Example 3.1 *Simple cyclic random walk.* Consider the regular, reversible Markov chain described by the graph in Figure 11.

Clearly the stationary distribution is \( \pi = \left[ \frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n} \right] \).

The conductance \( \Phi_A = F_A/C_A \) of a cut \( (A, \bar{A}) \) is minimised by choosing \( A \) to consist of any \( n/2 \) consecutive nodes on the cycle, e.g. \( A = \{1, 2, \ldots, n/2\} \). Then

\[
\Phi = \Phi_A = \frac{F_A}{C_A} = \frac{\sum_{i \in A} \pi_i p_{ij}}{\sum_{i \in A} \pi_i} = \frac{2 \cdot \frac{1}{n} \cdot \frac{1}{4}}{\frac{n}{2} \cdot \frac{1}{n}} = \frac{1/2n}{1/2} = \frac{1}{n}.
\]

Thus, by Theorem 3.6, the second eigenvalue satisfies:

\[
1 - \frac{2}{n} \leq \lambda_2 \leq 1 - \frac{1}{2n^2},
\]

by Corollary 3.7, the convergence rate satisfies

\[
\left(1 - \frac{2}{n}\right)^t \leq \Delta(t) \leq n \cdot \left(1 - \frac{1}{2n^2}\right)^t,
\]

and by Corollary 3.8, the mixing time satisfies:

\[
\frac{1 - 2/n}{2/n} \ln \frac{1}{\varepsilon} \leq \tau(\varepsilon) \leq 2n^2 \left(\ln \frac{1}{\varepsilon} + \ln n\right)
\]

\[
\Leftrightarrow \left(\frac{n}{2} - 1\right) \cdot \ln \frac{1}{\varepsilon} \leq \tau(\varepsilon) \leq 2n^2 \left(\ln n + \ln \frac{1}{\varepsilon}\right).
\]

Let us now return to the proof of Theorem 3.6, establishing the connection between the second-largest eigenvalue and the conductance of a Markov chain. Recall the statement of the Theorem:

**Theorem 3.6** Let \( \mathcal{M} \) be a finite, regular, reversible Markov chain and \( \lambda_2 \) the second-largest eigenvalue of its transition probability matrix. Then:
3. Estimating the Convergence Rate of a Markov Chain

(i) \( \lambda_2 \leq 1 - \frac{\Phi^2}{2} \).

(ii) \( \lambda_2 \geq 1 - 2\Phi \).

Proof. (i) The approach here is to bound \( \Phi \) in terms of the eigenvalue gap of \( \mathcal{M} \), i.e. to show that \( \Phi^2/2 \leq 1 - \lambda_2 \), from which the claimed result follows.

Thus, consider the eigenvalue \( \lambda = \lambda_2 \). (The following proof does not in fact depend on this particular choice of eigenvalue \( \lambda \neq 1 \), but since we are proving an upper bound of the form \( \Phi^2/2 \leq 1 - \lambda \), all other eigenvalues yield weaker bounds than \( \lambda_2 \).)

Let \( e \) be a left eigenvector \( e \neq 0 \) such that \( eP = \lambda e \). Since \( e \) is orthogonal to \( \pi \in [0, 1]^n \), \( e \) must contain both positive and negative components; in fact \( \sum_i e_i = 0 \) as can be seen:

\[
eP = \lambda e \iff \sum_i e_ip_{ij} = \lambda e_j \quad \forall j
\]

\[
\Rightarrow \sum_j \sum_i e_ip_{ij} = \sum_i e_i \sum_j p_{ij} = \lambda \sum_j e_j
\]

\[
\lambda \neq 1 \Rightarrow \sum_i e_i = 0.
\]

Define \( A = \{ i \mid e_i > 0 \} \). Assume, without loss of generality, that \( \pi(A) \leq 1/2 \). (Otherwise we may replace \( e \) by \( -e \) in the following proof.)

Define further a “\( \pi \)-normalised” version of \( e \mid A \):

\[
u_i = \begin{cases} e_i/\pi_i, & \text{if } i \in A \\ 0, & \text{if } i \notin A \end{cases}
\]

Without loss of generality we may again assume that the states are indexed so that \( u_1 \geq u_2 \geq \ldots \geq u_r > u_{r+1} = \ldots = u_n = 0 \), where \( r = |A| \).

