 2w\mathcal{E}(g,g). \quad \blacksquare \end{aligned}$$

Remarks. (a) Suppose the comparison weights $(\tilde{w}_{ij})$ are given by

$$\tilde{w}_{ij} = w_i w_j / w \quad \text{for } i, j \in I.$$

The corresponding discrete-time random walk is then the “trivial”walk with $\tilde{w} = w$ and

$$\tilde{w}_i = w_i, \quad \tilde{p}_{ij} = \pi_j, \quad \tilde{\pi}_j = \pi_j$$

for all i, j, and

$$\begin{aligned} \tilde{\mathcal{E}}(g,g) &= \frac{1}{2}\sum_i \sum_{j \neq i} \pi_i \pi_j (g(j) - g(i))^2 = \operatorname{var}_\pi g \\ &= \|g\|_2^2 \quad \text{provided } \textstyle\sum_i \pi_i g(i) = 0. \end{aligned}$$

In this case the conclusion of Theorem 1 reduces to

$$\|g\|_2^2 \leq \mathcal{E}(g,g) w \max_e \tfrac{1}{w_e} \sum_x \sum_{y \neq x} \pi_x \pi_y |\gamma_{xy}| 1_{(e \in \gamma_{xy})}.$$

This inequality was established in the proof of the distinguished path theorem (Chapter 4 Theorem yyy:32), and that theorem was an immediate consequence of the inequality. Hence the comparison Theorem 1 may be regarded as a generalization of the distinguished path theorem.

[xxx For NOTES: We've used simple Sinclair weighting. What about other weighting in use of Cauchy–Schwarz? Hasn't been considered, as far as I know.]

(b) When specialized to the setting of reversible random flights on Cayley graphs described in Chapter 7 Section yyy:1.9, Theorem 1 yields Theorem yyy:14 of Chapter 7. To see this, adopt the setup in Chapter 7 Section yyy:1.9, and observe that the word

$$x = g_1 g_2 \cdots g_d \quad \text{(with each } g_i \in \mathcal{G}) \tag{15}$$

corresponds uniquely to a path

$$\gamma_{\mathrm{id},x} = (\mathrm{id}, g_1, g_1 g_2, \ldots, g_1 g_2 \cdots g_d = x) \tag{16}$$

in the Cayley graph corresponding to the generating set $\mathcal{G}$ of interest. Having built paths $\gamma_{\mathrm{id},x}$ for each $x \in I$, we then can build paths γ_{yz} for $y, z \in I$ by exploiting vertex-transitivity, to wit, by setting

$$\gamma_{yz} = (y, y g_1, y g_1 g_2, \ldots, y g_1 g_2 \cdots g_d = z)$$

where $y^{-1} z = x$ and the path $\gamma_{\mathrm{id},x}$ is given by (16). In Theorem 1 we then have both stationary distributions π and $\tilde{\pi}$ uniform,

$$\tilde{w}_{xy} = \tilde{\mu}(x^{-1}y)/n, \quad \tilde{w} = 1, \quad |\gamma_{xy}| = |\gamma_{\mathrm{id},x^{-1}y}| = d(\mathrm{id}, x^{-1}y),$$

and, if $e = (v, vg)$ with $v \in I$ and $g \in \mathcal{G}$,

$$w_e = \mu(g)/n, \quad w = 1, \quad 1_{(e \in \gamma_{xy})} = 1_{((x^{-1}v, x^{-1}vg) \in \gamma_{\mathrm{id}, x^{-1}y})}.$$

Thus A of Theorem 1 equals

$$\max_{v \in I,\ g \in \mathcal{G}} \frac{1}{\mu(g)} \sum_x \sum_{y \neq x} \tilde{\mu}(x^{-1}y) d(\mathrm{id}, x^{-1}y) 1_{(x^{-1}v, x^{-1}vg \in \gamma_{\mathrm{id}, x^{-1}y})},$$

which reduces easily to K of Theorem yyy:14 of Chapter 7. Since π and $\tilde{\pi}$ are both uniform, the extremal characterization

$$\tau_2 = \sup\{\|g\|_2^2 / \mathcal{E}(g, g) : \sum_i \pi_i g(i) = 0\} \tag{17}$$

gives Theorem yyy:14 of Chapter 7.

Theorem 1 compares Dirichlet forms. To compare relaxation times using the extremal characterization (17), we compare L^2-norms using the same "direct" technique as for Chapter 3 Lemma yyy:26. For any g,

$$\|g\|_2^2 \leq \|g\|_{\tilde{2}}^2 \max_i (\pi_i / \tilde{\pi}_i) \tag{18}$$

where, as usual, $\pi_i := w_i / w$ and $\tilde{\pi}_i = \tilde{w}_i / \tilde{w}$. So if g has π-mean 0 and $\tilde{\pi}$-mean b, then

$$\frac{\|g\|_2^2}{\mathcal{E}(g, g)} \leq \frac{\|g - b\|_2^2}{\mathcal{E}(g - b, g - b)} \leq \frac{A \|g - b\|_{\tilde{2}}^2}{\tilde{\mathcal{E}}(g - b, g - b)} \max_i (\pi_i / \tilde{\pi}_i). \tag{19}$$

Thus

Corollary 2 (comparison of relaxation times) *In Theorem 1,*

$$\tau_2 \leq \frac{A}{a} \tilde{\tau}_2$$

where

$$\begin{aligned} A &:= \frac{w}{\tilde{w}} \max_e \frac{1}{w_e} \sum_x \sum_{y \neq x} \tilde{w}_{xy} |\gamma_{xy}| 1_{(e \in \gamma_{xy})}, \\ a &:= \min_i (\tilde{\pi}_i / \pi_i). \end{aligned}$$

xxx Perhaps restate as

$$\tau_2 \le \tilde{\tau}_2$$

where

$$B := \left(\max_e \frac{1}{w_e} \sum \sum \cdots \right) \left(\max_i \frac{w_i}{\tilde{w}_i} \right)$$

(and similarly for Corollaries 4 and 7)?

xxx Remark about best if $\pi = \tilde{\pi}$?

Here is a simple example, taken from [2], showing the improvement in Corollary 2 over Theorem yyy:32 of Chapter 4 provided by the freedom in choice of benchmark chain.

xxx NOTE: After the fact, I realized that the following example was already Example yyy:20 of Chapter 7; must reconcile.

Example 3 Consider a card shuffle which transposes the top two cards in the deck, moves the top card to the bottom, or moves the bottom card to the top, each with probability 1/3. This example fits the specialized group framework of Chapter 7 Section yyy:1.9 (see also Remark (b) following Theorem 1 above) with I taken to be the symmetric group on m letters and

$$\mathcal{G} := \{(1\ 2), (m\ m-1\ m-2\ \cdots\ 1), (1\ 2\ \cdots\ m)\}$$

in cycle notation. [If the order of the deck is represented by a permutation σ in such a way that $\sigma(i)$ is the position of the card with label i, and if permutations are composed left to right, then $\sigma \cdot (m\ m-1\ m-2\ \cdots\ 1)$ is the order resulting from σ by moving the top card to the bottom.]

We obtain a representation (15) for any given permutation x by writing

$$x = h_m h_{m-1} \cdots h_2$$

in such a way that

$$(h_m \cdots h_i)^{-1}(j) = x^{-1}(j) \quad \text{for } i \le j \le m \tag{20}$$

(i.e., $h_m \cdots h_i$ and x agree in positions i through m) and each h_i is explicitly represented as a product of generators. To accomplish this, we proceed inductively. Suppose that (20) holds for given $i \in \{3, \ldots, m+1\}$, and that $(h_m \cdots h_i)(x^{-1}(i-1)) = l_i = l$, with $1 \le l \le i-1$. Then let

$$\begin{aligned} h_{i-1} \; := \; & (m\ m-1\ m-2\ \cdots\ 1)^{l-1} [(1\ 2)(m\ m-1\ m-2\ \cdots\ 1)]^{i-l-1} \\ & \cdot (m\ m-1\ m-2\ \cdots\ 1)^{m-i+2}. \end{aligned}$$

In words, beginning with $h_m \cdots h_i$, we repeatedly move the top card to the bottom until card $x^{-1}(i-1)$ has risen to the top; then we repeatedly transpose and shift until the top $m-i+2$ cards, in order, are $x^{-1}(i-1), \ldots, x^{-1}(m)$; and finally we cut these $m-i+2$ cards to the bottom.

xxx Either revise Section 1.9 of Chapter 7 to delete requirement of *geodesic* paths, or explain one can erase cycles.

It follows that the diameter Δ of the Cayley graph associated with $\mathcal{G}$ satisfies

$$\Delta \leq \sum_{i=2}^{m+1} [(l_i - 1) + 2(i - l_i - 1) + (m - i + 2)] \leq 3\binom{m}{2}$$

and so by Chapter 7 Corollary yyy:15 that $\tau_2 \leq 27\binom{m}{2}^2 < \frac{27}{4}m^4$.

To improve this bound on the relaxation time we compare the chain of interest to the random transposition chain of Chapter 7 Example yyy:18 and employ Corollary 2, or rather its specialization, (yyy:Theorem 14) of Chapter 7.

xxx Continue as in Chapter 7 Example yyy:20 to get

$$\frac{\tau_2}{\tilde{\tau}_2} \leq 3\beta^2, \quad \tilde{\tau}_2 = \frac{m}{2}, \quad \tau_2 \leq \frac{27}{2}m^3.$$

xxx Test function on page 2139 of [2] shows this is right order.

Corollary 2 can be combined with the inequality

$$\hat{\tau} \leq \tau_2 \left(\frac{1}{2} \log \frac{1}{\pi_*} + 1 \right) \tag{21}$$

from (7) to bound the L^2 threshold parameter $\hat{\tau}$ for the chain of interest, but Theorem 1 sometimes affords a sharper result. From the Courant–Fischer "min–max" theorem ([9], Theorem 4.2.11) it follows along the same lines as in Chapter 3 Section yyy:6.3 that

$$\lambda^{-1} = \inf \rho(h_1, h_2, \ldots, h_{m-1}), \quad m = 2, \ldots, n, \tag{22}$$

where $h_1 \equiv 1$ and xxx Say the conditions better!

$$\begin{aligned} &\rho(h_1, h_2, \ldots, h_{m-1}) \\ &:= \sup\{\|g\|_2^2 / \mathcal{E}(g,g) : \sum_i \pi_i h_j(i) g(i) = 0 \quad \text{for } j = 1, \ldots, m-1\} \end{aligned}$$

and the inf in (22) is taken over all vectors $h_1, \ldots, h_{m-1}$ that are orthogonal in $L^2(\pi)$ (or, equivalently, that are linearly independent). Using (19), Corollary 2 now generalizes to

Corollary 4 (comparison of eigenvalues) *In Theorem 1, the eigenvalues λ_m and $\tilde{\lambda}_m$ in the respective spectral representations satisfy*

$$\lambda_m^{-1} \le \frac{A}{a} \tilde{\lambda}_m^{-1}$$

with A and a as defined in Corollary 2.

Here is a simple example [3] not possessing vertex-transitivity:

xxx NOTE: This is a DIRECT comparison!: see Chapter 3 Section 6.4.

