Chapter 2

Bounding the Mixing Time

2.1 Spectral Gap

Let P be the transition matrix of an ergodic, reversible Markov chain on state space Ω, Let π be its stationary distribution. Let $N = |\Omega|$ and assume w.l.o.g. that $\Omega = \{0, 1, \ldots, N-1\}$. Let the eigenvalues of P be $1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1}$. They are all real valued. Let $\lambda_{\max} = \max\{|\lambda_i| : i > 0\}$. The fact that $\lambda_{\max} < 1$ is a classical result of the theory of non-negative matrices. The *spectral gap* $1 - \lambda_{\max}$ determines the mixing rate of the chain in an essential way. The larger it is, the more rapidly does the chain mix. For $U \subseteq \Omega$ let

$$\Delta_U(t) = \max_{i,j \in U} \left\{ \frac{|P^t(i,j) - \pi(j)|}{\pi(j)} \right\}.$$

Theorem 2.1.1 *For all $U \subseteq \Omega$ and $t \geq 0$,*

$$\Delta_U \leq \frac{\lambda_{\max}^t}{\min\limits_{i \in U} \pi(i)}.$$

Proof Let $D^{1/2}$ be the diagonal $\Omega \times \Omega$ matrix with diagonal entries $\sqrt{\pi(\omega)}$, $\omega \in \Omega$ and let $D^{-1/2}$ be its inverse. Then the reversibility of of the chain (1.15) implies that the matrix $S = D^{1/2} P D^{-1/2}$ is symmetric. It has the same eigenvalues as P and its symmetry means that these are all real. We can select an orthonormal basis of column vectors $\mathbf{e}^{(i)}$, $i \in \Omega$ for $\mathbb{R}^\Omega$ consisting of left eigenvectors of S where $\mathbf{e}^{(i)}$ has associated eigenvalue λ_i and $\mathbf{e}^{(0)} = \pi^T D^{-1/2}$. S has the *spectral decomposition*

$$S = \sum_{i=0}^{N-1} \lambda_i \mathbf{e}^{(i)} \mathbf{e}^{(i)T} = \sum_{i=0}^{N-1} \lambda_i E^{(i)},$$

where $E^{(i)} = \mathbf{e}^{(i)}\mathbf{e}^{(i)^T}$. Note that $E^{(i)}E^{(j)} = 0$ for $i \neq j$ and $E^{(i)^2} = E^{(i)}$. It follows that for any $t = 0, 1, 2, \dots$, $S^t = \sum_{i=0}^{N-1} \lambda_i^t E^{(i)}$. Hence

$$\begin{aligned} P^t = D^{-1/2} S^t D^{1/2} &= \sum_{i=0}^{N-1} \lambda_i^t (D^{-1/2}\mathbf{e}^{(i)}))(\mathbf{e}^{(i)^T} D^{1/2}) \\ &= \mathbf{1}_N \pi^T + \sum_{i=1}^{N-1} \lambda_i^t (D^{-1/2}\mathbf{e}^{(i)}))(\mathbf{e}^{(i)^T} D^{1/2}), \end{aligned}$$

where $\mathbf{1}_N$ is the N-vector all of whose components are 1. In component form, we get with the help of the Cauchy-Schwartz inequality:

$$\begin{aligned} |P^t(j,k) - \pi_k| &= \left| \sqrt{\frac{\pi_k}{\pi_j}} \sum_{i=1}^{N-1} \lambda_i^t \mathbf{e}_j^{(i)} \mathbf{e}_k^{(i)} \right| \\ &\leq \sqrt{\frac{\pi_k}{\pi_j}} \lambda_{\max}^t \left(\sum_{i=0}^{N-1} \mathbf{e}_j^{(i)^2} \right)^{1/2} \left(\sum_{i=0}^{N-1} \mathbf{e}_k^{(i)^2} \right)^{1/2} \\ &= \sqrt{\frac{\pi_k}{\pi_j}} \lambda_{\max}^t. \end{aligned} \tag{2.1}$$

The theorem follows by substitution of the above inequality in the definition of Δ_U. □

In terms of mixing time we have

Corollary 2.1.1

$$\tau(\varepsilon) \leq \left\lceil \frac{\log \varepsilon \pi_{\min}}{\log \lambda_{\max}} \right\rceil.$$

Proof For $A \subseteq \Omega$ we have

$$p_t(A) - \pi(A) \leq \frac{\lambda_{\max}^t}{\pi_{\min}} \pi(A) \leq \frac{\lambda_{\max}^t}{\pi_{\min}}.$$

□

As an example we consider random walk $\mathcal{W}_n$ on the unit hypercube. Here the graph is the n-cube $Q_n = (X_n = \{0,1\}^n, E_n)$ where $x, y \in X_n$ are adjacent in Q_n if their Hamming distance is one i.e. if $|\{i \in [n] : x_i \neq y_i\}| = 1$. We add n self loops to each vertex to make the chain lazy.

If G is a d-regular graph without loops and A_G is its adjacency matrix then the probability transition matrix P_G of a random walk on G satisfies $P_G = d^{-1} A_G$.

For graphs $G_i = (V_i, E_i), i = 1, 2$ we can define their product $G = G_1 \times G_2 = (V, E)$ where $V = V_1 \times V_2$ and $E = \{((v_1, v_2), (w_1, w_2)) : v_1 = w_1 \text{ and } (v_2, w_2) \in E_2 \text{ or } v_2 = w_2 \text{ and } (v_1, w_1) \in E_1\}$. Then

$$Q_n = K_2 \times K_2 \times \cdots \times K_2 \qquad (n \text{ fold product}). \tag{2.2}$$

Theorem 2.1.2 *If $\mu_i, i = 1, 2, \ldots, m$ and $\nu_i, i = 1, 2, \ldots, n$ are the eigenvalues of matrices A_{G_1}, A_{G_2} respectively, then the eigenvalues of A_G are $\{\mu_i + \nu_j : 1 \leq i \leq m, 1 \leq j \leq n\}$.*

Proof A_G can be obtained from A_{G_1} by replacing each 1 by the $|V_2|$ identity matrix I_2, the off-diagonal 0's by the $|V_2| \times |V_2|$ matrix of 0's and replacing each diagonal entry by A_{G_2}. So if $p_G(\lambda) = \det(\lambda I - A_G)$ then

$$p_G(\lambda) = \det p_{G_1}(\lambda I_2 - A_{G_2}).$$

This follows from the following: Suppose the $mn \times mn$ matrix A is decomposed into an $m \times m$ matrix of $n \times n$ blocks $A_{i,j}$. Suppose also that the $A_{i,j}$ commute among themselves. Then

$$\det A = \det \left(\sum_\sigma (-1)^{\text{sign}(\sigma)} \prod_{i=1}^m A_{i,\sigma(i)} \right),$$

i.e. one can produce an $m \times m$ matrix by a "determinant" calculation and then take its determinant.

So

$$p_G(\lambda) = \det \prod_{i=1}^n (\lambda I_2 - A_{G_2} - \mu_i I_2) = \prod_{i=1}^n p_{G_2}(\lambda - \mu_i) = \prod_{i=1}^n \prod_{j=1}^n (\lambda - \mu_i - \nu_j).$$

□

The eigenvalues of K_2 are $\{1, -1\}$ and applying (2.2) we see that the eigenvalues of Q_n are $\{0, \pm 1, \pm 2, \ldots, \pm n\}$ (ignoring multiplicities). To get the eigenvalues for our random walk we (i) divide by n and then (ii) replace each eigenvalue λ by $\frac{1+\lambda}{2}$ to account for adding loops. Thus the second eigenvalue of the walk is $1 - \frac{1}{2n}$.

Applying Corollary 2.1.1 we obtain $\tau(\varepsilon) \leq \log(\varepsilon^{-1}) + O(n^2)$. This is a poor estimate, due to our use of the Cauchy-Schwartz inequality in the proof of Theorem 2.1.1. We get an easier and better estimate by using *coupling*.

2.1.1 Decomposition Theorem

2.2 Conductance

The conductance Φ of $\mathcal{M}$ is defined by

$$\Phi = \min\{\Phi_S : S \subseteq \Omega,\ 0 < \pi(S) \leq 1/2\}$$

where if $Q(\omega, \sigma) = \pi(\omega) P(\omega, \sigma)$ and $\bar{S} = \Omega \setminus S$,

$$\Phi_S = \pi(S)^{-1} Q(S, \bar{S}).$$

Thus Φ_S is the steady state probability of moving from S to $\bar{S}$ in one step of the chain, conditional on being in S.

Clearly $\Phi \leq \frac{1}{2}$ if $\mathcal{M}$ is lazy.

Note that

$$\Phi_S \pi(S) = Q(S, \bar{S}) = Q(\bar{S}, S) = \Phi_{\bar{S}} \pi(\bar{S}). \tag{2.3}$$

Indeed,

$$Q(S, \bar{S}) = Q(\Omega, \bar{S}) - Q(\bar{S}, \bar{S}) = \pi(\bar{S}) - Q(\bar{S}, \bar{S}) = Q(\bar{S}, S).$$

Let $\pi_{\min} = \min\{\pi(\omega) : \omega \in \Omega\} > 0$ and $\pi_{\max} = \max\{\pi(\omega) : \omega \in \Omega\}$.

2.2.1 Reversible Chains

In this section we show how conductance gives us an estimate of the spectral gap of a reversible chain.