In the remainder of the proof, the following quantity will be important:

\[
D = \frac{\sum_{i<j} w_{ij}(u_i^2 - u_j^2)}{\sum_i \pi_i u_i^2}.
\]

We shall prove the following claims:

(a) \( \Phi \leq D \),
(b) \( D^2/2 \leq 1 - \lambda \),

which suffice to establish our result.

**Proof of (a):** Denote \( A_k = \{1, \ldots, k\} \), for \( k = 1, \ldots, r \). The numerator in the definition of \( D \) may be expressed in terms of the ergodic flows out of the \( A_k \) as follows:

\[
\sum_{i < j} w_{ij}(u_i^2 - u_j^2) = \sum_{i < j} w_{ij} \sum_{i \leq k < j} (u_k^2 - u_{k+1}^2)
\]

\[
= \sum_{k=1}^r (u_k^2 - u_{k+1}^2) \sum_{i \in A_k \atop j \not\in A_k} w_{ij}
\]

\[
= \sum_{k=1}^r (u_k^2 - u_{k+1}^2) F_{A_k}.
\]

Now the capacities of the \( A_k \) satisfy \( \pi(A_k) \leq \pi(A) \leq 1/2 \), so by definition \( \Phi_{A_k} \geq \Phi \Rightarrow F_{A_k} \geq \Phi \cdot \pi(A_k) \). Thus,

\[
\sum_{i < j} w_{ij}(u_i^2 - u_j^2) = \sum_{k=1}^r (u_k^2 - u_{k+1}^2) F_{A_k}
\]

\[
\geq \Phi \cdot \sum_{k=1}^r (u_k^2 - u_{k+1}^2) \pi(A_k)
\]

\[
= \Phi \cdot \sum_{i=1}^r \sum_{k=i}^r (u_k^2 - u_{k+1}^2) \pi_i
\]

\[
= \Phi \cdot \sum_{i=1}^r \pi_i \sum_{k=i}^r (u_k^2 - u_{k+1}^2)
\]

\[
= \Phi \cdot \sum_{i \in A} \pi_i u_i^2.
\]

Hence,

\[
\Phi \leq \frac{\sum_{i < j} w_{ij}(u_i^2 - u_j^2)}{\sum_{i} \pi_i u_i^2} = D.
\]

**Proof of (b):** We introduce one more auxiliary expression:

\[
E = \frac{\sum_{i < j} w_{ij}(u_i - u_j)^2}{\sum_{i} \pi_i u_i^2},
\]
and establish that: (b’) $D^2 \leq 2E$, (b’’) $E \leq 1 - \lambda$. This will conclude the proof of Theorem 3.6 (i).

Proof of (b’): Observe first that

$$\sum_{i < j} w_{ij}(u_i + u_j)^2 \leq 2 \sum_{i < j} w_{ij}(u_i^2 + u_j^2) \leq 2 \sum_{i \in A} \pi_i u_i^2.$$ 

Then, by the Cauchy-Schwartz inequality:

$$D^2 = \left( \frac{\sum_{i < j} w_{ij}(u_i^2 - u_j^2)}{\sum_i \pi_i u_i^2} \right)^2 \leq \left( \frac{\sum_{i < j} w_{ij}(u_i + u_j)^2}{\sum_i \pi_i u_i^2} \right) \leq 2E.$$ 

Proof of (b’’): Denote $Q = I - P$. Then $eQ = (1 - \lambda)e$ and thus

$$eQu^T = (1 - \lambda)e u^T = (1 - \lambda) \sum_{i = 1}^r \pi_i u_i^2.$$ 

On the other hand, writing $eQu^T$ out explicitly:

$$eQu^T = \sum_{i = 1}^n \sum_{j = 1}^r q_{ij} e_i u_j$$

$$\geq \sum_{i = 1}^r \sum_{j = 1}^r q_{ij} e_i u_j$$

$$= - \sum_{i \in A} \sum_{j \notin A} w_{ij} u_i u_j + \sum_{i \in A} \sum_{j \notin A} w_{ij} u_i^2$$