Example 5 *Random walk on a d-dimensonal grid.*

To keep the notation simple, we let $d = 2$ and consider the grid $I := \{0, \ldots, m_1 - 1\} \times \{0, \ldots, m_2 - 1\}$ as an (unweighted) subgraph of $\mathbf{Z}^2$. The eigenvalues λ_l are not known in closed form. However, if we add self-loops to make a benchmark graph where I is regular with degree 4, then the eigenvalues $\tilde{\lambda}_l$ for the continuous-time walk are

$$1 - \frac{1}{2}\left(\cos\frac{\pi r}{m_1} + \cos\frac{\pi s}{m_2}\right), \quad 0 \le r \le m_1 - 1, \quad 0 \le s \le m_2 - 1.$$

xxx Product chain. Add discussion of *all* eigenvalues to Section yyy:6.2 of Chapter 4?

xxx P.S. See Chapter 5, (66).

In particular, assuming $m_1 \ge m_2$ we have

$$\tilde{\tau}_2 = 2\left(1 - \cos\frac{\pi}{m_1}\right)^{-1}. \tag{23}$$

Now we apply Corollary 4 to bound the eigenvalues λ_l. In Theorem 1, the two graphs agree except for self-loops, so

$$A = w/\tilde{w};$$

furthermore,

$$a = \min_i \frac{\tilde{\pi}_i}{\pi_i} = \frac{w}{\tilde{w}} \min_i \tilde{w}_i w_i,$$

so

$$\frac{A}{a} = \max_i \frac{w_i}{\tilde{w}_i} \le 1.$$

Thus $\lambda_l^{-1} \le \tilde{\lambda}_l^{-1}$ for $1 \le l \le n := m_1 m_2$; in particular,

$$\tau_2 \le \tilde{\tau}_2. \tag{24}$$

Comparing the other way around gives

$$\tilde{\lambda}_l^{-1} \le \left(\max_i \frac{\tilde{w}_i}{w_i} \right) \lambda_l^{-1} = 2\lambda_l^{-1}, \quad 1 \le l \le n$$

and in particular

$$\tau_2 \ge \tfrac{1}{2}\tilde{\tau}_2.$$

The result $\frac{1}{2}\tilde{\lambda}_l^{-1} \le \lambda_l^{-1} \le \tilde{\lambda}_l^{-1}$ extends to general d, for which (for example)

$$\tilde{\tau}_2 = d\left(1 - \cos\frac{\pi}{m}\right)^{-1}$$

where $I = \{0, \ldots, m_1 - 1\} \times \cdots \times \{0, \ldots, m_d - 1\}$ and $m := \max_i m_i$.

Example 6 *Random walk on a thinned grid.*

As a somewhat more interesting example, suppose we modify the grid in $\mathbf{Z}^2$ in Example 5 by deleting at most one edge from each unit square.

xxx Copy picture on page 700 in [3] as example?

Again we can apply Corollary 4, using the same benchmark graph as in Example 5. In Theorem 1, $\tilde{w}_{xy} > 0$ for $x \ne y$ if and only if x and y are neighboring vertices in the (unthinned) grid $\{0, \ldots, m_1 - 1\} \times \{0, \ldots, m_2 - 1\}$. We can choose γ_{xy} to have length 1 (if the edge joining x and y has not been deleted) or 3 (if it has). For any directed edge e in the grid, there are at most two paths of length 3 and at most one path of length 1 passing through e. Thus $A \le 7w/\tilde{w}$, and so $A/a \le 7\max_i(w_i/\tilde{w}_i) \le 7$; comparing the other way around is even easier (all paths have length 1), and we find

$$\tfrac{1}{4}\tilde{\lambda}_l^{-1} \le \lambda_l^{-1} \le 7\tilde{\lambda}_l^{-1}, \quad 2 \le l \le n.$$

xxx REMINDER: NOTES OR ELSEWHERE?: *Mention* exclusion process [7, 3].

Example 7 *The n-path with end self-loops.*

The comparison technique does not always provide results as sharp as those in the preceding two examples, even when the two chains are "close." For example, let the chain of interest be the n-path, with self-loops added at each end added to make the graph regular with degree 2, and let the benchmark graph be the n-cycle (Chapter 5, Example yyy:7). Use of Corollary 2 gives only $\tau_2 \le n\tilde{\tau}_2$, whereas in fact $\tau_2 = (1 - \cos\frac{\pi}{n})^{-1} \sim \frac{2}{\pi^2}n^2$ and $\tilde{\tau}_2 = \left(1 - \cos\frac{2\pi}{n}\right)^{-1} \sim \frac{1}{2\pi^2}n^2$.

It is difficult in general to use Corollary 4 to improve upon (21). However, when both the chain of interest and the benchmark chain are symmetric reversible chains (as defined in Chapter 7 Section yyy:1.1), it follows from Chapter 4 yyy:(14) by averaging over i that

$$\hat{d}(t) \leq \tilde{\hat{d}}\left(\frac{a}{A}t\right), \quad t \geq 0,$$

and hence from (1) we obtain

Corollary 8 (comparison of L^2 mixing times) *In Theorem 1, if both the graph of interest and the benchmark graph are vertex-transitive, then the L^2 mixing time parameters $\hat{\tau}$ and $\tilde{\hat{\tau}}$ satisfy*

$$\hat{\tau} \leq \frac{A}{a}\tilde{\hat{\tau}}.$$

Example 9 Returning to the slow card-shuffling scheme of Example 3 with random transpositions benchmark, it is known from group representation methods [1, 6] which make essential use of *all* the eigenvalues $\tilde{\lambda}_r$, not just $\tilde{\lambda}_2$, that

$$\tilde{\hat{\tau}} \sim \tfrac{1}{2} m \log m \quad \text{as } m \to \infty.$$

Since $a = 1$ and A ($= K$ of Chapter 7, Theorem yyy:14) $\leq 27m^2$, it follows that

$$\hat{\tau} \leq (1 + o(1))\tfrac{27}{2} m^3 \log m. \tag{25}$$

This improves upon Example 3, which combines with (21) to give only

$$\hat{\tau} \leq (1 + o(1))\tfrac{27}{4} m^4 \log m.$$

xxx Show truth is $\hat{\tau} = \Theta(m^3 \log m)$?

2 Improved bounds on L^2 distance

The central theme of the remainder of this chapter is that norms other than the L^1 norm (and closely related variation distance) and L^2 norm can be used to improve substantially upon the bound

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \leq \pi^{-1/2} e^{-t/\tau_2}. \tag{26}$$

2.1 L^q norms and operator norms

Our discussion here of L^q norms will parallel and extend the discussion in Chapter 2 Section yyy:6.2 of L^1 and L^2 norms. Given $1 \leq q \leq \infty$, both the L^q norm of a function and the L^q norm of a signed measure are defined with respect to some fixed reference probability distribution π on I, which for our purposes will be the stationary distribution of some irreducible but not necessarily reversible chain under consideration. For $1 \leq q < \infty$, the L^q norm of a *function* $f : I \to \mathbf{R}$ is

$$\|f\|_q := \left(\sum_i \pi_i |f(i)|^q \right)^{1/q},$$

and we define the L^q norm of a *signed measure* ν on I to be the L^q norm of its density function with respect to π:

$$\|\nu\|_q := \left(\sum_j \pi_j^{1-q} \, |\nu_j|^q \right)^{1/q}.$$

For $q = \infty$, the corresponding definitions are

$$\|f\|_\infty := \max_i |f(i)|$$

and

$$\|\nu\|_\infty := \max_j (|\nu_j| / \pi_j).$$

Any matrix $\mathbf{A} := (a_{ij} : i, j \in I)$ operates on functions $f : I \to \mathbf{R}$ by left-multiplication:

$$(\mathbf{A}f)(i) = \sum_j a_{ij} f(j), \tag{27}$$

and on signed measures ν by right-multiplication:

$$(\nu \mathbf{A})_j = \sum_i \nu_i a_{ij}. \tag{28}$$

For (27), fix $1 \leq q_1 \leq \infty$ and $1 \leq q_2 \leq \infty$ and regard $\mathbf{A}$ as a linear operator mapping L^{q_1} into L^{q_2}. The *operator norm* $\|\mathbf{A}\|_{q_1 \to q_2}$ is defined by

$$\|\mathbf{A}\|_{q_1 \to q_2} := \sup\{ \|\mathbf{A}f\|_{q_2} : \|f\|_{q_1} = 1 \}. \tag{29}$$

The *sup* in (29) is always achieved, and there are many equivalent reexpressions, including

$$\begin{aligned} \|\mathbf{A}\|_{q_1 \to q_2} &= \max\{\|\mathbf{A}f\|_{q_2} : \|f\|_{q_1} \le 1\} \\ &= \max\{\|\mathbf{A}f\|_{q_2}/\|f\|_{q_1} : f \ne 0\}. \end{aligned}$$

Note also that

$$\|\mathbf{BA}\|_{q_1 \to q_3} \le \|\mathbf{A}\|_{q_1 \to q_2} \|\mathbf{B}\|_{q_2 \to q_3}, \quad 1 \le q_1, q_2, q_3 \le \infty. \tag{30}$$

For (28), we may similarly regard $\mathbf{A}$ as a linear operator mapping signed measures ν, measured by $\|\nu\|_{q_1}$, to signed measures $\nu\mathbf{A}$, measured by $\|\nu\mathbf{A}\|_{q_2}$. The corresponding definition of operator norm, call it $|||\mathbf{A}|||_{q_1 \to q_2}$, is then

$$|||\mathbf{A}|||_{q_1 \to q_2} := \sup\{\|\nu\mathbf{A}\|_{q_2} : \|\nu\|_{q_1} = 1\}.$$

A brief calculation shows that

$$|||\mathbf{A}|||_{q_1 \to q_2} = \|\mathbf{A}^*\|_{q_1 \to q_2},$$

where $\mathbf{A}^*$ is the matrix with (i,j) entry $\pi_j a_{ji}/\pi_i$, that is, $\mathbf{A}^*$ is the adjoint operator to $\mathbf{A}$ (with respect to π).

Our applications in this chapter will all have $\mathbf{A} = \mathbf{A}^*$, so we will not need to distinguish between the two operator norms. In fact, all our applications will take $\mathbf{A}$ to be either $\mathbf{P}_t$ or $\mathbf{P}_t - \mathbf{E}$ for some $t \ge 0$, where

$$\mathbf{P}_t := (p_{ij}(t) : i, j \in I)$$

xxx notation $\mathbf{P}_t$ found elsewhere in book?

and $\mathbf{E} = \lim_{t \to \infty} \mathbf{P}_t$ is the transition matrix for the trivial discrete time chain that jumps in one step to stationarity:

$$\mathbf{E} = (\pi_j : i, j \in I),$$

and where we assume that the chain for $(\mathbf{P}_t)$ is *reversible*. Note that $\mathbf{E}$ operates on functions essentially as expectation with respect to π:

$$(\mathbf{E}f)(i) = \sum_j \pi_j f(j), \quad i \in I.$$

The effect of $\mathbf{E}$ on signed measures is to map ν to $(\sum_i \nu_i)\pi$, and

$$\mathbf{P}_t\mathbf{E} = \mathbf{E} = \mathbf{E}\mathbf{P}_t, \quad t \ge 0. \tag{31}$$

2.2 A more general bound on L^2 distance

The following preliminary result, a close relative to Chapter 3, Lemmas yyy:21 and 23, is used frequently enough in the sequel that we isolate it for reference. It is the simple identity in part (b) that shows why L^2-based techniques are so useful.