Lemma 2.2.1 *If $\mathcal{M}$ is lazy and ergodic then all eigenvalues are positive.*

Proof $Q = 2P - I \geq 0$ is stochastic and has eigenvalues $\mu_i = 2\lambda_i - 1$, $i = 0, 1, \ldots N-1$. The result follows from $\mu_i > -1$, $i = 0, 1, \ldots N-1$. □

For $y \in \mathbb{R}^N$ let

$$\mathcal{E}(y, y) = \sum_{i<j} \pi_i P_{i,j} (y_i - y_j)^2.$$

Lemma 2.2.2 *If $\mathcal{M}$ is reversible then*

$$1 - \lambda_1 = \min_{\pi^T y = 0} \frac{\mathcal{E}(y, y)}{\sum_i \pi_i y_i^2}.$$

Proof Let $D, S, \mathbf{e}^{(0)}$ be as in Section 2.1. Then by the Rayleigh principle,

$$\lambda_1 = \max_{\pi^T D^{-1/2} x = 0} \frac{x^T D^{1/2} P D^{-1/2} x}{x^T x}.$$

Thus

$$\begin{aligned} 1 - \lambda_1 &= \min_{\pi^T D^{-1/2} x = 0} \frac{x^T D^{1/2} (I - P) D^{-1/2} x}{x^T x} \\ &= \min_{\pi^T y = 0} \frac{y^T D (I - P) y}{y^T D y}. \end{aligned} \tag{2.4}$$

Now

$$
\begin{aligned}
y^T D(I-P)y &= -\sum_{i\neq j} y_i y_j \pi_i P_{i,j} + \sum_i \pi_i (1-P_{i,i}) y_i^2 \\
&= -\sum_{i\neq j} y_i y_j \pi_i P_{i,j} + \sum_{i\neq j} \pi_i P_{i,j} \frac{y_i^2+y_j^2}{2} \\
&= \sum_{i<j} \pi_i P_{i,j} (y_i - y_j)^2 \\
&= \mathcal{E}(y,y),
\end{aligned}
$$

and the lemma follows from (2.4). □

Theorem 2.2.1 *If $\mathcal{M}$ is a reversible chain then*

$$1-\lambda_1 \geq \frac{\Phi^2}{2}.$$

Proof Assume now that $\pi^T y = 0$, $y_1 \geq y_2 \geq \cdots \geq y_N$ and that

$$\pi_1 + \pi_2 + \cdots + \pi_{r-1} \leq \frac{1}{2} < \pi_1 + \pi_2 + \cdots + \pi_r.$$

Let $z_i = y_i - y_r$ for $i = 1, 2, \ldots, n$. Then

$$z_1 \geq z_2 \geq \cdots \geq z_r = 0 \geq z_{r+1} \geq \cdots \geq z_N,$$

and

$$
\begin{aligned}
\frac{\mathcal{E}(y,y)}{\sum_i \pi_i y_i^2} &= \frac{\mathcal{E}(z,z)}{-y_r^2 + \sum_i \pi_i z_i^2} \\
&\geq \frac{\mathcal{E}(z,z)}{\sum_i \pi_i z_i^2}. \qquad (2.5) \\
&= \frac{\left(\sum_{i<j} \pi_i P_{i,j} (z_i - z_j)^2\right)\left(\sum_{i<j} \pi_i P_{i,j} (|z_i| + |z_j|)^2\right)}{\left(\sum_i \pi_i z_i^2\right)\left(\sum_{i<j} \pi_i P_{i,j} (|z_i| + |z_j|)^2\right)} \\
&= \frac{A}{B}, \qquad \text{say.}
\end{aligned}
$$

Now,

$$
\begin{aligned}
A &\geq \left(\sum_{i<j} \pi_i P_{i,j} |z_i - z_j| (|z_i| + |z_j|)\right)^2 \qquad \text{by Cauchy-Schwartz} \\
&\geq \left(\sum_{i<j} \pi_i P_{i,j} \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2|\right)^2. \qquad (2.6)
\end{aligned}
$$

We verify (2.6) later. Also,

$$\sum_{i<j} \pi_i P_{i,j}(|z_i| + |z_j|)^2 \leq 2\sum_{i<j} \pi_i P_{i,j}(z_i^2 + z_j^2) \leq 2\sum_i \pi_i z_i^2.$$

So,

$$\frac{\mathcal{E}(y,y)}{\sum_i \pi_i y_i^2} \geq \frac{A}{B} \geq \frac{\left(\sum_{i<j} \pi_i P_{i,j} \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2|\right)^2}{2\left(\sum_i \pi_i z_i^2\right)^2.}$$

Now let $S_k = \{1, 2, \ldots, k\}$ and $C_k = \{(i,j) : i \leq k < j\}$. Then

$$\begin{aligned}
\sum_{i<j} \pi_i P_{i,j} \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2| &= \sum_{k=1}^{N-1} |z_{k+1}^2 - z_k^2| \sum_{(i,j)\in C_k} \pi_i P_{i,j} \\
&\geq \Phi\left(\sum_{k=1}^{r-1}(z_k^2 - z_{k+1}^2)\pi(S_k) + \sum_{k=r}^{N-1}(z_{k+1}^2 - z_k^2)(1 - \pi(S_k))\right) \\
&= \Phi\left(\sum_{k=1}^{N-1}(z_k^2 - z_{k+1}^2)\pi(S_k) + (z_N^2 - z_r^2)\right) \\
&= \Phi\left(\sum_{k=1}^{N} \pi_k z_k^2\right)
\end{aligned}$$

since $z_r = 0$.

Thus if $\pi^T y = 0$ then

$$\frac{\mathcal{E}(y,y)}{\sum_i \pi_i y_i^2} \geq \frac{\Phi^2}{2}$$

and Theorem 2.2.1 follows.

Proof of (2.6)

We show that if $i < j$ then

$$|z_i - z_j|(|z_i| + z_j|) \geq \sum_{k=i}^{j-1} |z_{k+1}^2 - z_k^2|. \tag{2.7}$$

If $r \notin \{i, i+1, \ldots, j\}$ i.e. if z_i, z_j have the same sign then LHS(2.7)=RHS(2.7)=$|z_i^2 - z_j^2|$. Otherwise LHS(2.7)=$(|z_i| + |z_j|)^2$ and RHS(2.7)=$z_i^2 + z_j^2$. □

In terms of mixing time we obtain from Corollary 2.1.1,

Corollary 2.2.1 *If $\mathcal{M}$ is a lazy ergodic chain then*

$$\tau(\varepsilon) \leq \left\lceil \frac{2|\log \varepsilon\pi_{\min}|}{\Phi^2} \right\rceil.$$

Proof Lemma 2.2.1 implies that $\lambda_1 = \lambda_{\max}$ and then

$$\frac{1}{\log \lambda_{\max}^{-1}} \leq \frac{1}{\log(1-\Phi^2/2)^{-1}} \leq \frac{2}{\Phi^2}.$$

□

Now consider the conductance of a random walk on a graph $G = (V, E)$. For $S, T \subseteq V$ let $E(S,T) = \{(v,w) \in E : \ v \in S, w \in T\}$ and $e(S,T) = |E(S,T)$. Then, by definition,

$$\Phi_S = \frac{\displaystyle\sum_{(v,w)\in E(S,\bar{S})} \frac{d_v}{2|E|}\frac{1}{d_v}}{\displaystyle\sum_{v\in S} \frac{d_v}{2|E|}} = \frac{e(S,\bar{S})}{\displaystyle\sum_{v\in S} d_v}.$$

In particular when G is an r-regular graph

$$\Phi = r^{-1} \min_{|S| \leq \frac{1}{2}|V|} \frac{e(S,\bar{S})}{|S|}. \tag{2.8}$$

The minimand above is referred to as the *expansion* of G. This graphs with good expansion (*expander graphs*) have large conductance and random walks on them mix rapidly.

As an example consider the n-cube Q_n. For $S \subseteq X_n$ let $in(S)$ denote the number of edges of Q_n which are wholly contained in S.

Lemma 2.2.3 *If $\emptyset \neq S \subseteq X_n$ then $in(S) \leq \frac{1}{2}|S| \log_2 |S|$.*

Proof We prove this by induction on n. It is trivial for $n = 1$. For $n > 1$ let $S_i = \{x \in S : \ x_n = i\}$ for $i = 1, 2$. Then

$$in(S) \leq in(S_0) + in(S_1) + \min\{|S_0|, |S_1|\}$$

since the term $\min\{|S_0|, |S_1|\}$ bounds the number of edges which are contained in S and join S_0, S_1. The lemma follows from the inequality

$$x \log_2 x + y \log_2 y + 2y \leq (x+y) \log_2(x+y)$$

for all $x \geq y \geq 0$. The proof is left as a simple exercise in calculus. □

By summing the degrees at each vertex of S we see that

$$e(S,\bar{S}) + 2in(S) = n|S|.$$

By the above lemma we have

$$e(S,\bar{S}) \geq n|S| - \tfrac{1}{2}|S| \log_2 |S| \geq |S|$$

assuming $|S| \leq 2^{n-1}$. It follows from (2.8) that $\Phi \geq \frac{1}{n}$. Adding the self-loops to delay the walk will halve the conductance – the denominator $\sum_{v \in S} d_v$ doubles without changing the numerator in the definition of Φ_S. This gives us the estimate of $\frac{1}{8n^2}$ for the spectral gap, which is off by a factor of n – see Section 2.1.

We finish this section by proving a sort of converse to Theorem 2.2.1.

Theorem 2.2.2 *If $\mathcal{M}$ is a reversible chain then*

$$1 - \lambda_{\max} \leq 2\Phi$$

Proof We use Lemma 2.2.2. Let S be a set of states which minimises Φ_S and define y by $y_j = \frac{1}{\pi(S)}$ if $j \in S$ and $y_j = -\frac{1}{\pi(\bar{S})}$ if $j \in \bar{S}$. It is easy to check that $\pi^T y = 0$. Then

$$\mathcal{E}(y, y) = \left(\frac{1}{\pi(S)} + \frac{1}{\pi(\bar{S})}\right)^2 Q(S, \bar{S}) \text{ and } \sum \pi_i y_i^2 = \frac{1}{\pi(S)} + \frac{1}{\pi(\bar{S})}.$$

Thus

$$1 - \lambda_{\max} \leq \Phi_S \pi(S) \left(\frac{1}{\pi(S)} + \frac{1}{\pi(\bar{S})}\right) \leq 2\Phi_S = 2\Phi.$$

□

2.2.2 General Chains

Theorem 2.2.3 *Suppose that $\mathcal{M}$ is lazy and*

$$\pi_{\max} \leq \frac{\Phi^2}{20}.$$

Then

$$|p_t(\omega) - \pi(\omega)| \leq \pi_{\min}^{-1/2} \left(1 - \tfrac{1}{2}\Phi^2\right)^t.$$

Proof For $0 \leq x \leq 1$ let

$$h_t(x) = \max\left\{\sum_{\omega \in \Omega} (p^t(\omega) - \pi(\omega))\xi(\omega) : \ \xi \in [0,1]^\Omega, \sum_{\omega \in \Omega} \xi(\omega)\pi(\omega) = x\right\}.$$

By putting $\xi = 1_{\omega \in S}$ in the above definition we see that

$$p_t(S) - \pi(S) \leq h_t(\pi(S))$$

for all $S \subseteq \Omega$.