$$= -2 \sum_{i < j} w_{ij} u_i u_j + \sum_{i < j} w_{ij} (u_i^2 + u_j^2)$$

$$= \sum_{i < j} w_{ij} (u_i^2 - u_j^2).$$

Thus,

$$\sum_{i < j} w_{ij}(u_i - u_j)^2 = E \sum_i \pi_i u_i^2 \leq eQu^T = (1 - \lambda) \sum_i \pi_i u_i^2 \Rightarrow E \leq 1 - \lambda.$$
(ii) Given the stationary distribution vector \( \pi \in \mathbb{R}^n \), define an inner product \( \langle \cdot, \cdot \rangle_\pi \) in \( \mathbb{R}^n \) as:

\[
\langle u, v \rangle_\pi = \sum_{i=1}^{n} \pi_i u_i v_i .
\]

By (a slight modification of) a standard result (the Courant-Fischer minimax theorem) in matrix theory, and the fact that \( P \) is reversible with respect to \( \pi \), one can characterise the eigenvalues of \( P \) as:

\[
\lambda_1 = \max \left\{ \frac{\langle u, Pu \rangle_\pi}{\langle u, u \rangle_\pi} \mid u \neq 0 \right\},
\]
\[
\lambda_2 = \max \left\{ \frac{\langle u, Pu \rangle_\pi}{\langle u, u \rangle_\pi} \mid u \perp \pi, u \neq 0 \right\}, \text{ etc.}
\]

In particular,

\[
\lambda_2 \geq \frac{\langle u, Pu \rangle_\pi}{\langle u, u \rangle_\pi} \quad \text{for any } u \neq 0 \text{ such that } \sum_i \pi_i u_i = 0. \quad (6)
\]

Given a set of states \( A \subseteq S \), \( 0 < \pi(A) \leq 1/2 \), we shall apply the bound (6) to the vector \( u \) defined as:

\[
u_i = \begin{cases} 
\frac{1}{\pi(A)}, & \text{if } i \in A \\
-\frac{1}{\pi(A)}, & \text{if } i \in \bar{A}
\end{cases}
\]

Clearly

\[
\sum_i \pi_i u_i = \sum_{i \in A} \frac{\pi_i}{\pi(A)} - \sum_{i \in \bar{A}} \frac{\pi_i}{\pi(A)} = 1 - 1 = 0, \text{ and}
\]

\[
\langle u, u \rangle_\pi = \sum_i \pi_i u_i^2 = \sum_{i \in A} \frac{\pi_i}{\pi(A)^2} + \sum_{i \in \bar{A}} \frac{\pi_i}{\pi(A)^2} = \frac{1}{\pi(A)} + \frac{1}{\pi(A)},
\]

so let us compute the value of \( \langle u, Pu \rangle_\pi \).

The task can be simplified by representing \( P \) as \( P = I_n - (I_n - P) \), and first com-
puting \(\langle u, (I - P)u\rangle_\pi\):

\[
\langle u, (I - P)u\rangle_\pi = \sum_i \pi_i u_i \sum_j (I - P)_{ij} u_j
\]

\[
= -\sum_i \sum_{j \neq i} \pi_i u_i p_{ij} u_j + \sum_i \sum_{j \neq i} \pi_i u_i p_{ij} u_i
\]

\[
= \sum_i \sum_{j \neq i} \pi_i p_{ij} (u_i^2 - u_i u_j)
\]

\[
= \sum_{i < j} \pi_i p_{ij} (u_i - u_j)^2
\]

\[
= \sum_{i \in F} \pi_i p_{ij} \left( \frac{1}{\pi(A)} + \frac{1}{\pi(A)} \right)^2
\]

\[
= \left( \frac{1}{\pi(A)} + \frac{1}{\pi(A)} \right)^2 F_A.
\]