Lemma 10 *(a) For any function f,*

$$\frac{d}{dt}||\mathbf{P}_t f||_2^2 = -2\mathcal{E}(\mathbf{P}_t f, \mathbf{P}_t f) \leq -\frac{2}{\tau_2}\mathrm{var}_\pi\, \mathbf{P}_t f \leq 0.$$

(b)

$$||\mathbf{P}_t - \mathbf{E}||_{2\to 2} = e^{-t/\tau_2}, \quad t \geq 0.$$

Proof. (a) Using the backward equations

$$\frac{d}{dt}p_{ij}(t) = \sum_k q_{ik}\, p_{kj}(t)$$

we find

$$\frac{d}{dt}(\mathbf{P}_t f)(i) = \sum_k q_{ik}[(\mathbf{P}_t f)(k)]$$

and so

$$\begin{aligned}
\frac{d}{dt}||\mathbf{P}_t f||_2^2 &= 2\sum_i\sum_k \pi_i[(\mathbf{P}_t f)(i)]q_{ik}[(\mathbf{P}_t f)(k)] \\
&= -2\mathcal{E}(\mathbf{P}_t f, \mathbf{P}_t f) \quad \text{by Chapter 3 yyy:(70)} \\
&\leq -\frac{2}{\tau_2}\mathrm{var}_\pi\, \mathbf{P}_t f \quad \text{by the extremal characterization of } \tau_2.
\end{aligned}$$

(b) From (a), for any f we have

$$\frac{d}{dt}||(\mathbf{P}_t - \mathbf{E})f||_2^2 = \frac{d}{dt}||\mathbf{P}_t(f - \mathbf{E}f)||_2^2 \leq -\frac{2}{\tau_2}||(\mathbf{P}_t - \mathbf{E})f||_2^2,$$

which yields

$$\begin{aligned}
||(\mathbf{P}_t - \mathbf{E})f||_2^2 &\leq ||(\mathbf{P}_0 - \mathbf{E})f||_2^2\, e^{-2t/\tau_2} = (\mathrm{var}_\pi\, f)e^{-2t/\tau_2} \\
&\leq ||f||_2^2\, e^{-2t/\tau_2}.
\end{aligned}$$

Thus $||\mathbf{P}_t - \mathbf{E}||_{2\to 2} \leq e^{-t/\tau_2}$. Taking f to be the eigenvector

$$f_i := \pi_i^{-1/2} u_{i2}, \quad i \in I,$$

of $\mathbf{P}_t - \mathbf{E}$ corresponding to eigenvalue $\exp(-t/\tau_2)$ demonstrates equality and completes the proof of (b). ■

The key to all further developments in this chapter is the following result.

Lemma 11 *For an irreducible reversible chain with arbitrary initial distribution and any $s, t \geq 0$,*

$$\|P(X_{s+t} \in \cdot) - \pi(\cdot)\|_2 \leq \|P(X_s \in \cdot)\|_2 \|\mathbf{P}_t - \mathbf{E}\|_{2 \to 2} = \|P(X_s \in \cdot)\|_2 \, e^{-t/\tau_2}.$$

Proof. The equality is Lemma 10(b), and

$$\|P(X_{s+t} \in \cdot) - \pi(\cdot)\|_2 = \|P(X_s \in \cdot)(\mathbf{P}_t - \mathbf{E})\|_2 \leq \|P(X_s \in \cdot)\|_2 \|\mathbf{P}_t - \mathbf{E}\|_{2 \to 2}$$

proves the inequality. ■

We have already discussed, in Section 1, a technique for bounding τ_2 when (as is usually the case) it cannot be computed exactly. To utilize Lemma 11, we must also bound $\|P(X_s \in \cdot)\|_2$. Since

$$\|P(X_s \in \cdot)\|_2 = \|P(X_0 \in \cdot)\mathbf{P}_s\|_2 \tag{32}$$

xxx For NOTES: By Jensen's inequality (for $1 \leq q < \infty$), any transition matrix contracts L^q for any $1 \leq q \leq \infty$.

and each $\mathbf{P}_t$ is contractive on L^2, i.e., $\|\mathbf{P}_t\|_{2 \to 2} \leq 1$ (this follows, for example, from Lemma 10(a); and note that $\|\mathbf{P}_t\|_{2 \to 2} = 1$ by considering constant functions), it follows that

$$\|P(X_s \in \cdot)\|_2 \text{ decreases monotonically to 1 as } s \uparrow \infty, \tag{33}$$

and the decrease is strictly monotone unless $P(X_0 \in \cdot) = \pi(\cdot)$. From (32) follows

$$\|P(X_s \in \cdot)\|_2 \leq \|P(X_0 \in \cdot)\|_{q^*} \|\mathbf{P}_s\|_{q^* \to 2} \text{ for any } 1 \leq q^* \leq \infty, \tag{34}$$

and again

$$\|\mathbf{P}_s\|_{q^* \to 2} \text{ decreases monotonically to 1 as } s \uparrow \infty. \tag{35}$$

The norm $\|\mathbf{P}_s\|_{q^* \to 2}$ decreases in q^* (for fixed s) and is identically 1 when $q^* \geq 2$, but in applications we will want to take $q^* < 2$. The following duality lemma will then often prove useful. Recall that $1 \leq q, q^* \leq \infty$ are said to be *(Hölder-)conjugate exponents* if

$$\frac{1}{q} + \frac{1}{q^*} = 1. \tag{36}$$

Lemma 12 *For any operator* $\mathbf{A}$*, let*

$$\mathbf{A}^* = (\pi_j a_{ji}/\pi_i : i, j \in I)$$

denote its adjoint with respect to π*. Then, for any* $1 \leq q_1, q_2 \leq \infty$*,*

$$||\mathbf{A}||_{q_1 \to q_2} = ||\mathbf{A}^*||_{q_2^* \to q_1^*}.$$

In particular, for a reversible chain and any $1 \leq q \leq \infty$ *and* $s \geq 0$*,*

$$||\mathbf{P}_s||_{2 \to q} = ||\mathbf{P}_s||_{q^* \to 2}. \tag{37}$$

Proof. Classical duality for L^q spaces (see, e.g., Chapter 6 in [11]) asserts that, given $1 \leq q \leq \infty$ and g on I,

$$||g||_{q^*} = \max\{|\langle f, g\rangle| : ||f||_q = 1\}$$

where

$$\langle f, g\rangle := \sum_i \pi_i f(i) g(i).$$

Thus

$$\begin{aligned} ||\mathbf{A}^* g||_{q_1^*} &= \max\{|\langle f, \mathbf{A}^* g\rangle| : ||f||_{q_1} = 1\} \\ &= \max\{|\langle \mathbf{A} f, g\rangle| : ||f||_q = 1\}, \end{aligned}$$

and also

$$|\langle \mathbf{A} f, g\rangle| \leq ||\mathbf{A} f||_{q_2} ||g||_{q_2^*} \leq ||\mathbf{A}||_{q_1 \to q_2} ||f||_{q_1} ||g||_{q_2^*},$$

so

$$||\mathbf{A}^* g||_{q_1^*} \leq ||\mathbf{A}||_{q_1 \to q_2} ||g||_{q_2^*}.$$

Since this is true for every g, we conclude $||\mathbf{A}^*||_{q_2^* \to q_1^*} \leq ||\mathbf{A}||_{q_1 \to q_2}$. Reverse roles to complete the proof. ■

As a corollary, if $q^* = 1$ then (34) and (37) combine to give

$$||P(X_s \in \cdot)||_2 \leq ||\mathbf{P}_s||_{1 \to 2} = ||\mathbf{P}_s||_{2 \to \infty}$$

and then

$$||P(X_{s+t} \in \cdot) - \pi(\cdot)||_2 \leq ||\mathbf{P}_s||_{2 \to \infty} \, e^{-t/\tau_2}$$

from Lemma 11. Thus

$$\sqrt{\hat{d}(2(s+t))} \leq ||\mathbf{P}_s||_{2 \to \infty} \, e^{-t/\tau_2}. \tag{38}$$

Here is a somewhat different derivation of (38):

Lemma 13 *For $0 \le s \le t$,*

$$\begin{aligned}\sqrt{\hat{d}(2t)} = \|\mathbf{P}_t - \mathbf{E}\|_{2\to\infty} &\le \|\mathbf{P}_s\|_{2\to\infty}\|\mathbf{P}_{t-s} - \mathbf{E}\|_{2\to 2} \\ &= \|\mathbf{P}_s\|_{2\to\infty}\, e^{-(t-s)/\tau_2}.\end{aligned}$$

Proof. In light of (31), (30), and Lemma 10(b), we need only establish the first equality. Indeed, $\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2$ is the L^2 norm of the *function* $(P_i(X_t \in \cdot)/\pi(\cdot)) - 1$ and so equals

$$\max\left\{\left|\sum_j (p_{ij}(t) - \pi_j) f(j)\right| : \|f\|_2 = 1\right\} = \max\{|((\mathbf{P}_t - \mathbf{E})f)(i)| : \|f\|_2 = 1\}.$$

Taking the maximum over $i \in I$ we obtain

$$\begin{aligned}\sqrt{\hat{d}(2t)} &= \max\left\{\max_i |((\mathbf{P}_t - \mathbf{E})f)(i)| : \|f\|_2 = 1\right\} \\ &= \max\{\|(\mathbf{P}_t - \mathbf{E})f\|_\infty : \|f\|_2 = 1\} \\ &= \|\mathbf{P}_t - \mathbf{E}\|_{2\to\infty}. \quad \blacksquare\end{aligned}$$

Choosing $s = 0$ in Lemma 11 recaptures (26), and choosing $s = 0$ in Lemma 13 likewise recaptures the consequence (5) of (26). The central theme for both Nash and log-Sobolev techniques is that one can improve upon these results by more judicious choice of s.