So, in particular $p_t(\omega) - \pi(\omega) \le h_t(\pi(\omega))$ $(x = \pi(\omega))$ and $\pi(\omega) - p_t(\omega) \le h_t(1-\pi(\omega))$ $(x = 1-\pi(\omega))$ and so

$$|p_t(\omega) - \pi(\omega)| \le \min\{h_t(\pi(\omega)), h_t(1-\pi(\omega))\}. \tag{2.9}$$

Order the elements $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$ so that

$$\frac{p_t(\omega_1)}{\pi(\omega_1)} \ge \frac{p_t(\omega_2)}{\pi(\omega_2)} \ge \cdots \ge \frac{p_t(\omega_N)}{\pi(\omega_N)},$$

and let $\theta_k = \sum_{i=1}^k \pi(\omega_i)$. Find the index k such that $\theta_{k-1} \le x < \theta_k$. Then

$$h_t(x) = \sum_{i=1}^{k-1}(p_t(\omega_i) - \pi(\omega_i)) + \frac{x-\theta_{k-1}}{\pi(\omega_k)}(p_t(\omega_k) - \pi(\omega_k)).$$

This is because putting

$$\xi(\omega_i) = \begin{cases} 1 & i < k \\ \frac{x-\theta_{k-1}}{\pi(\omega_k)} & i = k \\ 0 & i > k \end{cases}$$

yields an optimal basic feasible solution to the linear program in the definition of $h_t(x)$.

It follows that $h_t(x)$ is a concave piece-wise linear function on the interval $[0,1]$ with breakpoints at $0 = \theta_0 < \theta_1 < \cdots < \theta_N = 1$. Trivially, $h_t(0) = h_t(1) = 0$ and $0 \le h_t(x) \le 1$ for all t and x.

Now let

$$C = \max\left\{\frac{h_0(x)}{\min\{\sqrt{x}, \sqrt{1-x}\}} : 0 < x < 1\right\}.$$

If $a, b, c, d \ge 0$ then the function $f(\xi) = (a + b\xi)/\sqrt{c + d\xi}$ is monotone on $[0,1]$ and so the value of x defining C must occur at one of the breakpoints of h_t. It follows easily that

$$\begin{aligned} C &\le \max_{S\subseteq\Omega} \frac{|\pi_0(S) - \pi(S)|}{\min\left\{\sqrt{\pi(S)}, \sqrt{1-\pi(S)}\right\}} \\ &= \max_{\substack{S\subseteq\Omega \\ \pi(S)\le 1/2}} \frac{|\pi_0(S) - \pi(S)|}{\sqrt{\pi(S)}} \\ &\le \frac{1}{\sqrt{\pi_{\min}}}. \end{aligned} \tag{2.10}$$

(The second equation comes from considering $\Omega \setminus S$ when $\pi(S) \ge 1/2$.)

We now prove that for $t \ge 1$ and $x \in \{\theta_0, \theta_1, \dots, \theta_N\}$,

$$h_t(x) \le \tfrac{1}{2}(h_{t-1}(x - 2\Phi\min\{x, 1-x\}) + h_{t-1}(x + 2\Phi\min\{x, 1-x\})) \tag{2.11}$$

Fix k and let $u_i = \sum_{j \le k} p_{i,j}$ where $p_{i,j} = P(\omega_i, \omega_j)$. Clearly

$$1 \ge u_i \ge p_{i,i} \ge \tfrac{1}{2} \ (i \le k) \quad \text{and} \quad 0 \le u_i \le 1 - p_{i,i} \le \tfrac{1}{2} \ (i > k).$$

Now

$$\pi(\omega_j) = \sum_{i=1}^{N} \pi(\omega_i) p_{i,j} \text{ and } \sum_{i=1}^{N} p_{t-1}(\omega_i) p_{i,j} = p_t(\omega_j)$$

and so if $x = \theta_k$

$$\begin{aligned} h_t(x) &= \sum_{j=1}^{k} (p_t(\omega_j) - \pi(\omega_j)) = \sum_{j=1}^{k} \sum_{i=1}^{N} p_{i,j} (p_{t-1}(\omega_i) - \pi(\omega_i)) \\ &= \sum_{i=1}^{N} (p_{t-1}(\omega_i) - \pi(\omega_i)) u_i. \end{aligned} \tag{2.12}$$

Moreover, $0 \le u_i \le 1$ and

$$\sum_{i=1}^{N} \pi(\omega_i) u_i = \sum_{i=1}^{N} \sum_{j=1}^{k} \pi(\omega_i) p_{i,j} = \sum_{j=1}^{k} \sum_{i=1}^{N} \pi(\omega_i) p_{i,j} = \sum_{j=1}^{k} \pi(\omega_j) = x. \tag{2.13}$$

Now let

$$u_i' = \begin{cases} 2u_i - 1 & i \le k \\ 0 & i > k \end{cases} \quad \text{and } u_i'' = \begin{cases} 1 & i \le k \\ 2u_i & i > k \end{cases}$$

Then $0 \le u_i', u_i'' \le 1$ and $u_i' + u_i'' = 2u_i$. Let $x' = \sum_{i=1}^{N} \pi(\omega_i) u_i'$ and $x'' = \sum_{i=1}^{N} \pi(\omega_i) u_i''$. Then (2.13) implies $x' + x'' = 2x$ and so by (2.12)

$$\begin{aligned} h_t(x) &= \tfrac{1}{2} \sum_{i=1}^{N} (p_{t-1}(\omega_i) - \pi(\omega_i)) u_i' + \tfrac{1}{2} \sum_{i=1}^{N} (p_{t-1}(\omega_i) - \pi(\omega_i)) u_i'' \\ &\le \tfrac{1}{2} h_{t-1}(x') + \tfrac{1}{2} h_{t-1}(x''). \end{aligned}$$

Furthermore,

$$\begin{aligned} x - x' &= \sum_{i=1}^{N} \pi(\omega_i)(u_i - u_i') = \sum_{i=1}^{k} \pi(\omega_i)(1 - u_i) + \sum_{i=k+1}^{N} \pi(\omega_i) u_i \\ &= \sum_{i=1}^{k} \pi(\omega_i) \left(1 - \sum_{j=1}^{k} p_{i,j} \right) + \sum_{i=k+1}^{N} \pi(\omega_i) \sum_{j=1}^{k} p_{i,j} \\ &= \sum_{i=1}^{k} \sum_{j=k+1}^{N} \pi(\omega_i) p_{i,j} + \sum_{i=k+1}^{N} \pi(\omega_i) \sum_{j=1}^{k} p_{i,j} \\ &\ge 2\Phi \min\{x, 1 - x\}, \end{aligned}$$

using (2.3) and the definition of Φ. So, $x' \le x - 2\Phi\min\{x, 1-x\}$ and similarly $x'' \ge x + 2\Phi\min\{x, 1-x\}$. (2.11) now follows from the concavity of h_{t-1}.

Now consider an x, k such that $q_{k-1} < x < q_k < \frac{1}{2}$. Let $x = \alpha q_{k-1} + (1-\alpha)q_k$. Then

$$\begin{aligned} h_t(x) &= \alpha h_t(q_{k-1}) + (1-\alpha)h_t(q_k) \\ &\le \tfrac{1}{2}(\alpha(h_{t-1}(q_{k-1}(1-2\Phi)) + h_{t-1}(q_{k-1}(1+2\Phi))) \\ &\qquad\qquad + (1-\alpha)(h_{t-1}(q_k(1-2\Phi)) + h_{t-1}(q_k(1+2\Phi)))) \\ &= \tfrac{1}{2}(\alpha(h_{t-1}(q_{k-1}(1-2\Phi))) + (1-\alpha)(h_{t-1}(q_k(1-2\Phi)))) \\ &\qquad\qquad + \tfrac{1}{2}(\alpha(h_{t-1}(q_{k-1}(1+2\Phi))) + (1-\alpha)(h_{t-1}(q_k(1+2\Phi)))) \\ &\le \tfrac{1}{2}(h_{t-1}(x(1-2\Phi)) + h_{t-1}(x(1+2\Phi))) \end{aligned}$$

from the concavity of h_{t-1}. Thus (2.11) holds for such an x. A similar argument shows that (2.11) holds for an x, k such that $\frac{1}{2} < q_{k-1} < x < q_k$. So let ℓ be such that $q_{\ell-1} \le \frac{1}{2} < q_\ell$ and suppose $q_{\ell-1} < x < q_\ell$. For such x we can only prove

$$h_t(x) \le \tfrac{1}{2}(h_{t-1}(x - x\gamma_x\Phi) + h_{t-1}(x + x\gamma_x\Phi)) \tag{2.14}$$

where $\gamma_x \ge 2 - 4\pi_{\max}$.

$$\begin{aligned} h_t(x) &= \alpha h_t(q_{\ell-1}) + (1-\alpha)h_t(q_\ell) \\ &\le \tfrac{1}{2}(\alpha(h_{t-1}(q_{\ell-1}(1-2\Phi)) + h_{t-1}(q_{\ell-1}(1+2\Phi))) \\ &\qquad\qquad + (1-\alpha)((h_{t-1}(q_\ell - 2\Phi(1-q_\ell))) + h_{t-1}(q_\ell + 2\Phi(1-q_\ell)))) \\ &\le \tfrac{1}{2}(h_{t-1}(x - 2\Phi(x - (1-\alpha)(2q_\ell - 1))) + h_{t-1}(x + 2\Phi(x - (1-\alpha)(2q_\ell - 1))) \end{aligned}$$

Thus (2.14) holds with

$$\begin{aligned} \gamma_x &= 2 - \frac{(1-\alpha)(2q_\ell - 1)}{x} \\ &\ge 2 - \frac{2q_\ell - 1}{q_\ell} \\ &= 2 - \frac{4\varepsilon}{1+2\varepsilon} \end{aligned}$$

where $\varepsilon = q_\ell - \frac{1}{2} \le \pi_{\max}$ and (2.14) follows.