Thus,

\[
\lambda_2 \geq \frac{\langle u, Pu \rangle_\pi}{\langle u, u \rangle_\pi} = \frac{1}{\langle u, u \rangle_\pi} \left( \langle u, u \rangle_\pi - \langle u, (I - P)u \rangle_\pi \right)
\]

\[
= 1 - \frac{1}{\langle u, u \rangle_\pi} \cdot \langle u, (I - P)u \rangle_\pi
\]

\[
= 1 - \left( \frac{1}{\pi(A)} + \frac{1}{\pi(A)} \right)^{-1} \left( \frac{1}{\pi(A)} + \frac{1}{\pi(A)} \right)^2 \cdot F_A
\]

\[
= 1 - \left( \frac{1}{\pi(A)} + \frac{1}{\pi(A)} \right) \cdot F_A
\]

\[
\geq 1 - \frac{2}{\pi(A)} \cdot F_A = 1 - 2\Phi_A.
\]

Since the bound (7) holds for any \(A \subseteq S\) such that 0 < \(\pi(A) \leq 1/2\), it follows that it holds also for the conductance

\[
\Phi = \min_{0 < \pi(A) \leq 1/2} \Phi_A.
\]

Thus, we have shown that \(\lambda_2 \geq 1 - 2\Phi\), which completes the proof. □

Despite the elegance of the conductance approach, it can be sometimes (often?) difficult to apply in practice – computing graph conductance can be quite difficult. Also the bounds obtained are not necessary the best possible; in particular the square in the upper bound \(\lambda_2 \leq 1 - \Phi^2/2\) is unfortunate.
An alternative approach, which is sometimes easier to apply, and can even yield better bounds, is based on the construction of so-called “canonical paths” between states of a Markov chain.

Consider again a regular, reversible Markov chain with stationary distribution $\pi$, represented as a weighted graph with node set $S$ and edge set $E = \{(i, j) \mid p_{ij} > 0\}$. The weight $w_e$ associated to edge $e = (i, j)$ corresponds to the ergodic flow $\pi_i p_{ij}$ between states $i$ and $j$.

Specify for each pair of states $k, l \in S$ a canonical path $\gamma_{kl}$ connecting them. The paths should intuitively be chosen as short and as nonoverlapping as possible. (For precise statements, see Theorems 3.9 and 3.11 below.)

Denote $\Gamma = \{\gamma_{kl} \mid k, l \in S\}$ and define the unweighted and weighted maximum edge loading induced by $\Gamma$ as:

$$
\rho = \rho(\Gamma) = \max_{e \in E} \frac{1}{w_e} \sum_{\gamma_{kl} \ni e} \pi_k \pi_l
$$

$$
\tilde{\rho} = \tilde{\rho}(\Gamma) = \max_{e \in E} \frac{1}{w_e} \sum_{\gamma_{kl} \ni e} \pi_k \pi_l |\gamma_{kl}|,
$$

where $|\gamma_{kl}|$ is the length (number of edges) of path $\gamma_{kl}$. Note that here the edges are considered to be oriented, so that only paths crossing an edge $e = (i, j)$ in the direction from $i$ to $j$ are counted in determining the loading of $e$.

**Theorem 3.9** For any regular, reversible Markov chain and any choice of canonical paths,

$$
\Phi \geq \frac{1}{2\rho}.
$$

**Proof.** Represent the chain as a weighted graph $G$, where the weight on edge $e = (i, j)$ corresponds to the ergodic flow between states $i$ and $j$:

$$
w_{ij} = \pi_i p_{ij} = \pi_j p_{ji}
$$

Every set of states $A \subseteq S$ determines a cut $(A, \bar{A})$ in $G$, and the conductance of the cut corresponds to its relative weight:

$$
\Phi_A = \frac{w(A, \bar{A})}{\pi(A)} = \frac{1}{\pi(A)} \sum_{i \in A, j \in \bar{A}} w_{ij}.
$$

Let then $A$ be a set with $0 < \pi(A) \leq \frac{1}{2}$ that minimises $\Phi_A$, and thus has $\Phi_A = \Phi$. Assume some choice of canonical paths $\Gamma = \{\gamma_{ij}\}$, and assign to each path $\gamma_{ij}$ a
“flow” of value \( \pi_i \pi_j \). Then the total amount of flow crossing the cut \((A, \bar{A})\) is
\[
\sum_{i \in A, j \in \bar{A}} \pi_i \pi_j = \pi(A) \pi(\bar{A}),
\]
but the cut edges (edges crossing the cut) have only total weight \( w(A, \bar{A}) \). Thus, some cut edge \( e \) must have loading
\[
\rho_e = \frac{1}{w_e} \sum_{i,j \in e} \pi_i \pi_j \geq \frac{\pi(A) \pi(\bar{A})}{w(A, A)} \geq \frac{\pi(A)}{2w(A, A)} = \frac{1}{2\Phi}.
\]
The result follows. \( \square \)