2.3 Exact computation of $N(s)$

The proof of Lemma 13 can also be used to show that

$$N(s) := \|\mathbf{P}_s\|_{2\to\infty} = \max_i \|P_i(X_s \in \cdot)\|_2 = \max_i \sqrt{\frac{p_{ii}(2s)}{\pi_i}}, \tag{39}$$

as at (9). In those rare instances when the spectral representation is known explicitly, this gives the formula

xxx Also useful in conjunction with comparison method—see Section 3.

xxx If we can compute this, we can compute $\hat{d}(2t) = N^2(t) - 1$. But the point is to test out Lemma 13.

$$N^2(s) = 1 + \max_i \pi_i^{-1} \sum_{m=2}^{n} u_{im}^2 \exp(-2\lambda_m s), \tag{40}$$

and the techniques of later sections are not needed to compute $N(s)$. In particular, in the vertex-transitive case

$$N^2(s) = 1 + \sum_{m=2}^{n} \exp(-2\lambda s).$$

The norm $N(s)$ clearly behaves nicely under the formation of products:

$$N(s) = N^{(1)}(s) N^{(2)}(s). \tag{41}$$

Example 14 *The two-state chain and the d-cube.*

For the two-state chain, the results of Chapter 5 Example yyy:4 show

$$N^2(s) = 1 + \frac{\max(p,q)}{\min(p,q)} e^{-2(p+q)s}.$$

In particular, for the continuized walk on the 2-path,

$$N^2(s) = 1 + e^{-4s}.$$

By the extension of (41) to higher-order products, we therefore have

$$N^2(s) = (1 + e^{-4s/d})^d$$

for the continuized walk on the d-cube. This result is also easily derived from the results of Chapter 5 Example yyy:15. For $d \geq 2$ and $t \geq \frac{1}{4} d \log(d-1)$, the optimal choice of s in Lemma 13 is therefore

$$s = \tfrac{1}{4} d \log(d-1)$$

and this leads in straightforward fashion to the bound

$$\hat{\tau} \leq \tfrac{1}{4} d (\log d + 3).$$

While this is a significant improvement on the bound [cf. (12)]

$$\hat{\tau} \leq \tfrac{1}{4} (\log 2) d^2 + \tfrac{1}{2} d$$

obtained by setting $s = 0$, i.e., obtained using only information about τ_2, it is not

xxx REWRITE!, in light of corrections to my notes.

as sharp as the upper bound

$$\hat{\tau} \leq (1 + o(1)) \tfrac{1}{4} d \log d$$

in (13) that will be derived using log-Sobolev techniques.

Example 15 *The complete graph.*

For this graph, the results of Chapter 5 Example yyy:9 show

$$N^2(s) = 1 + (n-1)\exp\left(-\frac{2ns}{n-1}\right).$$

It turns out for this example that $s = 0$ is the optimal choice in Lemma 13. This is not surprising given the sharpness of the bound in (7) in this case. See Example 32 below for further details.

Example 16 *Product random walk on a d-dimensional grid.*

Consider again the benchmark product chain (i.e., the "tilde chain") in Example 5. That chain has relaxation time

$$\tau_2 = d\left(1 - \cos\left(\frac{\pi}{m}\right)\right)^{-1} \le \frac{1}{2}dm^2,$$

so choosing $s = 0$ in Lemma 13 gives

$$\begin{aligned} \hat{\tau} &\le \frac{d}{1-\cos(\pi/m)}\left(\frac{1}{2}\log n + 1\right) \\ &\le \tfrac{1}{4}dm^2(\log n + 2). \end{aligned} \tag{42}$$

This bound can be improved using $N(\cdot)$. Indeed, if we first consider continuized random walk on the m-path with self-loops added at each end, the stationary distribution is uniform, the eigenvalues are

$$\lambda_l = 1 - \cos(\pi(l-1)/m), \quad 1 \le l \le m,$$

and the eigenvectors are given by

$$u_{il} = (2/m)^{1/2}\cos(\pi(l-1)(i-\tfrac{1}{2})/m), \quad 0 \le i \le m-1, \quad 2 \le l \le m.$$

According to (40) and simple estimates, for $s > 0$

$$N^2(s) - 1 \le 2\sum_{l=2}^{m}\exp[-2s(1-\cos(\pi(l-1)/m)] \le 2\sum_{l=1}^{m-1}\exp(-4sl^2/m^2)$$

and

$$\begin{aligned}
\sum_{l=2}^{m-1} \exp(-4sl^2/m^2) &\le \int_{x=1}^{\infty} \exp(-4sx^2/m^2)\,dx \\
&= m\left(\frac{\pi}{4s}\right)^{1/2} P\left(Z \ge \frac{2(2s)^{1/2}}{m}\right) \\
&\le \left(\frac{\pi/4}{4s/m^2}\right)^{1/2} \exp(-4s/m^2) \\
&\le (4s/m^2)^{-1/2} \exp(-4s/m^2)
\end{aligned}$$

when Z is standard normal; in particular, we have used the well-known (xxx: point to Ross book exercise) bound

$$P(Z \ge z) \le \tfrac{1}{2} e^{-z^2/2}, \quad z \ge 0.$$

Thus

$$N^2(s) \le 1 + 2[1 + (4s/m^2)^{-1/2}] \exp(-4s/m^2), \quad s > 0.$$

Return now to the "tilde chain" of Example 5, and assume for simplicity that $m_1 = \cdots = m_d = m$. Since this chain is a slowed-down d-fold product of the path chain, it has

$$N^2(s) \le \left[1 + 2\left(1 + \left(\frac{4s}{dm^2}\right)^{-1/2}\right) \exp\left(-\frac{4s}{dm^2}\right)\right]^d, \quad s > 0. \tag{43}$$

In particular, since $\hat{d}(2t) = N^2(t) - 1$, it is now easy to see that

$$\hat{\tau} \le K m^2 d \log d = K n^{2/d} d \log d \tag{44}$$

for a universal constant K.

xxx We've improved on (42), which gave order $m^2 d^2 \log m$.

xxx When used to bound $d(t)$ at (4), the bound (44) is "right": see Theorem 4.1, p. 481, in [5].

The optimal choice of s in Lemma 13 cannot be obtained explicitly when (43) is used to bound $N(s) = \|\mathbf{P}_s\|_{2\to\infty}$. For this reason, and for later purposes, it is useful to use the simpler, but more restricted, bound

$$N^2(s) \le (4dm^2/s)^{d/2} \quad \text{for } 0 \le s \le dm^2/16. \tag{45}$$

To verify this bound, simply notice that $u+2ue^{-u^2}+2e^{-u^2} \le 4$ for $0 \le u \le \frac{1}{2}$. When $s = dm^2/16$, (45), and $\tau_2 \le dm^2/2$ are used in Lemma 13, we find

$$\hat{\tau} \le \tfrac{3}{4}m^2 d^2(\log 2 + \tfrac{3}{4}d^{-1}).$$

xxx Improvement over (42) by factor $\Theta(\log m)$, but display following (43) shows still off by factor $\Theta(d/\log d)$.

3 Nash inequalities

xxx For NOTES: Nash vs. Sobolev

A Nash inequality for a chain is an inequality of the form

$$\|g\|_2^{2+\frac{1}{D}} \le C\,[\mathcal{E}(g,g) + \tfrac{1}{T}\|g\|_2^2]\,\|g\|_1^{1/D} \tag{46}$$

that holds for some positive constants C, D, and T and for all functions g. We connect Nash inequalities to mixing times in Section 3.1, and in Section 3.2 we discuss a comparison method for establishing such inequalities.

3.1 Nash inequalities and mixing times

A Nash inequality implies a useful bound on the quantity

$$N(t) = \|\mathbf{P}_t\|_{2\to\infty} \tag{47}$$

appearing in the mixing time Lemma 13. This norm is continuous in t and decreases to $\|\mathbf{E}\|_{2\to\infty} = 1$ as $t \uparrow \infty$. Here is the main result:

Theorem 17 *If the Nash inequality (46) holds for a continuous-time reversible chain, some $C, D, T > 0$, and all g, then the norm $N(s)$ at (47) satisfies*

$$N(t) \le e(DC/t)^D \quad \textit{for } 0 < t \le T.$$

Proof. First note $N(t) = \|\mathbf{P}_t\|_{1\to 2}$ by Lemma 12. Thus we seek a bound on $h(t) := \|\mathbf{P}_t g\|_2^2$ independent of g satisfying $\|g\|_1 = 1$; the square root of such a bound will also bound $N(t)$.

Substituting $\mathbf{P}_t g$ for g in (46) and utilizing the identity in Lemma 10(a) and the fact that $\mathbf{P}_t$ is contractive on L^1, we obtain the differential inequality

$$h(t)^{1+\frac{1}{2D}} \le C\,[-\tfrac{1}{2}h'(t) + \tfrac{1}{T}h(t)], \quad t \ge 0.$$

Writing

$$H(t) := \left[\tfrac{1}{2}Ch(t)e^{-2t/T}\right]^{-1/(2D)},$$

the inequality can be equivalently rearranged to

$$H'(t) \geq \left[2D(C/2)^{1+\frac{1}{2D}}\right]^{-1} e^{\frac{t}{DT}}, \quad t \geq 0.$$

Since $H(0) > 0$, it follows that

$$H(t) \geq T\left[2(C/2)^{1+\frac{1}{2D}}\right]^{-1}\left(e^{\frac{t}{DT}} - 1\right), \quad t \geq 0,$$

or equivalently

$$h(t) \leq \left[\frac{T}{C}\left(1 - e^{-t/(DT)}\right)\right]^{-2D}, \quad t \geq 0.$$

But

$$e^{-t/(DT)} \leq 1 - \frac{t}{T}(1 - e^{-1/D}) \quad \text{for } 0 < t \leq T,$$

so for these same values of t we have

$$\begin{aligned} h(t) &\leq \left[\frac{t}{C}\left(1 - e^{-1/D}\right)\right]^{-2D} \\ &= e^2\left[\frac{t}{C}\left(e^{1/D} - 1\right)\right]^{-2D} \leq [e(DC/t)^D]^2, \end{aligned}$$

as desired. ■

We now return to Lemma 13 and, for $t \geq T$, set $s = T$. (Indeed, using the bound on $N(s)$ in Theorem 17, this is the optimal choice of s if $T < D\tau_2$.) This gives

xxx NOTE: In next theorem, only need *conclusion* of Theorem 17, not hypothesis!

Theorem 18 *In Theorem 17, if $c \geq 1$ and*

$$t \geq T + \tau_2\left(D\log\left(\frac{DC}{T}\right) + c\right),$$

then $\sqrt{\hat{d}(2t)} \leq e^{1-c}$; in particular,

$$\hat{\tau} \leq T + \tau_2\left(D\log\left(\frac{DC}{T}\right) + 2\right).$$

The following converse of sorts to Theorem 17 will be very useful in conjunction with the comparison method.

Theorem 19 *If a continuous time reversible chain satisfies*

$$N(t) \le C t^{-D} \quad \textit{for } 0 < t \le T,$$

then it satisfies the Nash inequality

$$\|g\|_2^{2+\frac{1}{D}} \le C' \, [\mathcal{E}(g,g) + \tfrac{1}{2T}\|g\|_2^2] \, \|g\|_1^{1/D} \quad \textit{for all } g$$

with

$$C' := 2\,(1 + \tfrac{1}{2D})\,[(1+2D)^{1/2} C]^{1/D} \le 2^{2+\frac{1}{2D}}\, C^{1/D}.$$

xxx Detail in proof to be filled in (I have notes): Show $\mathcal{E}(\mathbf{P}_t g, \mathbf{P}_t g) \downarrow$ as $t \uparrow$, or at least that it's maximized at $t = 0$. Stronger of two statements is equivalent to assertion that $\|\mathbf{P}_t f\|_2^2$ is convex in t.