Combining (2.11) and (2.14) we get that for $0 \le x \le 1$

$$h_t(x) \le \tfrac{1}{2}(h_{t-1}(x - \gamma_x\Phi\min\{x, 1-x\}) + h_{t-1}(x + \gamma_x\Phi\min\{x, 1-x\})) \tag{2.15}$$

We now prove inductively that

$$h_t(x) \le C\min\left\{\sqrt{x}, \sqrt{1-x}\right\}\left(1 - \frac{\Phi^2}{2}\right)^t. \tag{2.16}$$

For $t = 0$ (2.16) follows trivially from $0 \leq h_0(x) \leq 1$. Let $t \geq 1$ and suppose for example $0 \leq x \leq \frac{1}{2}$. Then (2.15) implies

$$\begin{aligned} h_t(x) &\leq \tfrac{1}{2} C \left(1 - \tfrac{1}{2}\Phi^2\right)^{t-1} \left(\sqrt{x - \gamma_x \Phi x} + \sqrt{x + \gamma_x \Phi x}\right) \\ &= C \left(1 - \tfrac{1}{2}\Phi^2\right)^{t-1} \sqrt{x} \left(\sqrt{1 - \gamma_x \Phi} + \sqrt{1 + \gamma_x \Phi}\right) /2. \end{aligned}$$

The last factor can be estimated by

$$\begin{aligned} \tfrac{1}{2}(\sqrt{1 - \gamma_x \Phi} + \sqrt{1 + \gamma_x \Phi}) &= \tfrac{1}{2} \left(\sum_{r=0}^{\infty} (-1)^r \binom{\frac{1}{2}}{r} (\gamma_x \Phi)^r + \sum_{r=0}^{\infty} \binom{\frac{1}{2}}{r} (\gamma_x \Phi)^r \right) \\ &= \sum_{r=0}^{\infty} \binom{\frac{1}{2}}{2r} (\gamma_x \Phi)^r = 1 - \tfrac{1}{8}(\gamma_x \Phi)^2 - \tfrac{5}{128}(\gamma_x \Phi)^4 - \cdot \\ &\leq 1 - \tfrac{1}{2}\Phi^2. \end{aligned}$$

This completes the induction for $x \leq \frac{1}{2}$. For $x > \frac{1}{2}$ we put $x = 1 - y$ and define $\hat{h}_t(y) = h_t(1 - y)$. Then (2.15) gives

$$\hat{h}_t(y) \leq \tfrac{1}{2}(\hat{h}_{t-1}(y - \gamma_x \Phi y) + \hat{h}_{t-1}(y + \gamma_x \Phi y))$$

from which we obtain

$$\hat{h}_t(y) \leq C\sqrt{y} \left(1 - \tfrac{1}{2}\Phi^2\right)^t$$

as before. □

Suppose now that we define the following "distance" M between measures $\hat{\pi}$ and π on space Ω.

$$M(\hat{\pi}, \pi) = \max_{\emptyset \neq A \subseteq \Omega} \frac{|\hat{\pi}(A) - \pi(A)|}{\sqrt{\pi(A)}}. \tag{2.17}$$

Corollary 2.2.1 *Let a lazy ergodic Markov chain with steady state π be started with distribution π_0 and let π_t denote the distribution after t steps. If $\pi_{\max} \leq \frac{\Phi^2}{20}$ then*

$$M(\pi_t, \pi) \leq M(\pi_0, \pi) \left(1 - \tfrac{1}{2}\Phi^2\right)^t.$$

Proof Fix $S \subseteq \Omega$. It follows from (2.10) and (2.16) that

$$|\pi_t(S) - \pi(S)| \leq \sqrt{\pi(S)} M(\pi_0, \pi) \left(1 - \frac{\Phi^2}{2}\right)^t.$$

□

2.2.3 Path Congestion

Supose that for each pair $(x,y) \in \Omega \times \Omega$ we have a *canonical path* γ_{xy} from x to y in the digraph $D_{\mathcal{M}} = (\Omega, A)$ (defined in Section 1.3). Let

$$\bar{\rho} = \max_{e \in A} \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y)|\gamma_{xy}|$$

where if $e = (\sigma, \tau)$ then $Q(e) = \pi(\sigma)P(\sigma, \tau)$ and $|\gamma_{xy}|$ is the number of arcs in γ_{xy}.

Theorem 2.2.4

$$1 - \lambda_{\max} \geq \frac{1}{\bar{\rho}}.$$

Proof We use Lemma 2.2.2. Assume $\sum_i \pi_i y_i = 0$. Then

$$\begin{aligned} 2\sum_{i=1}^{N} \pi_i y_i^2 &= \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i\pi_j(y_i - y_j)^2 \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i\pi_j \left(\sum_{e \in \gamma_{ij}} (y_{e^+} - y_{e^-})\right)^2 \end{aligned}$$

where edge $e = (e^-, e^+)$

$$\leq \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_i\pi_j|\gamma_{ij}| \sum_{e \in \gamma_{ij}} (y_{e^+} - y_{e^-})^2$$

by Cauchy-Schwartz

$$\begin{aligned} &= \sum_{e \in A} (y_{e^+} - y_{e^-})^2 \sum_{\gamma_{xy} \ni e} \pi_i\pi_j|\gamma_{ij}| \\ &\leq \sum_{e \in A} (y_{e^+} - y_{e^-})^2 Q(e)\bar{\rho} \\ &= 2\bar{\rho}\mathcal{E}(y,y). \end{aligned}$$

□

This theorem often gives stronger bounds on the spectral gap than Theorem 2.2.1. We apply it now to our example of a random walk $\mathcal{W}_n$ on the cube.

Let $x = (x_0, x_1, \ldots, x_{n-1})$ and $y = (y_0, y_1, \ldots, y_{n-1})$ be arbitrary members of X_n. The canonical path γ_{xy} from x to y is composed of n edges, 0 to $n-1$, where edge i is simply

$$\big((y_0, \ldots, y_{i-1}, x_i, x_{i+1}, \ldots x_{n-1}),\ (y_0, \ldots, y_{i-1}, y_i, x_{i+1}, \ldots x_{n-1})\big),$$

i.e., we change the ith component from x_i to y_i. Note that some of the edges may be loops (if $x_i = y_i$). To compute $\bar{\varrho}$, fix attention on a particular (oriented) edge

$$t = (w, w') = \big((w_0, \ldots, w_i, \ldots w_{n-1}), (w_0, \ldots, w'_i, \ldots w_{n-1})\big),$$

and consider the number of canonical paths γ_{xy} that include t. The number of possible choices for x is 2^i, as the final $n-i$ positions are determined by $x_j = w_j$, for $j \geq i$; and by a similar argument the number of possible choices for y is 2^{n-i-1}. Thus the total number of canonical paths using a particular edge t is 2^{n-1}; furthermore, $Q(w, w') = \pi(w)P(w, w') \geq 2^{-n}(2n)^{-1}$, and the length of every canonical path is exactly n. Plugging all these bounds into the definition of $\bar{\rho}$ yields $\bar{\rho} \leq n^2$. Thus, by Theorem 2.2.4, the mixing time of $\mathcal{W}_n$ is $\tau(\varepsilon) \leq n^2(n \ln q + \ln \varepsilon^{-1})$.

2.2.4 Comparison Theorems

2.2.5 Decomposition Theorem

2.3 Coupling

A *coupling* $\mathcal{C}(\mathcal{M})$ for $\mathcal{M}$ is a stochastic process (X_t, Y_t) on $\Omega \times \Omega$ such that each of X_t, Y_t is marginally a copy of $\mathcal{M}$,

$$\begin{aligned} \mathbf{Pr}(X_t = \sigma_1 \mid X_{t-1} = \omega_1) &= P(\omega_1, \sigma_1), \\ \mathbf{Pr}(Y_t = \sigma_2 \mid Y_{t-1} = \omega_2) &= P(\omega_2, \sigma_2), \end{aligned} \qquad (\forall t > 0). \tag{2.18}$$

The following simple but powerful inequality then follows easily from these definitions.

Lemma 2.3.1 (Coupling Lemma) *Let X_t, Y_t be a coupling for $\mathcal{M}$ such that Y_0 has the stationary distribution π. Then, if X_t has distribution p_t,*

$$D_{\text{tv}}(p_t, \pi) \leq \mathbf{Pr}(X_t \neq Y_t). \tag{2.19}$$

Proof Suppose $A_t \subseteq \Omega$ maximizes in (1.3). Then, since Y_t has distribution π,

$$\begin{aligned} D_{\text{tv}}(p_t, \pi) &= \mathbf{Pr}(X_t \in A_t) - \mathbf{Pr}(Y_t \in A_t) \\ &\leq \mathbf{Pr}(X_t \in A_t, Y_t \notin A_t) \\ &\leq \mathbf{Pr}(X_t \neq Y_t). \end{aligned}$$

□

It is important to remember that the Markov chain Y_t is simply a proof construct, and X_t the chain we actually observe. We also require that $X_t = Y_t$ implies $X_{t+1} = Y_{t+1}$, since this makes the right side of (2.19) nonincreasing. Then the earliest epoch T at which $X_T = Y_T$ is called *coalescence*, making T a random variable. A *successful coupling* is such that $\lim_{t \longrightarrow \infty} \mathbf{Pr}(X_t \neq Y_t) = 0$. Clearly we are only interested in successful couplings.

As an example consider our random walk on the cube Q_n. We can define a coupling as follows: Given (X_t, Y_t) we

(a) Choose i uniformly at random from $[n]$.

(b) Put $X_{t+1,j} = X_{t,j}$ and $Y_{t+1,j} = Y_{t,j}$ for $j \neq i$.

(c) If $X_{t,i} = Y_{t,i}$ then

$$X_{t+1,i} = Y_{t+1,i} = \begin{cases} X_{t,i} & \text{prob } \frac{1}{2} \\ \\ 1 - X_{t,i} & \text{prob } \frac{1}{2} \end{cases}$$

(d) otherwise

$$(X_{t+1,i}, Y_{t+1,i}) = \begin{cases} (X_{t,i}, 1 - Y_{t,i}) \text{ prob } \frac{1}{2} \\ \\ (1 - X_{t,i}, Y_{t,i}) \text{ prob } \frac{1}{2} \end{cases}$$

It should hopefully be clear that this is a coupling i.e. the marginals are correct and $X_t = Y_t$ implies $X_{t+1} = Y_{t+1}$.

Now let $I_t = \{j : i$ is chosen in (a) of steps $1, 2, \ldots, t$. Then $I_t = [n]$ implies that $X_\tau = Y_\tau$ for $\tau \geq t$. So

$$\begin{aligned} \mathbf{Pr}(X_t \neq Y_t) &\leq \mathbf{Pr}(I_t \neq [n]) \\ &= \mathbf{Pr}(\bar{I}_t \neq \emptyset) \\ &\leq \mathbf{E}(|\bar{I}_t|) \\ &= n\left(1 - \tfrac{1}{n}\right)^t. \end{aligned}$$

So if $t = n(\log n + \log \epsilon^{-1})$ we have $d_{TV}(p_t, \pi) \leq \epsilon$.

A coupling is a *Markovian coupling* if the process $\mathcal{C}(\mathcal{M})$ is a Markov chain on $\Omega \times \Omega$. There always exists a *maximal coupling*, which gives equality in (2.19). This maximal coupling is in general non-Markovian, and is seemingly not constructible without knowing p_t $(t = 1, 2, \ldots)$. But coupling has little algorithmic value if we already know p_t. More generally, it seems difficult to prove mixing properties of non-Markovian couplings in our setting. Therefore we restrict attention to Markovian couplings, at the (probable) cost of sacrificing equality in (2.19).