Proof. As in the proof of Theorem 17, we note $N(t) = \|\mathbf{P}_t\|_{1\to 2}$. Hence, for any g and any $0 < t \le T$,

$$\begin{aligned}
\|g\|_2^2 &= \|\mathbf{P}_t g\|_2^2 - \int_{s=0}^{t} \tfrac{d}{ds}\|\mathbf{P}_s g\|_2^2 \, ds \\
&= \|\mathbf{P}_t g\|_2^2 + 2\int_{s=0}^{t} \mathcal{E}(\mathbf{P}_s g, \mathbf{P}_s g)\, ds \quad \text{by Lemma 10(a)} \\
&\le \|\mathbf{P}_t g\|_2^2 + 2\mathcal{E}(g,g)t \quad \text{xxx see above} \\
&\le 2\mathcal{E}(g,g)t + C^2 t^{-2D}\|g\|_1^2.
\end{aligned}$$

This gives

$$\|g\|_2^2 \le t[2\mathcal{E}(g,g) + \tfrac{1}{T}\|g\|_2^2] + C^2 t^{-2D}\|g\|_1^2$$

for *any* $t > 0$. The righthand side here is convex in t and minimized (for $g \ne 0$) at

$$t = \left(\frac{2DC^2\|g\|_1^2}{2\mathcal{E}(g,g) + T^{-1}\|g\|_2^2} \right)^{1/(2D+1)}.$$

Plugging in this value, raising both sides to the power $1 + \frac{1}{2D}$ and simplifying yields the desired Nash inequality. The upper bound for C' is derived with a little bit of calculus. ■

3.2 The comparison method for bounding $N(\cdot)$

In Section 1 we compared relaxation times for two chains by comparing Dirichlet forms and variances. The point of this subsection is that the comparison can be extended to the norm function $N(\cdot)$ of (47) using Nash inequalities. Then results on $N(\cdot)$ like those in Section 2.3 can be used to bound mixing times for other chains on the same state space.

xxx For NOTES?: Can even use different spaces. New paragraph:

To see how this goes, suppose that a benchmark chain is known to satisfy

$$\tilde{N}(t) \le \tilde{C} t^{-\tilde{D}} \quad \text{for } 0 < t \le \tilde{T}. \tag{48}$$

By Theorem 19, it then satisfies a Nash inequality. The L^1- and L^2-norms appearing in this inequality can be compared in the obvious fashion [cf. (18)] and the Dirichlet forms can be compared as in Theorem 1. This shows that the chain of interest also satisfies a Nash inequality. But then Theorem 17 gives a bound like (48) for the chain of interest, and Theorem 18 can then be used to bound the L^2 threshold time $\hat{\tau}$.

Here is the precise result; the details of the proof are left to the reader.

Theorem 20 (comparison of bounds on $N(\cdot)$) *If a reversible benchmark chain satisfies*

$$\tilde{N}(t) \le \tilde{C} t^{-\tilde{D}} \quad \textit{for } 0 < t \le \tilde{T}$$

for constants $\tilde{C}, \tilde{D}, \tilde{T} > 0$, then any other reversible chain on the same state space satisfies

$$N(t) \le e(DC/t)^D \quad \textit{for } 0 < t \le T,$$

where, with a and A as defined in Corollary 2, and with

$$a' := \max_i (\tilde{\pi}_i / \pi_i),$$

we set

$$\begin{aligned} D &= \tilde{D}, \\ C &= a^{-(2+\frac{1}{D})} a'^{1/D} A \times 2(1 + \tfrac{1}{2D})[(1+2D)^{1/2} \tilde{C}]^{1/D} \\ &\le a^{-(2+\frac{1}{D})} a'^{1/D} A \times 2^{2+\frac{1}{2D}} \tilde{C}^{1/D}, \\ T &= \frac{2A}{a'^2} \tilde{T}. \end{aligned}$$

xxx Must correct this slightly. Works for *any* A such that $\tilde{\mathcal{E}} \leq A\mathcal{E}$, not just minimal one. This is important since we need a *lower* bound on T but generally only have an *upper* bound on $\tilde{\mathcal{E}}/\mathcal{E}$. The same goes for a' (only need *upper bound* on $\tilde{\pi}_i/\pi_i$): we also need an upper bound on T.

Example 21 *Random walk on a d-dimensional grid.*

As in Example 5, consider the continuized walk on the d-dimensional grid $I = \{0, \ldots, m-1\}^d$. In Example 5 we compared the Dirichlet form and variance for this walk to the d-fold product of random walk on the m-path with end self-loops to obtain

$$\tau_2 \leq \tilde{\tau}_2 = d(1 - \cos\tfrac{\pi}{m})^{-1} \leq \tfrac{1}{2}dm^2; \tag{49}$$

using the simple bound $\sqrt{\hat{d}(2t)} \leq \pi_*^{-1/2}e^{-t/\tau_2} \leq (2n)^{-1/2}\exp\left(-\frac{2t}{dm^2}\right)$ we then get

$$\hat{\tau} \leq \tfrac{1}{4}m^2 d[\log(2n) + 2], \tag{50}$$

which is of order $m^2d^2\log m$. Here we will see how comparing $N(\cdot)$, too, gives a bound of order $m^2d^2\log d$. In Example 43, we will bring log-Sobolev techniques to bear, too, to lower this bound to order $m^2 d\log d$

xxx which is correct, at least for TV. New paragraph:

Recalling (45), we may apply Theorem 20 with

$$\tilde{D} = d/4, \quad \tilde{C} = (4dm^2)^{d/4}, \quad \tilde{T} = dm^2/16,$$

and, from the considerations in Example 5,

$$a \geq \tfrac{1}{2}, \quad A = 1, \quad a' = 2.$$

xxx See xxx following Theorem 20. Same paragraph:

This gives

$$D = d/4, \quad C \leq 2^{6+\frac{10}{d}}dm^2 \leq 2^{16}dm^2, \quad T = dm^2/32.$$

Plugging these into Theorem 18 yields

$$\hat{\tau} \leq \tfrac{1}{8}m^2d^2\log d + (\tfrac{19}{8}\log 2)m^2d^2 + \tfrac{33}{32}m^2 d, \tag{51}$$

which is $\leq 3m^2d^2\log d$ for $d \geq 2$.

Other variants of the walk, including the thinned-grid walk of Example 6, can be handled in a similar fashion.

xxx Do moderate growth and local Poincaré? Probably *not*, to keep length manageable. Also, will need to rewrite intro a little, since not doing Δ^2-stuff (in any detail).

4 Logarithmic Sobolev inequalities

xxx For NOTES: For history and literature, see ([4], first paragraph and end of Section 1).

xxx For NOTES: Somewhere mention relaxing to *nonreversible* chains.

4.1 The log-Sobolev time τ_l

Given a probability distribution π on a finite set I, define

xxx For NOTES: Persi's $\mathcal{L}(g)$ is double ours.

$$\mathcal{L}(g) := \sum_i \pi_i g^2(i) \log(|g(i)|/\|g\|_2) \tag{52}$$

for $g \not\equiv 0$, recalling $\|g\|_2^2 = \sum_i \pi_i g^2(i)$ and using the convention $0 \log 0 = 0$. By Jensen's inequality,

$$\mathcal{L}(g) \geq 0, \text{ with equality if and only if } |g| \text{ is constant.}$$

Given a finite, irreducible, reversible Markov chain with stationary distribution π, define the *logarithmic Sobolev* (or *log-Sobolev*) *time* by

xxx For NOTES: Persi's α is $1/(2\tau_l)$.

xxx Note $\tau_l < \infty$. (Show?) See also Corollary 27.

$$\tau_l := \sup\{\mathcal{L}(g)/\mathcal{E}(g,g) : g \not\equiv \text{constant}\}. \tag{53}$$

Notice the similarity between (53) and the extremal characterization of τ_2 (Chapter 3, Theorem yyy:22):

$$\tau_2 = \sup\{\|g\|_2^2/\mathcal{E}(g,g) : \sum_i \pi_i g(i) = 0, \quad g \not\equiv 0\}.$$

We discuss exact computation of τ_l in Section 4.3, the behavior of τ_l for product chains in Section 4.4, and a comparison method for bounding τ_l in Section 4.5. In Section 4.2 we focus on the connection between τ_l and mixing times. A first such result asserts that the relaxation time does not exceed the log-Sobolev time:

Lemma 22 $\tau_2 \leq \tau_l$.

xxx Remarks about how "usually" equality?

xxx For NOTES: Proof from [10], via [4].

Proof. Given $g \not\equiv$ constant and ϵ, let $f := 1 + \epsilon g$. Then, writing $\bar{g} = \sum_i \pi_i g(i)$, and with all asymptotics as $\epsilon \to 0$,

$$\begin{aligned}\log |f|^2 &= 2\epsilon g - \epsilon^2 g^2 + O(\epsilon^3), \\ \log \|f\|_2^2 &= 2\epsilon \bar{g} + \epsilon^2 \|g\|_2^2 - 2\epsilon^2 \bar{g}^2 + O(\epsilon^3), \\ \log \frac{|f|^2}{\|f\|_2^2} &= 2\epsilon(g - \bar{g}) + \epsilon^2(2\bar{g}^2 - \|g\|_2^2 - g^2) + O(\epsilon^3).\end{aligned}$$

Also,

$$f^2 = 1 + 2\epsilon g + \epsilon^2 g^2;$$

thus

$$f^2 \log(|f|^2/\|f\|_2^2) = 2\epsilon(g - \bar{g}) + \epsilon^2(3g^2 - \|g\|_2^2 - 4g\bar{g} + 2\bar{g}^2) + O(\epsilon^3)$$

and so

$$\mathcal{L}(f) = \epsilon^2(\|g\|_2^2 - \bar{g}^2) + O(\epsilon^3) = \epsilon^2 \text{var}_\pi g + O(\epsilon^3).$$

Furthermore, $\mathcal{E}(f, f) = \epsilon^2 \mathcal{E}(g, g)$; therefore

$$\tau_l \geq \frac{\mathcal{L}(f)}{\mathcal{E}(f, f)} = \frac{\text{var}_\pi g}{\mathcal{E}(g, g)} + O(\epsilon).$$

Finish by letting $\epsilon \to 0$ and then taking the supremum over g. ■

4.2 τ_l, mixing times, and hypercontractivity

In this subsection we discuss the connection between the L^2 threshold time parameter

$$\hat{\tau} = \inf\{t > 0 : \sqrt{\hat{d}(2t)} = \max_i \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \leq e^{-1}\} \qquad (54)$$

and the log-Sobolev time τ_l. As in Section 3, we again consider the fundamental quantity

$$N(s) = \|\mathbf{P}_s\|_{2 \to \infty}$$

arising in the bound on $\sqrt{\hat{d}(2t)}$ in Lemma 13, and recall from Section 3.1 that

$$N(s) \text{ decreases strictly monotonically from } \pi_*^{-1/2} \text{ at } s = 0 \text{ to } 1 \text{ as } s \uparrow \infty.$$

The function N is continuous. It would be nice (especially for use in conjunction with the comparison technique) if we could characterize, in terms of the Dirichlet form $\mathcal{E}$, the value of s, call it s^*, such that $N(s)$ equals 2 (say), but such a characterization is not presently available.

xxx For NOTES?: A partial result is Theorem 3.9 in [4], taking $q = \infty$.

Open Problem 23 Characterize s^* in terms of $\mathcal{E}$.