Let $\mathcal{C}(\mathcal{M})$ be a Markovian coupling, with Q its transition matrix, i.e. the probability of a joint transition from (ω_1, ω_2) to (σ_1, σ_2) is $Q^{\omega_1 \omega_2}_{\sigma_1 \sigma_2}$. The precise conditions required of

Q are then

$$Q^{\omega\,\omega}_{\sigma_1\sigma_2} \neq 0 \qquad \text{implies } \sigma_1 = \sigma_2 \qquad\qquad (\forall \omega \in \Omega), \tag{2.20}$$

$$\sum_{\sigma_2\in\Omega} Q^{\omega_1\omega_2}_{\sigma_1\sigma_2} = P^{\omega_1}_{\sigma_1} \quad (\forall \omega_2 \in \Omega), \ \sum_{\sigma_1\in\Omega} Q^{\omega_1\omega_2}_{\sigma_1\sigma_2} = P^{\omega_2}_{\sigma_2} \quad (\forall \omega_1 \in \Omega). \tag{2.21}$$

Here (2.20) implies equality after coalescence, and (2.21) implies the marginals are copies of $\mathcal{M}$. Our goal is to design Q so that $\mathbf{Pr}(X_t \neq Y_t)$ quickly becomes small. We need only specify Q to satisfy (2.21) for $\omega_1 \neq \omega_2$. The other entries are completely determined by (2.20) and (2.21).

In general, to prove rapid mixing using coupling, it is usual to map $\mathcal{C}(\mathcal{M})$ to a process on $\mathbb{N}$ by defining a function $\psi : \Omega \times \Omega \longrightarrow \mathbb{N}$ such that $\psi(\omega_1, \omega_2) = 0$ implies $\omega_1 = \omega_2$. We call this a *proximity function*. Then $\mathbf{Pr}(X_t \neq Y_t) \leq \mathbf{E}(\psi(X_t, Y_t))$, by Markov's inequality, and we need only show that $\mathbf{E}(\psi(X_t, Y_t))$ converges quickly to zero.

2.4 Path coupling

A major difficulty with coupling is that we are obliged to to specify it, and show improvement in the proximity function, for every pair of states. The idea of *path coupling*, where applicable, can be a major saving in this respect. We describe the approach below.

As a simple example of this approach consider a Markov chain where $\Omega \subseteq S^m$ for some set S and positive integer m. Suppose also that if $\omega, \sigma \in \Omega$ and $h(\omega, \sigma) = d$ (Hamming distance) then there exists a sequence $\omega = x_0, x_1, \ldots, x_d = \sigma$ of members of Ω such that (i) $\{x_0, x_1, \ldots, x_d\} \subseteq \Omega$, (ii) $h(x_i, x_{i+1}) = 1$, $i = 0, 1, \ldots, d-1$ and (iii) $P(x_i, x_{i+1}) > 0$.

Now suppose we define a coupling of the chains (X_t, Y_t) *only* for the case where $h(X_t, Y_t) = 1$. Suppose then that

$$\mathbf{E}(h(X_{t+1}, Y_{t+1}) \mid h(X_t, Y_t) = 1) \leq \beta \tag{2.22}$$

for some $\beta < 1$. Then

$$\mathbf{E}(h(X_{t+1}, Y_{t+1})) \leq \beta h(X_t, Y_t), \tag{2.23}$$

in all cases. It then follows that

$$d_{TV}(p_t, \pi) \leq \mathbf{Pr}(X_t \neq Y_t) \leq n\beta^t.$$

Equation (2.23) is shown by choosing a sequence $X_t = Z_0, Z_1, \ldots, Z_d = Y_t, d = h(X_t, Y_t)$ $Z_0, Z_1, \ldots, Z_d$ satisfy (i),(i),(iii) above. Then we can couple Z_i and Z_{i+1}, $1 \leq i < d$ so that $X_{t+1} = Z'_0, Z'_1, \dots, Z'_d = Y_{t+1}$ and (i) $\mathbf{Pr}(Z'_i = \sigma \mid Z_i = \omega) = P(\omega, \sigma)$ and (ii) $\mathbf{E}(h(Z'_i, Z'_{i+1})) \leq \beta$. Therefore

$$\mathbf{E}(h(X_{t+1}, Y_{t+1})) \leq \sum_{i=1}^{d} \mathbf{E}(h(Z'_i, Z'_{i+1})) \leq \beta d$$

and (2.23) follows.

As an example, let $G = (V, E)$ be a graph with maximum degree Δ and let $k \geq 2\Delta + 1$ be an integer. Let Ω_k be the set of proper k-vertex colourings of G i.e. $\{c : V \to [k]\}$ such that $(v, w) \in E$ implies $c(v) \neq c(w)$. We describe a chain which provides a good sampler for the uniform distribution over Ω_k. We let $\Omega = V^k$ be all k-colourings, including improper ones and describe a chain on Ω for which only proper colourings have a positive steady state probability.

To describe a general step of the chain asume $X_t \in \Omega$. Then

Step 1 Choose w uniformly from V and x uniformly from $[k]$.

Step 2 Let $X_{t+1}(v) = X_t(v)$ for $v \in V \setminus \{w\}$.

Step 3 If no neighbour of w in G has colour x then put $X_{t+1}(w) = x$, otherwise put $X_{t+1}(w) = x$.

Note that $P(\omega, \sigma) = P(\sigma, \omega) = \frac{1}{nk}$ for two proper colourings which can be obtained from each other by a single move of the chain. It follows from (1.15) that the steady state is uniform over Ω_k.

We first describe a coupling which is extremely simple but needs $k > 3\Delta$ in order for (2.22) to be satisfied. Let $h(X_t, Y_t) = 1$ and let v_0 be the unique vertex of V such that $X_t(v) \neq Y_t(v)$. In our coupling we choose w, x as in Step 1 and try to colour w with x in both chains.

We claim that

$$\mathbf{E}(h(X_{t+1}, Y_{t+1}) \leq 1 - \frac{1}{n}\left(1 - \frac{\Delta}{k}\right) + \frac{\Delta}{n}\frac{2}{k} = 1 - \frac{k - 3\Delta}{kn}. \tag{2.24}$$

and so we can take $\beta \leq 1 - \frac{1}{kn}$ in (2.23) if $k > 3\Delta$.

The term $\frac{1}{n}\left(1 - \frac{\Delta}{k}\right)$ in (2.24) lower bounds the probability that $w = v_0$ and that x is not used in the neighbourhood of v_0. In which case we will have $X_{t+1} = Y_{t+1}$. Next let $c_X \neq c_Y$ be the colours of v_0 in X_t, Y_t respectively. The term $\frac{\Delta}{n}\frac{2}{k}$ in (2.24) is an upper bound for the probability that w is in the neighbourhood of v_0 and $x \in \{c_X, c_Y\}$ and in which case we might have $h(X_{t+1}, Y_{t+1}) = 2$. In all other cases we find that $h(X_{t+1}, Y_{t+1}) = h(X_t, Y_t) = 1$.

A better coupling gives the desired result. We proceed as above except for the case where w is a neighbour of v_0 and $x \in \{c_X, c_Y\}$. In this case with probability $\frac{1}{2}$ we try to colour w with c_X in X_t and colour w with c_Y in Y_t, and fail in both cases. With probability $\frac{1}{2}$ we try to colour w with c_Y in X_t and colour w with c_X in Y_t, in which case the hamming distance may increase by one. Thus for this coupling we have

$$\mathbf{E}(h(X_{t+1}, Y_{t+1}) \leq 1 - \frac{1}{n}\left(1 - \frac{\Delta}{k}\right) + \frac{1}{2}\frac{\Delta}{n}\frac{2}{k} = 1 - \frac{k - 2\Delta}{kn}$$

and we can take $\beta \leq 1 - \frac{1}{kn}$ in (2.23) if $k > 2\Delta$.

We now give a more general framework for the definition of path coupling. Recall that a *quasi-metric* satisfies the conditions for a metric except possibly the symmetry condition. Any metric is a quasi-metric, but a simple example of a quasi-metric which is not a metric is directed edge distance in a digraph.

Suppose we have a relation $S \subseteq \Omega \times \Omega$ such that S has transitive closure $\Omega \times \Omega$, and suppose that we have a proximity function defined for all pairs in S, i.e. $\psi : S \longrightarrow \mathbb{N}$. Then we may lift ψ to a quasi-metric $\phi(\omega, \omega')$ on Ω as follows. For each pair $(\omega, \omega') \in \Omega \times \Omega$, consider the set $\mathcal{P}(\omega, \omega')$ of all sequences

$$\omega = \omega_1, \omega_2, \ldots, \omega_{r-1}, \omega_r = \omega' \quad \text{with} \quad (\omega_i, \omega_{i+1}) \in S \quad (i = 1, \ldots, r-1). \tag{2.25}$$

Then we set

$$\phi(\omega, \omega') \;=\; \min_{\mathcal{P}(\omega,\omega')} \sum_{i=1}^{r-1} \psi(\omega_i, \omega_{i+1}). \tag{2.26}$$

It is easy to prove that ϕ is a quasi-metric. We call a sequence minimizing (2.26) *geodesic*. We now show that, without any real loss, we may define the (Markovian) coupling only on pairs in S. Such a coupling is a called a path coupling. We give a detailed development below. Clearly $S = \Omega \times \Omega$ is always a relation whose transitive closure is $\Omega \times \Omega$, but path coupling is only useful when we can define a suitable S which is "much smaller" than $\Omega \times \Omega$. A relation of particular interest is $\mathcal{R}_\sigma$ from Section 1.4, but this is not always the best choice.

As in Section 2.3, we use σ (or σ_i) to denote a state obtained by performing a single transition of the chain from the state ω (or ω_i). Let P_σ^ω denote the probability of a transition from state ω to state σ in the Markov chain, and let $Q_{\sigma\sigma'}^{\omega\omega'}$ denote the probability of a joint transition from (ω, ω') to (σ, σ'), where $(\omega, \omega') \in S$, as specified by the path coupling. Since this coupling has the correct marginals, we have

$$\sum_{\sigma' \in \Omega} Q_{\sigma\sigma'}^{\omega\omega'} = P_\sigma^\omega, \qquad \sum_{\sigma \in \Omega} Q_{\sigma\sigma'}^{\omega\omega'} = P_{\sigma'}^{\omega'} \qquad (\forall (\omega, \omega') \in S). \tag{2.27}$$

We extend this to all pairs $(\omega, \omega') \in \Omega \times \Omega$, as follows. For each pair, fix a sequence $(\omega_1, \omega_2, \ldots, \omega_r) \in \mathcal{P}(\omega, \omega')$. We do not assume the sequence is geodesic here, or indeed the existence of any proximity function, but this is our eventual purpose. The implied global coupling $\bar{Q}^{\omega_1\omega_r}_{\sigma_1\sigma_r}$ is then defined along this sequence by successively conditioning on the previous choice. Using (2.27), this can be written explicitly as

$$\bar{Q}^{\omega_1\omega_r}_{\sigma_1\sigma_r} = \sum_{\sigma_2\in\Omega}\sum_{\sigma_3\in\Omega}\cdots\sum_{\sigma_{r-1}\in\Omega} Q^{\omega_1\omega_2}_{\sigma_1\sigma_2}\frac{Q^{\omega_2\omega_3}_{\sigma_2\sigma_3}}{P^{\omega_2}_{\sigma_2}}\cdots\frac{Q^{\omega_{r-1}\omega_r}_{\sigma_{r-1}\sigma_r}}{P^{\omega_{r-1}}_{\sigma_{r-1}}}. \tag{2.28}$$

Summing (2.28) over σ_r or σ_1, and again applying (2.27), causes the right side to successively simplify, giving

$$\sum_{\sigma_r\in\Omega}\bar{Q}^{\omega_1\omega_r}_{\sigma_1\sigma_r} = P^{\omega_1}_{\sigma_1} \quad (\forall\omega_r\in\Omega), \qquad \sum_{\sigma_1\in\Omega}\bar{Q}^{\omega_1\omega_r}_{\sigma_1\sigma_r} = P^{\omega_r}_{\sigma_r} \quad (\forall\omega_1\in\Omega). \tag{2.29}$$

Hence the global coupling satisfies (2.21), as we would anticipate from the properties of conditional probabilities.