To carry on along these general lines, it turns out to be somewhat more convenient to substitute use of

$$\|P(X_t \in \cdot) - \pi(\cdot)\|_2 \leq \|P(X_0 \in \cdot)\|_{\frac{q}{q-1}} \, \|\mathbf{P}_s\|_{2\to q} \, e^{-(t-s)/\tau_2}, \quad 2 \leq q < \infty, \tag{55}$$

an immediate consequence of Lemmas 11 and 12 and (34), for use of Lemma 13. The reason is that, like $N(s)$, $\|\mathbf{P}_s\|_{2\to q}$ decreases monotonically to 1 as $s \uparrow \infty$; but, unlike $N(s)$, it turns out that

$$\text{for each } q \geq 2,\ \|\mathbf{P}_s\|_{2\to q} \textit{ equals } 1 \text{ for all sufficiently large } s. \tag{56}$$

The property (56) is called *hypercontractivity*, in light of the facts that, for fixed s, $\mathbf{P}_s$ is a contraction on L^2 and $\|\mathbf{P}_s\|_{2\to q}$ is increasing in q. Let

$$s_q := \inf\{s \geq 0 : \|\mathbf{P}_s\|_{2\to q} \leq 1\} = \inf\{s : \|\mathbf{P}_s\|_{2\to q} = 1\};$$

then $s_2 = 0 < s_q$, and we will see presently that $s_q < \infty$ for $q \geq 2$. The following theorem affords a connections with the log-Sobolev time τ_l (and hence with the Dirichlet form $\mathcal{E}$).

Theorem 24 *For any finite, irreducible, reversible chain,*

$$\tau_l = \sup_{2<q<\infty} \frac{2s_q}{\log(q-1)}.$$

Proof. The theorem is equivalently rephrased as follows:

$$\|\mathbf{P}_t\|_{2\to q} \leq 1 \text{ for all } t \geq 0 \text{ and } 2 \leq q < \infty \text{ satisfying } e^{2t/u} \geq q-1 \tag{57}$$

if and only if $u \geq \tau_l$. The proof will make use of the generalization

$$\mathcal{L}_q(g) := \sum_i \pi_i |g(i)|^q \log(|g(i)|/\|g\|_q)$$

of (52). Fixing $0 \not\equiv g \geq 0$ and $u > 0$, we will also employ the notation

$$\begin{aligned} q(t) &:= 1 + e^{2t/u}, \quad G(t) := \|\mathbf{P}_t g\|_{q(t)}^{q(t)}, \qquad (58)\\ F(t) &:= \|\mathbf{P}_t g\|_{q(t)} = \exp\left[\frac{1}{q(t)} \log G(t)\right] \end{aligned}$$

for $t \geq 0$.

As a preliminary, we compute the derivative of F. To begin, we can proceed as at the start of the proof of Lemma 10(a) to derive

$$G'(t) = -q(t)\,\mathcal{E}\left(\mathbf{P}_t g, (\mathbf{P}_t g)^{q(t)-1}\right) + \frac{q'(t)}{q(t)} E_\pi \left[(\mathbf{P}_t g)^{q(t)} \log\left((\mathbf{P}_t g)^{q(t)}\right)\right].$$

Then

$$\begin{aligned} F'(t) &= \left[\frac{G'(t)}{q(t)G(t)} - \frac{q'(t)\log G(t)}{q^2(t)}\right] F(t) \\ &= F(t)^{-(q(t)-1)} \left[\frac{q'(t)}{q(t)} \mathcal{L}_{q(t)}(\mathbf{P}_t g) - \mathcal{E}\left(\mathbf{P}_t g, (\mathbf{P}_t g)^{q(t)-1}\right)\right]. \quad (59) \end{aligned}$$

For the first half of the proof we suppose that (57) holds and must prove $\tau_l \leq u$, that is, we must establish the log-Sobolev inequality

$$\mathcal{L}(g) \leq u\,\mathcal{E}(g,g) \text{ for every } g. \quad (60)$$

To establish (60) it is enough to consider $0 \not\equiv g > 0$,

xxx Do we actually use $g \geq 0$ here?

since for arbitrary g we have

$$\mathcal{L}(g) = \mathcal{L}(|g|) \text{ and } \mathcal{E}(g,g) \geq \mathcal{E}(|g|,|g|). \quad (61)$$

Plugging the specific formula (58) for $q(t)$ into (59) and setting $t = 0$ gives

$$F'(0) = \|g\|_2^{-1}(u^{-1}\mathcal{L}(g) - \mathcal{E}(g,g)). \quad (62)$$

Moreover, since

$$\begin{aligned} F(t) = \|\mathbf{P}_t g\|_{q(t)} &\leq \|\mathbf{P}_t\|_{2\to q(t)} \|g\|_2 \leq \|g\|_2 \quad \text{by (57)} \\ &= \|\mathbf{P}_0 g\|_2 = F(0), \end{aligned}$$

the (right-hand) derivative of F at 0 must be nonpositive. The inequality (60) now follows from (62).

For the second half of the proof, we may assume $u = \tau_l$ and must establish (57). For $g \geq 0$, (53) and Lemma 25 (to follow) give

$$\frac{q}{2}\mathcal{L}_q(g) = \mathcal{L}(g^{q/2}) \leq \tau_l\,\mathcal{E}(g^{q/2}, g^{q/2}) \leq \frac{q^2 \tau_l}{4(q-1)} \mathcal{E}(g, g^{q-1}) \quad (63)$$

for any $1 < q < \infty$. With $q(t) := 1 + e^{2t/\tau_l}$, we have $q'(t) = \frac{2}{\tau_l}(q(t) - 1)$, and replacing g by $\mathbf{P}_t g$ in (63) we obtain

$$\frac{q'(t)}{q(t)} \mathcal{L}_{q(t)}(\mathbf{P}_t g) - \mathcal{E}\left(\mathbf{P}_t g, (\mathbf{P}_t g)^{q(t)-1}\right) \leq 0.$$

From (59) we then find $F'(t) \leq 0$ for all $t \geq 0$. Since $F(0) = \|g\|_2$, this implies

$$\|\mathbf{P}_t g\|_{q(t)} \leq \|g\|_2. \tag{64}$$

We have assumed $g \geq 0$, but (64) now extends trivially to general g, and therefore

$$\|\mathbf{P}_t\|_{2 \to q(t)} \leq 1.$$

This gives the desired hypercontractivity assertion (57). ■

Here is the technical Dirichlet form lemma that was used in the proof of Theorem 24.

Lemma 25 $\mathcal{E}(g, g^{q-1}) \geq \frac{4(q-1)}{q^2} \mathcal{E}(g^{q/2}, g^{q/2})$ *for* $g \geq 0$ *and* $1 < q < \infty$.

xxx Do we somewhere have the following?:

$$\mathcal{E}(f, g) = \frac{1}{2} \sum_i \sum_{j \neq i} \pi_i q_{ij} (f(i) - f(j))(g(i) - g(j)). \tag{65}$$

Proof. For any $0 \leq a < b$

$$\begin{aligned} \left(\frac{b^{q/2} - a^{q/2}}{b - a}\right)^2 &= \left(\frac{q}{2(b-a)} \int_a^b t^{\frac{q}{2}-1}\, dt\right)^2 \\ &\leq \frac{q^2}{4(b-a)} \int_a^b t^{q-2}\, dt = \frac{q^2}{4(q-1)} \frac{b^{q-1} - a^{q-1}}{b - a}. \end{aligned}$$

This shows that

$$(b^{q-1} - a^{q-1})(b - a) \geq \frac{4(q-1)}{q^2} (b^{q/2} - a^{q/2})^2$$

and the lemma follows easily from this and (65). ■

Now we are prepared to bound $\hat{\tau}$ in terms of τ_l.

Theorem 26 *(a) If $c \geq 0$, then for any state i with $\pi_i \leq e^{-1}$,*

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \leq e^{1-c} \text{ for } t \geq \tfrac{1}{2}\tau_l \log\log \tfrac{1}{\pi_i} + c\tau_2.$$

(b)

$$\hat{\tau} \leq \tfrac{1}{2}\tau_l \log\log \tfrac{1}{\pi_*} + 2\tau_2 \leq \tau_l(\tfrac{1}{2}\log\log \tfrac{1}{\pi_*} + 2).$$

Proof. Part (b) follows immediately from (54), part (a), and Lemma 22. To prove part (a), we begin with (55):

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \leq \pi_i^{-1/q} \|\mathbf{P}_s\|_{2\to q}\, e^{-(t-s)/\tau_2}.$$

As in the second half of the proof of Theorem 24, let $q = q(s) := 1 + e^{2s/\tau_l}$. Then $\|\mathbf{P}_s\|_{2\to q(s)} \leq 1$. Thus

$$\|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \leq \pi_i^{-1/q(s)}\, e^{-(t-s)/\tau_2}, \quad 0 \leq s \leq t.$$

Choosing $s = \frac{1}{2}\tau_l \log\log(\frac{1}{\pi_i})$ we have $q(s) = 1 + \log(\frac{1}{\pi_i})$ and thus

$$\|\mathbf{P}_i(X_t \in \cdot) - \pi(\cdot)\|_2 \leq \exp(1 - \tfrac{t-s}{\tau_2}) \text{ for } t \geq s. \quad \blacksquare$$

We have established the upper bound in the following corollary; for the lower bound, see Corollary 3.11 in [4].

Corollary 27

$$\tau_l \leq \hat{\tau} \leq \tau_l(\tfrac{1}{2}\log\log \tfrac{1}{\pi_*} + 2).$$

Examples illustrating the improvement Corollary 27 affords over the similar result (7) in terms of τ_2 are offered in Examples 37 and 40.

4.3 Exact computation of τ_l

Exact computation of τ_l is exceptionally difficult—so difficult, in fact, that τ_l is known only for a handful of examples. We present some of these examples in this subsection.

Example 28 *Trivial two-state chains.*

We consider a discrete-time chain on $\{0, 1\}$ that jumps in one step to stationarity (or, since the value of τ_l is unaffected by continuization, the corresponding continuized chain). Thus $(p_{00}, p_{01}, p_{10}, p_{11}) = (\theta, 1-\theta, \theta, 1-\theta)$ with $\theta = \pi_0 = 1 - \pi_1$. We also assume $0 < \theta \leq 1/2$. The claim is that

$$\tau_l = \begin{cases} \frac{\log[(1-\theta)/\theta]}{2(1-2\theta)} & \text{if } \theta \neq 1/2 \\ 1 & \text{if } \theta = 1/2. \end{cases} \tag{66}$$

Note that this is continuous and decreasing for $\theta \in (0, 1/2]$.