Now suppose the global coupling is determined by geodesic sequences. We bound the expected value of $\phi(\sigma_1,\sigma_r)$. This is

$$\begin{aligned}
\mathbf{E}(\phi(\sigma_1,\sigma_r)) &= \sum_{\sigma_1}\cdots\sum_{\sigma_r}\phi(\sigma_1,\sigma_r)\frac{Q^{\omega_1\omega_2}_{\sigma_1\sigma_2}Q^{\omega_2\omega_3}_{\sigma_2\sigma_3}\cdots Q^{\omega_{r-1}\omega_r}_{\sigma_{r-1}\sigma_r}}{P^{\omega_2}_{\sigma_2}\cdots P^{\omega_{r-1}}_{\sigma_{r-1}}} \\
&\le \sum_{\sigma_1}\cdots\sum_{\sigma_r}\sum_{i=1}^{r-1}\phi(\sigma_i,\sigma_{i+1})\frac{Q^{\omega_1\omega_2}_{\sigma_1\sigma_2}Q^{\omega_2\omega_3}_{\sigma_2\sigma_3}\cdots Q^{\omega_{r-1}\omega_r}_{\sigma_{r-1}\sigma_r}}{P^{\omega_2}_{\sigma_2}\cdots P^{\omega_{r-1}}_{\sigma_{r-1}}} \\
&= \sum_{i=1}^{r-1}\sum_{\sigma_1}\cdots\sum_{\sigma_r}\phi(\sigma_i,\sigma_{i+1})\frac{Q^{\omega_1\omega_2}_{\sigma_1\sigma_2}Q^{\omega_2\omega_3}_{\sigma_2\sigma_3}\cdots Q^{\omega_{r-1}\omega_r}_{\sigma_{r-1}\sigma_r}}{P^{\omega_2}_{\sigma_2}\cdots P^{\omega_{r-1}}_{\sigma_{r-1}}} \\
&= \sum_{i=1}^{r-1}\sum_{\sigma_i}\sum_{\sigma_{i+1}}\phi(\sigma_i,\sigma_{i+1})Q^{\omega_i\omega_{i+1}}_{\sigma_i\sigma_{i+1}},
\end{aligned} \tag{2.30}$$

where we have used the triangle inequality for a quasi-metric and the same observation as that leading from (2.28) to (2.29).

Suppose we can find $\beta \le 1$, such that, for all $(\omega,\omega')\in S$,

$$\mathbf{E}(\phi(\sigma,\sigma')) \;=\; \sum_{\sigma}\sum_{\sigma'}\phi(\sigma,\sigma')\,Q^{\omega\omega'}_{\sigma\sigma'} \;\le\; \beta\,\phi(\omega,\omega'). \tag{2.31}$$

Then, from (2.30), (2.31) and (2.26) we have

$$\mathbf{E}(\phi(\sigma_1,\sigma_r)) \;\le\; \sum_{i=1}^{r-1}\beta\,\phi(\omega_i,\omega_{i+1}) \;=\; \beta\sum_{i=1}^{r-1}\phi(\omega_i,\omega_{i+1}) \;=\; \beta\,\phi(\omega_1,\omega_r). \tag{2.32}$$

Thus we can show (2.31) for every pair, merely by showing that this holds for all pairs in S. To apply path coupling to a particular problem, we must find a relation S and proximity function ψ so that this is possible. In particular we need $\phi(\omega,\omega')$ for $(\omega,\omega')\in S$ to be easily deducible from ψ.

Suppose that Ω has *diameter* D, i.e. $\phi(\omega,\omega')\le D$ for all $\omega,\omega'\in\Omega$. Then, $\mathbf{Pr}(X_t\ne Y_t)\le\beta^t D$ and so if $\beta<1$ we have, since $\log\beta^{-1}\ge 1-\beta$,

$$D_{\rm tv}(p_t,\pi)\le\varepsilon \quad\text{for}\quad t\ge\log(D\varepsilon^{-1})/(1-\beta). \tag{2.33}$$

This bound is polynomial even when D is exponential in the problem size. It is also possible to prove a bound when $\beta=1$, provided we know the quasi-metric cannot "get stuck". Specifically, we need an $\alpha>0$ (inversely polynomial in the problem size) such that, in the above notation,

$$\mathbf{Pr}(\phi(\sigma,\sigma')\ne\phi(\omega,\omega'))\ge\alpha \qquad (\forall\omega,\omega'\in\Omega). \tag{2.34}$$

Observe that it is not sufficient simply to establish (2.34) for pairs in S. However, the structure of the path coupling can usually help in proving it. In this case, we can show that

$$D_{\rm tv}(p_t,\pi)\le\varepsilon \quad\text{for}\quad t\ge\lceil eD^2/\alpha\rceil\lceil\ln(\varepsilon^{-1})\rceil. \tag{2.35}$$

This is most easily shown using a martingale argument. Here we need D to be polynomial in the problem size.

Consider a sequence $(\omega_0,\omega_0'),(\omega_1,\omega_1')\dots,(\omega_t,\omega_t')$ and define the random time $T^{\omega,\omega'}=\min\{t:\ \phi(\omega_t,\omega_t')=0\}$, assuming that $\omega_0=\omega,\omega_0'=\omega'$. We prove that

$$\mathbf{E}(T^{\omega,\omega'})\le D^2/\alpha. \tag{2.36}$$

Let

$$Z(t)=\phi(\omega_t,\omega_t')^2-2D\phi(\omega_t,\omega_t')-\alpha t$$

and let

$$\delta(t)=\phi(\omega_{t+1},\omega_{t+1}')-\phi(\omega_t,\omega_t').$$

Then

$$\begin{aligned}&\mathbf{E}(Z(t+1)\mid Z(0),Z(1),\dots,Z(t))-Z(t)=\\ &\qquad 2(\phi(\omega_t,\omega_t')-D)\mathbf{E}(\delta(t)\mid\omega_t,\omega_t')+(\mathbf{E}(\delta(t)^2\mid\omega_t,\omega_t')-\alpha)\ge 0.\end{aligned}$$

Hence $Z(t)$ is a submartingale. The stopping time $T^{\omega,\omega'}$ has finite expectation and $|Z(t+1)-Z(t)|\le D^2$. We can therefore apply the Optional Stopping Theorem for submartingales to obtain

$$\mathbf{E}(Z(T^{\omega,\omega'}))\ge Z(0).$$

This implies

$$-\alpha\mathbf{E}(T^{\omega,\omega'})\ge\delta(0)^2-2D\delta(0)$$

and (2.36) follows.

So for any ω, ω'

$$\mathbf{Pr}(T^{\omega,\omega'} \geq eD^2/\alpha) \leq e^{-1}$$

and by considering k consecutive time intervals of length k we obtain

$$\mathbf{Pr}(T^{\omega,\omega'} \geq keD^2/\alpha) \leq e^{-k}$$

and (2.35) follows.

2.5 Hitting Time Lemmas

For a finite Markov chain $\mathcal{M}$ let $\mathbf{Pr}_i, \mathbf{E}_i$ denote probability and expectation, given that $X_0 = i$.

For a set $A \subseteq \Omega$ let

$$T_A = \min\{t \geq 0 : \; X_t \in A\}.$$

Then for $i \neq j$ the *hitting time*

$$H_{i,j} = \mathbf{E}_i(T_j)$$

is the expected number of steps needed to get from state i to state j.

The *commute time*

$$C_{i,j} = H_{i,j} + H_{j,i}.$$

Lemma 2.5.1 *Assume $X_0 = i$ and S is a stopping time. Let j be an arbitrary state. Then*

$$\mathbf{E}_i(\textit{number of visits to state } j \textit{ before time } S) = \pi_j \mathbf{E}_i(S).$$

Proof Consider the renewal process whose inter-renewal time is distributed as S. The reward-renewal theorem states that the asymptotic proportion of time spent in state j is given by

$$\mathbf{E}_i(\text{number of visits to } j \text{ before time } S)/\mathbf{E}_i(S).$$

This also equal to π_j, by the ergodic theorem. □

Lemma 2.5.2

$$\mathbf{E}_j(\textit{number of visits to } j \textit{ before } T_i) = \pi_j C_{i,j}.$$

Proof Let S be the time of the first return to i after the first visit to j. Apply Lemma 2.5.1. □

The *cover time* $C(\mathcal{M})$ of $\mathcal{M}$ is $\max_i C_i(\mathcal{M})$ where $C_i(\mathcal{M}) = \mathbf{E}_i(\max_j T_j)$ is the expected time to visit all states starting at i.

Let $\mathcal{M}_G$ denote a random walk on the connected graph $G = (V, E)$. Here $|V| = n$ and $|E| = m$.

Lemma 2.5.3 *For Markov chain $\mathcal{M}_G$ and $e = \{u, v\} \in E$, $C_{u,v} \leq 2m$.*

Proof The random walk on G induces a Markov chain on $A = \{(x, y) : \{x, y\}\}$ the set of oriented edges obtainable by replacing each edge $\{x, y\} \in E$ by a pair of oppositely oriented edges. It is can be easily checked that the all 1's vector satisfies (1.13) and hence the steady state of the induced walk is uniform. It follows from (1.14) the expected time between traversals of (v, u) is $\frac{1}{2m}$. So conditional on entering u from v the expected time to visit v and subsequently visit u is at most $\frac{1}{2m}$. Conditioning on initially traversing (v, u) is irrelevant to the time to subsequently visit v and then u and the lemma follows. □

We can use this to obtain a bound on the cover time of $\mathcal{M}_G$.