To prove (66), we need to show that $\mathcal{L}(g)/\mathcal{E}(g, g) \leq \tau(\theta)$ for every nonconstant g on $\{0, 1\}$, where $\tau(\theta)$ denotes the righthand side of (66), with equality for some g_0. First suppose $\theta \neq 1/2$. For the inequality, as at (60)–(61) we may suppose $g \geq 0$ and, by homogeneity,

$$E_\pi g = \theta g(0) + (1-\theta) g(1) = 1.$$

We will work in terms of the single variable

$$x := 1/(g(0) - g(1)),$$

so that

$$g(0) = 1 + \frac{1-\theta}{x}, \quad g(1) = 1 - \frac{\theta}{x}$$

and we must consider $x \in (-\infty, -(1-\theta)] \cup [\theta, \infty)$. We calculate

$$\begin{aligned}
\mathcal{E}(g, g) &= \theta(1-\theta)(g(0) - g(1))^2 = \theta(1-\theta)/x^2, \\
\|g\|_2^2 &= \theta\left(1 + \frac{1-\theta}{x}\right)^2 + (1-\theta)\left(1 - \frac{\theta}{x}\right)^2 \\
&= [\theta(x + 1 - \theta)^2 + (1-\theta)(x-\theta)^2]/x^2 = 1 + \frac{\theta(1-\theta)}{x^2}, \\
\ell(x) &:= \mathcal{L}(g) = \theta\left(1 + \frac{1-\theta}{x}\right)^2 \log\left(1 + \frac{1-\theta}{x}\right) \\
&\quad + (1-\theta)\left(1 - \frac{\theta}{x}\right)^2 \log\left(1 - \frac{\theta}{x}\right) \\
&\quad - \frac{1}{2}\theta\left(1 + \frac{\theta(1-\theta)}{x^2}\right) \log\left(1 + \frac{\theta(1-\theta)}{x^2}\right) \\
&= \Big[\theta(x + 1 - \theta)^2 \log(x + 1 - \theta)^2 + (1-\theta)(x-\theta)^2 \log(x-\theta)^2 \\
&\quad - (x^2 + \theta(1-\theta)) \log(x^2 + \theta(1-\theta))\Big] / (2x^2),
\end{aligned}$$

$$
\begin{aligned}
r(x) &:= 2\theta(1-\theta)\frac{\mathcal{L}(g)}{\mathcal{E}(g,g)} = 2x^2\ell(x) \\
&= \theta(x+1-\theta)^2\log(x+1-\theta)^2 + (1-\theta)(x-\theta)^2\log(x-\theta)^2 \\
&\qquad -(x^2+\theta(1-\theta))\log(x^2+\theta(1-\theta)).
\end{aligned}
$$

¿From here, a straightforward but very tedious calculus exercise shows that r decreases over $(-\infty, -(1-\theta)]$, with $r(-\infty) = 2\theta(1-\theta)$, and that r is strictly unimodal over $[\theta, \infty)$, with $r(\theta) = 0$ and $r(\infty) = 2\theta(1-\theta)$. It follows that $r(x)$ is maximized over $(-\infty, -(1-\theta)] \cup [\theta, \infty)$ by taking x to be the unique root to

$$
\begin{aligned}
0 &= r'(x) = 4\theta(x+(1-\theta))\log(x+(1-\theta)) \\
&\qquad + 4(1-\theta)(x-\theta)\log(x-\theta) - 2x\log(x^2+\theta(1-\theta))
\end{aligned} \tag{67}
$$

over (θ, ∞).

There is on hope for solving (67) explicitly unless

$$
x^2 + \theta(1-\theta) = (x+1-\theta)(x-\theta),
$$

i.e., $x = 2\theta(1-\theta)/(1-2\theta)$. Fortunately, this *is* a solution to (67), and it falls in (θ, ∞). The corresponding value of r is $\frac{\theta(1-\theta)}{1-2\theta}\log\frac{1-\theta}{\theta}$, so (66) follows, and we learn furthermore that the function g maximizing $\mathcal{L}(g)/\mathcal{E}(g,g)$ is g_0, with $g_0(0) = \frac{1}{2\theta}$ and $g_0(1) = \frac{1}{2(1-\theta)}$.

For $\theta = 1/2$, the major change is that now r is *increasing*, rather than unimodal, over $[\theta, \infty)$. Thus $r_{\sup} = 2\theta(1-\theta) = 1/2$, and (66) again follows.

Example 29 *Two-state chains.*

Now consider *any* irreducible chain (automatically reversible) on $\{0, 1\}$, with stationary distribution π. Without loss of generality we may suppose $\pi_0 \leq \pi_1$. We claim that

$$
\tau_l = \begin{cases} \frac{\pi_1 \log(\pi_1/\pi_0)}{2p_{01}(1-2\pi_0)} & \text{if } \pi_0 \neq 1/2 \\ \\ 1/(2p_{01}) & \text{if } \pi_0 = 1/2. \end{cases}
$$

The proof is easy. The functional $\mathcal{L}(g)$ depends only on π and so is unchanged from Example 28, and the Dirichlet form changes from $\mathcal{E}(g,g) = \pi_0\pi_1(g(0)-g(1))^2$ in Example 28 to $\mathcal{E}(g,g) = p_{01}(g(0)-g(1))^2$ here.

Remark. Recall from Chapter 5, Example yyy:4 that $\tau_2 = 1/(p_{01}+p_{10}) = \pi_1/p_{01}$. It follows that

$$\frac{\tau_l}{\tau_2} = \begin{cases} \frac{\log(\pi_1/\pi_0)}{2(1-2\pi_0)} & \text{if } 0 < \pi_0 < 1/2 \\ 1 & \text{if } \pi_0 = 1/2 \end{cases}$$

is a continuous and decreasing function of π_0. In particular, we have equality in Lemma 22 for a two-state chain if and only if $\pi_0 = 1/2$. Moreover,

$$\tau_l/\tau_2 \sim \tfrac{1}{2}\log(1/\pi_0) \to \infty \text{ as } \pi_0 \to 0.$$

Example 30 *Trivial chains.*

The proof of Lemma 22 and the result of Example 28 can be combined to prove the following result: For the "trivial" chain with $p_{ij} \equiv \pi_j$, the log-Sobolev time τ_l is given (when $\pi_* < 1/2$) by

$$\tau_l = \frac{\log(\frac{1}{\pi_*} - 1)}{2(1 - 2\pi_*)}.$$

We omit the details, referring the reader to Theorem 5.1 of [4].

As an immediate corollary, we get a reverse-inequality complement to Lemma 22:

Corollary 31 *For any reversible chain (with $\pi_* < 1/2$, which is automatic for $n \geq 3$),*

$$\tau_l \leq \tau_2 \frac{\log(\frac{1}{\pi_*} - 1)}{2(1 - 2\pi_*)}.$$

Proof. The result of Example 30 can be written

$$\mathcal{L}(g) \leq (\mathrm{var}_\pi\, g)\frac{\log(\frac{1}{\pi_*} - 1)}{2(1 - 2\pi_*)},$$

and $\mathrm{var}_\pi\, g \leq \tau_2 \mathcal{E}(g,g)$ by the extremal characterization of τ_2. ■

Example 32 *The complete graph.*

It follows readily from Example 30 that the continuized walk of the complete graph has

$$\tau_l = \frac{(n-1)\log(n-1)}{2(n-2)} \sim \frac{1}{2}\log n.$$

Since $\tau_2 = (n-1)/n$, equality holds in Corollary 31 for this example.

xxx Move the following warning to follow Corollary 27, perhaps? *Warning.* Although the ratio of the upper bound on $\hat{\tau}$ to lower bound in Corollary 27 is smaller than that in (7), the upper bound in Corollary 27 is sometimes of larger order of magnitude than the upper bound in (7). For the complete graph, (7) says

$$\tfrac{n-1}{n} \le \hat{\tau} \le \tfrac{n-1}{n}(\tfrac{1}{2}\log n + 1)$$

and Corollary 27 yields

$$(1+o(1))\tfrac{1}{2}\log n \le \hat{\tau} \le (1+o(1))\tfrac{1}{4}(\log n)(\log\log n),$$

while, from Chapter 5, yyy:(33) it follows that

$$\hat{\tau} = \tfrac{1}{2}\log n + O\left(\frac{\log n}{n}\right).$$

As another example, the product chain development in the next subsection together with Example 29 will give τ_l exactly for the d-cube. On the other hand, the exact value of τ_l is unknown even for many of the simplest examples in Chapter 5. For instance,

Open Problem 33 Calculate τ_l for the n-cycle (Chapter 5 Example yyy:7) when $n \ge 4$.

xxx For NOTES: $n = 3$ is complete graph K_3, covered by Example 32. ($\tau_l = \log 2$ for $n = 3$.)

Notwithstanding Open Problem 33, the value of τ_l is known *up to multiplicative constants*. Indeed, it is shown in Section 4.2 in [4] that

$$\frac{1}{4\pi^2}n^2 \le \tau_l \le \frac{25}{16\pi^2}n^2.$$

Here is a similar result we will find useful later in dealing with our running example of the grid.

Example 34 *The m-path with end self-loops.*

For this example, discussed above in Example 16, we claim

$$\frac{2}{\pi^2} m^2 \leq \tau_l \leq m^2.$$

The lower bound is easy, using Lemma 22:

$$\tau_l \geq \tau_2 = (1 - \cos(\pi/m))^{-1} \geq \frac{2}{\pi^2} m^2.$$

For the upper bound we use Corollary 27 and estimation of $\hat{\tau}$. Indeed, in Example 16 it was shown that

$$\hat{d}(2t) = N^2(t) - 1 \leq \left[1 + (4t/m^2)^{-1/2}\right] \exp(-4t/m^2), \quad t > 0.$$

Substituting $t = m^2$ gives $\sqrt{\hat{d}(2t)} \leq \sqrt{3/2}\, e^{-2} < e^{-1}$, so $\tau_l \leq \hat{\tau} \leq m^2$.

xxx P.S. Persi (98/07/02) points out that H. T. Yau showed $\tau_l = \Theta(n \log n)$ for random transpositions by combining $\tau_l \geq \tau_2$ (Lemma 22) and $\tau_l \leq \mathcal{L}(g_0)/\mathcal{E}(g_0, g_0)$ with g_0 = delta function. I have written notes generalizing and discussing this and will incorporate them into a later version.

4.4 τ_l and product chains

xxx Remind reader of definition of product chain in continuous time given in Chapter 4 Section yyy:6.2.

xxx Motivate study as providing benchmark chains for comparison method.

xxx Recall from Chapter 4, yyy:(42):

$$\tau_2 = \max(\tau_2^{(1)}, \tau_2^{(2)}). \tag{68}$$

xxx Product chain has transition rates equal (off diagonal) to

$$q_{(i_1,i_2),(j_1,j_2)} = \begin{cases} q^{(1)}_{i_1,j_1} & \text{if } i_1 \neq j_1 \text{ and } i_2 = j_2 \\ q^{(2)}_{i_2,j_2} & \text{if } i_1 = j_1 \text{ and } i_2 \neq j_2 \\ 0 & \text{otherwise.} \end{cases} \tag{69}$$

xxx Dirichlet form works out very nicely for products:

Lemma 35

$$\mathcal{E}(g,g) = \sum_{i_2} \pi^{(2)}_{i_2} \mathcal{E}^{(1)}(g(\cdot,i_2), g(\cdot,i_2)) + \sum_{i_1} \pi^{(1)}_{i_1} \mathcal{E}^{(2)}(g(i_1,\cdot), g(i_1,\cdot)).$$

Proof. This follows easily from (69) and the definition of $\mathcal{E}$ in Chapter 3 Section yyy:6.1 (cf. (68)). ■

The analogue of (68) for the log-Sobolev time is also true:

xxx For NOTES?: Can give analagous proof of (68): see my notes, page 8.4.24A.