Lemma 2.5.4

$$c(\mathcal{M}_G) \leq 2m(n-1).$$

Proof Let T be any spanning tree of G and let $v_0, v_1, \ldots, v_{2n-2} = v_0$ be a traversal of G which crosses each edge of T in each direction. Now consider the expected time for the random walk, started at v_0, to make journeys from v_0 to v_1, then from v_1 onto v_2 and so on until $v_0, v_1, \ldots, v_{2n-2}$ have been visited. This journey visits every vertex of G and so its expected length is an upper bound on the cover time, from v_0. Thus

$$C_{v_0}(\mathcal{M}_G) \leq \sum_{i=0}^{2n-3} H_{v_i, v_{i+1}} = \sum_{\{u,v\} \in T} C_{u,v}.$$

The result now follows from Lemma 2.5.3. □

2.6 Optimal Stopping Rules

Lovász and Winkler, see for example [?] have been studying *optimal stopping rules.* We need a little of that theory here. For us a stopping rule is a function $\rho : \Omega^* :\rightarrow [0, 1]$ where $\Omega^* = \{(X_0, X_1, \ldots, X_t) : t \geq 0\}$ is the set of possible sequences of states generated by our Markov chain. $\rho(X_0, X_1, \ldots, X_t)$ is the probability that we stop the chain at time t. If X_0 is chosen with probability distribution σ and τ is the distribution of the state where we stop then we say that ρ is a stopping rule from σ to τ and write $\rho \in SR(\sigma, \tau)$. We are naturally mainly interested in the case where $\tau = \pi$.

For a stopping rule ρ let T_ρ be the random number of steps taken until we stop. Let

$$H(\sigma,\tau) = \inf\{\mathbf{E}(T_\rho):\ \rho \in SR(\sigma,\tau)\}$$

denote the minimum expected number of steps in a stopping rule from σ to τ. If σ is concentrated on a single state s then we write $H(s,\tau)$.

For a stopping rule ρ and $j \in \Omega$ let $x_j = x_j(\rho)$ be the expected number of exits from i before stopping i.e. the expected number of times that the chain leaves i.

Lemma 2.6.1 *If $\rho \in SR(\sigma,\tau)$ then*

$$x_j + \tau_j = \sum_{i\in\Omega} x_i P(i,j) + \sigma_j.$$

Proof Let $T = T_\rho$ and consider the identity

$$\sum_{t=0}^{T-1} 1_{X_t=j} + 1_{T<\infty, X_T=j} = 1_{X_0=j} + \sum_{t=1}^{T} 1_{X_t=j}$$

where both sides count the number of times that $X_t = j$.

Taking expectations we have

$$\begin{aligned} x_j + \tau_j = \sigma_j + \sum_{t=1}^{T}\mathbf{Pr}(X_t=j) &= \sigma_j + \sum_{t=0}^{T-1}\sum_{i\in\Omega}\mathbf{Pr}(X_t=i)P(i,j) \\ &= \sum_{i\in\Omega}\sum_{t=0}^{T-1} 1_{X_t=i}P(i,j) = \sum_{i\in\Omega} x_i P(i,j). \end{aligned}$$

□

Corollary 2.6.1 *Let $\rho_1, \rho_2 \in SR(\sigma,\tau)$. Then for all $i \in \Omega$*

$$x_i(\rho_1) - x_i(\rho_2) = D\pi_i$$

where $D = \mathbf{E}(T_{\rho_1} - T_{\rho_2})$.

Proof It follows from Lemma 2.6.1 that $\xi = x(\rho_1) - x(\rho_2)$ satisfies

$$\xi_j = \sum_{i\in\Omega} \xi_i P(i,j).$$

Therefore $\xi = A\pi$ for some $A \geq 0$. Now for $k = 1,2$

$$T_{\rho_k} = \sum_{j\in\Omega} x_j(\rho_k)$$

and the result follows. □

A state j is a halting state for rule ρ if $x_j(\rho) = 0$. This implies that if the chain ever enters state j then it stops. Using Corollary 2.6.1 we can prove the following remarkable theorem:

Theorem 2.6.1 *A stopping rule $\rho \in SR(\sigma, \tau)$ has a minimum mean expected stopping time iff there is a halting state.*

Proof If there exists j such that $x_j = 0$ then Corollary 2.6.1 implies that for $\rho' \in SR(\sigma, \tau)$

$$\mathbf{E}(T_{\rho'} - T_\rho) = \frac{x_j(\rho')}{\pi_j} \geq 0$$

implying that $\mathbf{E}(T_\rho)$ is minimal. It only remains to show that there exists a stopping rule in $SR(\sigma, \tau)$ which has at least one halting state.

The rule we define has a particular format. We define a nested sequence of sets $S_i = \{v_i, v_{i+1}, \ldots, v_n\}$ where $\Omega = \{v_1, v_2, \ldots, v_n\}$. For each i we will define $q^{(i)}$ by

$$q_j^{(i)} = \mathbf{Pr}(v_j \text{ is the first vertex of } S_i \text{ visited}).$$

In particular $q^{(1)} = \sigma$. We choose $S_1, S_2, \ldots, S_n$ so that we can write

$$\tau = \alpha_1 q^{(1)} + \alpha_2 q^{(2)} + \cdots + \alpha_n q^{(n)} \tag{2.37}$$

where $\alpha \geq 0$ and $\alpha_1 + \alpha_2 + \cdots + \alpha_n = 1$. Our stopping rule ρ is then:

(i) Choose i with probability α_i.

(ii) Choose X_0 with probability σ and then run the chain until S_i is reached and then stop.

It should be clear that $\rho \in SR(\sigma, \tau)$. If $S_1, S_2, \ldots, S_n$ can be constructed so that (2.37) holds then we are done: v_n is a halting state.

Assume inductively that we have found $S_1, S_2, \ldots, S_i$ and $\alpha_1, \alpha_2, \ldots, \alpha_{i-1} \geq 0$ such that

$$\tau^{(i-1)} = \tau - (\alpha_1 q^{(1)} + \alpha_2 q^{(2)} + \cdots + \alpha_{i-1} q^{(i-1)}) \geq 0 \tag{2.38}$$

and

$$\alpha_1 + \alpha_2 + \cdots + \alpha_{i-1} \leq 1.$$

Putting $S_1 = \Omega$ does this for $i = 1$ and then for general i let

$$\alpha_i = \min_{j \in S_i} \frac{\tau_j^{(i)}}{q_j^{(i)}}$$

and let v_i be a state of S_i which achieves the minimum. Clearly $\alpha_i \geq 0$ and

$$\tau^{(i)} = \tau^{(i-1)} - \alpha_i q^{(i)} \geq 0 \tag{2.39}$$

from the definition of α_i.

Furthermore

$$\sum_{j=1}^{i} \alpha_j = \sum_{j=1}^{i} \alpha_j \sum_{k=1}^{n} q_k^{(j)} \leq \sum_{k=1}^{n} \tau_k = 1 \tag{2.40}$$

completing the induction.

Finally note that when $i = n$ the construction yields equality in (2.39) and then (2.37) holds and we obtain equality in (2.40). □

We now relate optimal stopping rules and mixing time. Let

$$T_{\text{mix}} = \max_{s \in \Omega} H(s, \pi).$$

Theorem 2.6.2

$$\tau(\epsilon) \leq 8 T_{\text{mix}} \log_2(1/\epsilon).$$

Proof Let $s \in \Omega$ and let ρ be an optimal stopping rule from s to π. Consider a modification: Follow ρ until it stops after $T = T_\rho$ steps and then generate $\xi \in \{0, 1, \ldots, t-1\}$ uniformly and independently of the previous walk, and then walk ξ more steps. Let the walk be $v_1, v_2, \ldots, v_{T+\xi}$. Then let $\eta = T + \xi \pmod t$ and note that η is uniformly distributed over $\{0, 1, \ldots, t-1\}$. Then for $i \in \Omega$

$$\mathbf{Pr}(v_\eta = i) \geq \mathbf{Pr}(v_{T+\xi} = i) - \mathbf{Pr}(v_{T+\xi} = i, v_\eta \neq i) \geq \pi_i - \mathbf{Pr}(v_{T+\xi} = i, T + \xi \geq t)$$

since $v_{T+\xi}$ is in the stationary distribution and $T + \xi < t$ implies $\eta = T + \xi$.

Hence, for every $A \subseteq \Omega$,

$$\pi(A) - \mathbf{Pr}(v_\eta \in A) \leq \mathbf{Pr}(v_{T+\xi} \in A, T + \xi \geq t) \leq \mathbf{Pr}(T + \xi \geq t).$$

Now for any fixed value of T, $\mathbf{Pr}(T + \xi \geq t) \leq \frac{T}{t}$ and so

$$\mathbf{Pr}(T + \xi \geq t) \leq \frac{\mathbf{E}(T)}{t} = \frac{H(s, \pi)}{t}$$

and

$$\pi(A) - \mathbf{Pr}(v_\eta \in A) \leq \frac{H(s, \pi)}{t}.$$

It follows from Lemma 1.3.1(d) that

$$d(t) \leq \frac{T_{\text{mix}}}{t}$$

and so

$$d(4T_{\rm mix}) \le \frac{1}{4}.$$

Applying Lemma 1.3.1(b) we see that

$$d(8T_{\rm mix} \log_2 \epsilon^{-1}) \le \epsilon.$$

□

We can now prove a refinement of the usual conductance bound on the mixing time (Corollary 2.2.1) due to Kannan and Lovász [?]. Thus for $0 \le x \le \frac{1}{2}$ let

$$\Phi(x) = \min_{\substack{S \subseteq \Omega \\ \pi(S) \le x}} \frac{Q(S,\bar S)}{\pi(S)\pi(\bar S)}$$

and let $\Phi(x) = \Phi(\frac{1}{2})$ for $\frac{1}{2} < x \le 1$. Note that $\Phi(x) \le 2$.

Theorem 2.6.3 *If* $0 \le \xi \le 1$ *then*

$$T_{\rm mix} \le \frac{30}{\xi^2} + 30 \int_{x=\pi_\xi}^{1} \frac{dx}{x\Phi(x)^2}$$

where $\pi_\xi = \inf\{y:\ \exists S$ *such that* $\pi(S) \le y$ *and* $\Phi(S) < \xi\}$.