Theorem 36 *For a continuous-time product chain,*

$$\tau_l = \max(\tau_l^{(1)}, \tau_l^{(2)}).$$

Proof. The keys to the proof are Lemma 35 and the following "law of total $\mathcal{L}$-functional." Given a function $g \not\equiv 0$ on the product state space $I = I_1 \times I_2$, define a function $G_2 \not\equiv 0$ on I_2 by

$$G_2(i_2) := \|g(\cdot,i_2)\|_2 = \left(\sum_{i_1} \pi_{i_1} g^2(i_1,i_2) \right)^{1/2}.$$

Then

$$\begin{aligned} \mathcal{L}(g) &= \sum_{i_1,i_2} \pi_{i_1,i_2}\, g^2(i_1,i_2)\, [\log(|g(i_1,i_2)|/G_2(i_2)) + \log(G_2(i_2)/\|g\|_2)] \\ &= \sum_{i_2} \pi^{(2)}_{i_2} \mathcal{L}^{(1)}(g(\cdot,i_2)) + \mathcal{L}^{(2)}(G_2), \end{aligned}$$

where we have used

$$\|G_2\|_2^2 = \|g\|_2^2.$$

Thus, using the extremal characterization (definition) (53) of $\tau_l^{(1)}$ and $\tau_l^{(2)}$,

$$\mathcal{L}(g) \le \tau_l^{(1)} \sum_{i_2} \pi^{(2)}_{i_2} \mathcal{E}^{(1)}(g(\cdot,i_2), g(\cdot,i_2)) + \tau_l^{(2)} \mathcal{E}^{(2)}(G_2,G_2). \qquad (70)$$

But from

$$|G_2(j_2) - G_2(i_2)| = |\|g(\cdot,j_2)\|_2 - \|g(\cdot,i_2)\|_2| \le \|g(\cdot,j_2) - g(\cdot,i_2)\|_2$$

follows

$$\mathcal{E}^{(2)}(G_2, G_2) \le \sum \pi^{(1)}_{i_1} \mathcal{E}^{(2)}(g(i_1, \cdot), g(i_1, \cdot)). \tag{71}$$

From (70), (71), Lemma 35, and the extremal characterization of τ_l we conclude $\tau_l \le \max(\tau_l^{(1)}, \tau_l^{(2)})$. Testing on functions that depend only on one of the two variables shows that $\tau_l = \max(\tau_l^{(1)}, \tau_l^{(2)})$. ■

Theorem 36 extends in the obvious fashion to higher-dimensional products.

Example 37 *The d-cube.*

The continuized walk on the d-cube (Chapter 5, Example yyy:15) is simply the product of d copies of the continuized walk on the 2-path, each run at rate $1/d$. Therefore, since the log-Sobolev time for the 2-path equals $1/2$ by Example 29, the corresponding time for the d-cube is

$$\tau_l = d/2 = \tau_2.$$

¿From this and the upper bound in Corollary 27 we can deduce

$$\hat{\tau} \le \tfrac{1}{4} d \log d + (1 - \tfrac{1}{4} \log \tfrac{1}{\log 2}) d.$$

As discussed in this chapter's introduction, this bound is remarkably sharp and improves significantly upon the analogous bound that uses only knowledge of τ_2. xxx Recall corrections marked on pages 8.2.11–12 of my notes.

4.5 The comparison method for bounding τ_l

In Section 1 we compared relaxation times for two chains by using the extremal characterization and comparing Dirichlet forms and variances. For comparing variances, we used the characterization

$$\mathrm{var}_\pi\, g = \min_{c \in \mathbf{R}} \|g - c\|_2.$$

To extend the comparison method to log-Sobolev times, we need the following similar characterization of $\mathcal{L}$.

xxx For NOTES: Cite [8].

Lemma 38 *The functional $\mathcal{L}$ in (52) satisfies*

$$\mathcal{L}(g) = \min_{c>0} \sum_i \pi_i L(g(i), c), \quad g \not\equiv 0, \tag{72}$$

with

$$L(g(i), c) := g^2(i) \log(|g(i)|/c) - \tfrac{1}{2}(g^2(i) - c^2) \ge 0. \tag{73}$$

Proof. We compute

$$\begin{aligned} f(c) &:= 2\sum_i \pi_i L(g(i), c^{1/2}) = E_\pi(g^2 \log |g|^2) - \|g\|_2^2 \log c - \|g\|_2^2 + c, \\ f'(c) &= 1 - c^{-1}\|g\|_2^2, \quad f''(c) = c^{-2}\|g\|_2^2 > 0. \end{aligned}$$

Thus f is strictly convex and minimized by the choice $c = \|g\|_2^2$, and so

$$\min_{c>0} \sum_i \pi_i L(g(i), c) = \tfrac{1}{2} \min_{c>0} f(c) = \tfrac{1}{2} f(\|g\|_2^2) = \mathcal{L}(g).$$

This proves (72). Finally, applying the inequality

$$x \log(x/y) - (x - y) \ge 0 \quad \text{for all } x \ge 0,\, y > 0$$

to $x = g^2(i)$ and $y = c^2$ gives the inequality in (73). ■

Now it's easy to see how to compare log-Sobolev times, since, adopting the notation of Section 1, Lemma 38 immediately yields the analogue

$$\mathcal{L}(g) \le \tilde{\mathcal{L}}(g) \max_i (\pi_i/\tilde{\pi}_i)$$

of (18). In the notation of Corollary 2, we therefore have

Corollary 39 (comparison of log-Sobolev times)

$$\tau_l \le \frac{A}{a} \tilde{\tau}_l.$$

Example 40 *Random walk on a d-dimensional grid.*

xxx Remarked in Example 34 that $\tau_l \le m^2$ for m-path with end self-loops.

xxx So by Theorem 36, benchmark product chain has $\tilde{\tau}_l \le dm^2$.

Recalling $A \le 1$ and $a \ge 1/2$ from Example 5, we therefore find

$$\tau_l \le 2m^2 d \tag{74}$$

for random walk on the grid. Then Theorem 26(b) gives

$$\hat{\tau} \le m^2 d(\log\log(2n) + 4),$$

which is of order $m^2 d(\log d + \log\log m)$. This is an improvement on the τ_2-only bound $O(m^2 d^2 \log m)$ of (50) and may be compared with the Nash-based bound $O(m^2 d^2 \log d)$ of (51). In Example 43 we will combine Nash-inequality and log-Sobolev techniques to get a bound of order $m^2 d \log d$

xxx right for TV.

5 Combining the techniques

To get the maximum power out of the techniques of this chapter, it is sometimes necessary to combine the various techniques. Before proceeding to a general result in this direction, we record a simple fact. Recall (36).

Lemma 41 *If q and q^* are conjugate exponents with $2 \le q \le \infty$, then*

$$\|f\|_{q^*} \le \|f\|_1^{1-\frac{2}{q}} \|f\|_2^{\frac{2}{q}} \quad \textit{for all } f.$$

Proof. Apply Hölder's inequality

$$\|gh\|_1 \le \|g\|_p \, \|h\|_{p^*}$$

with

$$g = |f|^{(q-2)/(q-1)}, \quad h = |f|^{2/(q-1)}, \quad p = \frac{q-1}{q-2}. \quad \blacksquare$$

Theorem 42 *Suppose that a continuous-time reversible chain satisfies*

$$N(t) \le Ct^{-D} \quad \textit{for } 0 < t \le T \tag{75}$$

for some constants C, T, D satisfying $CT^{-D} \ge e$. If $c \ge 0$, then

$$\sqrt{\hat{d}(2t)} = \max_i \|P_i(X_t \in \cdot) - \pi(\cdot)\|_2 \le e^{2-c}$$

for

$$t \ge T + \tfrac{1}{2}\tau_l \log\left[\log(CT^{-D}) - 1\right] + c\tau_2,$$

where τ_2 is the relaxation time and τ_l is the log-Sobolev time.

Proof. ¿From Lemma 11 and a slight extension of (34), for any $s, t, u \ge 0$ and any initial distribution we have

$$\|P(X_{s+t+u} \in \cdot) - \pi(\cdot)\|_2 \le \|P(X_s \in \cdot)\|_{q^*} \, \|\mathbf{P}_t\|_{q^* \to 2} \, e^{-u/\tau_2}$$

for any $1 \le q^* \le \infty$. Choose $q = q(t) = 1 + e^{2t/\tau_l}$ and q^* to be its conjugate. Then, as in the proof of Theorem 26(a),

$$\|\mathbf{P}_t\|_{q^* \to 2} = \|\mathbf{P}_t\|_{2 \to q} \le 1.$$

According to Lemma 41, (39), and (75), if $0 < s \le T$ then

$$\|P(X_s \in \cdot)\|_{q^*} \le \|P(X_s \in \cdot)\|_2^{2/q} \le N(s)^{2/q} \le (Cs^{-D})^{2/q}.$$

Now choose $s = T$. Combining everything so far,

$$\sqrt{\hat{d}(2(T+t+u))} \le (CT^{-D})^{2/q(t)}\, e^{-u/\tau_2} \quad \text{for } t, u \ge 0.$$

The final idea is to choose t so that the first factor is bounded by e^2. ¿From the formula for $q(t)$, the smallest such t is

$$\tfrac{1}{2}\tau_l \log\left[\log(CT^{-D}) - 1\right].$$

With this choice, the theorem follows readily. ■

Example 43 *Random walk on a d-dimensional grid.*

Return one last time to the walk of interest in Example 5. Example 21 showed that (75) holds with

$$D = d/4, \quad C = e(2^{14}d^2m^2)^{d/4} = e(2^7 dm)^{d/2}, \quad T = dm^2/32.$$

Also recall $\tau_2 \le \frac{1}{2}dm^2$ from (49) and $\tau_l \le 2dm^2$ from Example 40. Plugging these into Theorem 42 with $c = 2$ yields

$$\hat{\tau} \le \tfrac{49}{32}m^2 d \log[\tfrac{1}{4}d\log d + \tfrac{19}{4}d\log 2], \text{ which is } \le 5m^2 d\log d \text{ for } d \ge 2.$$

xxx Finally of right order of magnitude.

6 Notes on Chapter 8

xxx The chapter notes go here. Currently, they are interspersed throughout the text.

xxx Also cite and plug [12].

References

[1] P. Diaconis. *Group Representations in Probability and Statistics.* Institute of Mathematical Statistics, Hayward CA, 1988.

[2] P. Diaconis and L. Saloff-Coste. Comparison theorems for random walk on finite groups. *Ann. Probab.*, 21:2131–2156, 1993.

[3] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Ann. Appl. Probab.*, 3:696–730, 1993.

[4] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6:695–750, 1996.

[5] P. Diaconis and L. Saloff-Coste. Nash inequalities for finite Markov chains. *J. Theoretical Probab.*, 1996.

[6] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57:159–179, 1981.

[7] J.A. Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.*, 1:62–87, 1991.

[8] R. Holley and D. Stroock. Logarithmic Sobolev inequalitites and stochastic Ising models. *J. Statist. Phys.*, 46:1159–1194, 1987.

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

[10] O. S. Rothaus. Diffusion on compact Riemannian manifolds and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 42:102–109, 1981.

[11] W. Rudin. *Real and Complex Analysis*. McGraw–Hill Book Co., New York, 3rd edition, 1987.

[12] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, pages 301–413. Springer, Berlin, 1997.