Proof Let $s \in \Omega$ and ρ be an optimal stopping rule from s to π. let $y_i = x_i/\pi_i$, $i = 1,2,\ldots,n$ be the *scaled exit frequencies* of ρ. Now order the states so that $y_1 \le y_2 \le \cdots \le y_n$. We first claim that with this ordering

$$y_1 = 0 \text{ and } n = s. \tag{2.41}$$

The first assertion comes from Theorem 2.6.1. For the second we use Lemma 2.6.1 and write, for $j \in \Omega$,

$$\sum_{i \in \Omega} \pi_i P(i,j) y_i - \pi_j y_j = \pi_j - 1_{j=s}.$$

Putting $j = n$ we obtain

$$\pi_n - 1_{n=s} \le \sum_{i \in \Omega} \pi_i P(i,n) y_n - \pi_n y_n = 0$$

and (2.41) follows.

Now fix $1 \le k < m \le n$ and let $A = \{1,2,\ldots,k\}$, $B = \{k+1,k+2,\ldots,m-1\}$ and $C = \{m,m+1,\ldots,n\}$. We show next that

$$y_m - y_k \le \frac{\pi(A)}{Q(C,A)}. \tag{2.42}$$

We start with the identity

$$\sum_{i=1}^{k}\sum_{j=k+1}^{n} y_j Q(j,i) - \sum_{i=1}^{k}\sum_{j=k+1}^{n} y_i Q(i,j) = \pi(A). \tag{2.43}$$

The left hand side counts the expected number of steps from $V \setminus A$ to A less the expected number of steps from A to $V \setminus A$, when following an optimal rule. Since we do not start in A ($s = n$) and stop in A with probability $\pi(A)$, (2.43) follows.

Now we estimate the left hand side of (2.43) as follows:

$$\begin{aligned}\sum_{i=1}^{k}\sum_{j=k+1}^{n} y_j Q(j,i) &\geq \sum_{i=1}^{k}\sum_{j=m+1}^{n} y_m Q(j,i) + \sum_{i=1}^{k}\sum_{j=k+1}^{m-1} y_k Q(j,i) \\ &= y_m Q(C,A) + y_k Q(B,A)\end{aligned}$$

and

$$\sum_{i=1}^{k}\sum_{j=k+1}^{n} y_i Q(i,j) \leq \sum_{i=1}^{k}\sum_{j=k+1}^{n} y_k Q(i,j) = y_k Q(A, B\cup C) = y_k Q(B\cup C, A).$$

Substituting into (2.43) we get

$$y_m Q(C,A) + y_k Q(B,A) - y_k Q(B\cup C, A) = (y_m - y_k) Q(C,A) \leq \pi(A)$$

which proves (2.42).

We now observe that since $y_1 = 0$,

$$H(s,\pi) = \sum_{i=1}^{n} \pi_i y_i = \sum_{j=1}^{n-1} (y_{j+1} - y_j)\pi_{>j} \tag{2.44}$$

where $\pi_{>j} = \sum_{r=j+1}^{n} \pi_j$.

We now define a sequence $1 = m_0 < m_1 < \cdots m_k < m_{k+1}$ so that if $T_i = \{1, 2, \ldots, m_i\}$, $\bar{T}_i = \Omega \setminus T_i$ and $a_i = \pi(T_i)$ then

$$a_{i+1} - \pi_{m_{i+1}} < a_i \left(1 + \frac{\Phi(a_i)}{4}\right) \leq a_{i+1} \tag{2.45}$$

and

$$a_k \leq \frac{1}{2} < a_{k+1}. \tag{2.46}$$

This definition can be justified as follows: Given m_i with $a_i \leq \frac{1}{2}$ we let m_{i+1} be the first integer such that (2.45) holds. Since $a_n = 1$ and $a_i \left(1 + \frac{\Phi(a_i)}{4}\right) \leq \frac{3}{2} a_i$, such an m_{i+1} exists. k exists for the same reason.

We bound a portion of the sum in the right hand side of (2.44) by

$$\sum_{j=m_i}^{m_{i+1}-1} (y_{j+1} - y_j)\pi_{>j} \le (1-a_i)(y_{m_{i+1}} - y_{m_i}) \le \frac{a_i(1-a_i)}{Q(\bar{T}_{i+1} \cup \{m_{i+1}\}, T_i)} \tag{2.47}$$

where the second inequality follows from (2.43). Now,

$$\begin{aligned} Q(\bar{T}_{i+1} \cup \{m_{i+1}\}, T_i) = Q(T_i, \bar{T}_i) - Q(\bar{T}_i \setminus (\bar{T}_{i+1} \cup \{m_{i+1}\}, T_i)) \ge \\ Q(T_i, \bar{T}_i) - \pi(\bar{T}_i \setminus (\bar{T}_{i+1} \cup \{m_{i+1}\})) \ge \\ \Phi(a_i)a_i(1-a_i) - a_{i+1} + \pi_{m_{i+1}} + a_i > \Phi(a_i)a_i(1-a_i)/2. \end{aligned}$$

Hence we obtain from (2.47) that

$$\sum_{j=m_i}^{m_{i+1}-1} (y_{j+1} - y_j)\pi_{>j} \le \frac{2}{\Phi(a_i)}. \tag{2.48}$$

Now define i_0 by $\Phi(a_i) \ge \xi$ iff $i \le i_0$. It follows from (2.45) that

$$i_0 \le \frac{\ln 2}{\ln\left(1 + \frac{\xi}{4}\right)} \le \frac{5}{\xi}.$$

So from (2.48) we see that

$$\sum_{j=1}^{m_{i_0+1}-1} (y_{j+1} - y_j)\pi_{>j} \le \sum_{i=1}^{i_0} \frac{2}{\Phi(a_i)} \le \frac{10}{\xi^2}. \tag{2.49}$$

In general we have

$$\begin{aligned} \int_{a_i}^{a_{i+1}} \frac{dx}{x\Phi(x)^2} \ge \frac{1}{\Phi(a_i)^2} \int_{a_i}^{a_{i+1}} \frac{dx}{x} = \frac{1}{\Phi(a_i)^2} \ln(a_{i+1}/a_i) \\ \ge \frac{1}{\Phi(a_i)^2} \ln\left(1 + \frac{\Phi(a_i)}{4}\right) \ge \frac{1}{5\Phi(a_i)} \end{aligned} \tag{2.50}$$

since $\Phi(a_i) \le 2$.

So from (2.48), (2.49) and (2.50) we have

$$\sum_{j=1}^{m_{k+1}-1} (y_{j+1} - y_j)\pi_{>j} \le \frac{10}{\xi^2} + 10 \int_{\pi_\xi}^{1} \frac{dx}{x\Phi(x)^2}. \tag{2.51}$$

The estimate for the other half of the sum on the right hand sise of (2.44) is similar. We define a sequence $n_0 = n > n_1 > \cdots > n_r$ and sets $S_i = \{n_i, n_i + 1, \ldots, n\}$, $\bar{S}_i = \Omega \setminus S_i$ and $b_i = \pi(S_i)$ for $i = 1, 2, \ldots, r+1$ such that

$$b_{i+1} - \pi_{n_{i+1}} < b_i\left(1 + \frac{\Phi(b_i)}{4}\right) \le b_{i+1}$$

and

$$b_r \leq \frac{1}{2} < b_{r+1}.$$

As beofe we consider the partial sum

$$\sum_{j=n_{i+1}}^{n_i-1} (y_{j+1} - y_j)\pi_{>j} \leq (b_{i+1} - \pi_{n_{i+1}})(y_{n_i} - y_{n_{i+1}}) \leq \frac{(b_{i+1} - \pi_{n_{i+1}})(1 - b_{i+1} + \pi_{n_{i+1}})}{Q(S_i, \bar{S}_{i+1} \cup \{n_{i+1}\})}$$

where the second inequality follows from (2.42).

Now

$$\begin{aligned} Q(S_i, \bar{S}_{i+1} \cup \{n_{i+1}\}) &= Q(\bar{S}_{i+1} \cup \{n_{i+1}\}, S_i) = \\ Q(\bar{S}_i, S_i) - Q(\bar{S}_i \setminus (\bar{S}_{i+1} \cup \{n_{i+1}\}), S_i) &\geq Q(\bar{S}_i, S_i) - \pi(\bar{S}_{i+1} \setminus \bar{S}_i) + \pi_{n_{i+1}} \geq \\ \Phi(b_i)b_i(1 - b_i) - b_{i+1} + \pi_{n_{i+1}} + b_i &> \Phi(b_i)b_i(1 - b_i)/2. \end{aligned}$$

Hence

$$\sum_{j=n_{i+1}}^{n_i-1} (y_{j+1} - y_j)\pi_{>j} \leq \frac{2(b_{i+1} - \pi_{n_{i+1}})}{b_i} \frac{1}{\Phi(b_i)} \leq \frac{4}{\Phi(b_i)}$$

since $b_{i+1} - \pi_{n_{i+1}} \leq b_i \left(1 + \frac{\Phi(b_i)}{4}\right) \leq 2b_i$. So as before we get

$$\sum_{j=n_{r+1}}^{n-1} (y_{j+1} - y_j)\pi_{>j} \leq \frac{20}{\xi^2} + 20 \int_{\pi_\xi}^{1} \frac{dx}{x\Phi(x)^2}$$

and combined with (2.49) and (2.44) we have the theorem. □

Of paticular interest to us is the case where for some $A = A(n) < B = B(n)$, $\Phi(x)$ satisfies

$$\Phi(x) \geq \min\left\{A \log \frac{1}{x(1-x)}, B\right\} \tag{2.52}$$

for $x \leq 1/2$.

Theorem 2.6.4 *If (2.52) holds then the mixing time*

$$\tau(\epsilon) \leq cA^{-2}$$

for some absolute constant $c > 0$.

Proof It follows from (2.52) that for $\xi \leq B$ we have $\pi_\xi \geq e^{-\xi/A}$. Define x_0 by $A \log \frac{1}{x_0(1-x_0)} = B$. Then for $\xi \leq B$ we have by Theorems 2.6.2 and 2.6.3 that

$$\begin{aligned}
\tau(\epsilon) &= O\left(\frac{1}{\xi^2} + \frac{1}{B^2}\int_{e^{-\xi/A}}^{x_0} \frac{dx}{x} + \frac{1}{A^2}\int_{x_0}^{1/2} \frac{dx}{x(\log x)^2} + \frac{1}{A^2(\log 4)^2}\int_{1/2}^{1} \frac{dx}{x}\right) \\
&= O\left(\frac{1}{\xi^2} + \frac{1}{B^2}\left(\log x_0 + \frac{\xi}{A}\right) + \frac{1}{A^2}\left(\frac{1}{\log x_0} + \frac{1}{\log 2}\right) + \frac{1}{A^2}\right) \\
&= O(A^{-2})
\end{aligned}$$

where we use $\log x_0 = \Theta(B/A)$ and take $\xi = (AB^2/2)^{1/3}$ and absorb terms of order $(AB)^{-1}$ or B^{-2}. $\square$

2.7 Coupling from the Past