**Corollary 3.10** With notations and assumptions as above,
\[
\lambda_2 \leq 1 - \frac{1}{8\rho^2}.
\]

*Proof.* From Theorems 3.6 and 3.9. \( \square \)
A more advanced proof yields a tighter result:

**Theorem 3.11** With notations and assumptions as above:

(i) \( \lambda_2 \leq 1 - \frac{1}{\bar{\rho}} \)

(ii) \( \Delta(t) \leq \left(1 - 1/\bar{\rho}\right)^t \min_{i \in A} \pi_i \)

(iii) \( \tau(\varepsilon) \leq \bar{\rho} \left(\ln \frac{1}{\varepsilon} + \ln \frac{1}{\pi_{\min}}\right) \).

**Example 3.2** Cyclic random walk. Let us consider again the cyclic random walk of Figure 11. Clearly the stationary distribution is \( \pi = [\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}] \), and the ergodic flow on each edge \( e = (i, i+1) \) is
\[
w_e = \pi_i p_{i,i+1} = \frac{1}{n} \cdot \frac{1}{4} = \frac{1}{4n}.
\]
An obvious choice for a canonical path connecting nodes \( k, l \) is the shortest one, with length
\[
|\gamma_{kl}| = \min\{|l-k|, |l-k+n|, |k-l+n|\}.
\]
Part I. Markov Chains and Stochastic Sampling

Figure 12: Transition graph for three-element permutations.

It is fairly easy to see that each (oriented) edge is now travelled by 1 canonical path of length 1, 2 of length 2, 3 of length 3, \ldots, \frac{n}{2} of length \frac{n}{2} (actually the last one is just an upper bound). Thus:

$$\rho = \max_{e} \frac{1}{w_{e}} \sum_{\gamma_{ij} \in e} \pi_{i} \pi_{j} | \gamma_{ij} | \leq 4n \sum_{r=1}^{n/2} \frac{1}{r^2} \cdot r^{2}$$

$$= \frac{4}{n} \cdot \frac{1}{6} \cdot n \cdot \left(\frac{n}{2} + 1\right) \cdot (n + 1) = \frac{1}{6} (n + 1) (n + 2)$$

$$\Rightarrow \tau(\epsilon) \leq \frac{1}{6} (n + 1) (n + 2) \left(\ln n + \ln \frac{1}{\epsilon}\right)$$

$$= \frac{1}{6} n^{2} \left(\ln n + \frac{1}{\epsilon}\right) + O \left(n \left(\ln n + \frac{1}{\epsilon}\right)\right).$$

Example 3.3 Sampling permutations. Let us consider the Markov chain whose states are all possible permutations of \([n] = \{1, 2, \ldots, n\}\), and for any permutations \(s, t \in S_{n}\):

$$p_{st} = \begin{cases} \frac{1}{n!}, & \text{if } s = t, \\ \frac{1}{2} \cdot \left(\binom{n}{2}\right)^{-1}, & \text{if } s \text{ can be changed to } t \text{ by transposing two elements,} \\ 0, & \text{otherwise} \end{cases}$$

Thus, e.g. for \(n = 3\) we obtain the transition graph in Figure 12. Clearly, the stationary distribution for this chain is \(\pi = \left[\frac{1}{n!}, \frac{1}{n!}, \ldots, \frac{1}{n!}\right]\), and the ergodic flow on each edge \(\tau = (s, t)\), with \(s \neq t, p_{st} > 0\), is:

$$w_{\tau} = \pi_{s} p_{st} = \frac{1}{n!} \cdot \frac{1}{2} \cdot \left(\frac{n}{2}\right)^{-1}.$$ 

A natural canonical path connecting permutation \(s\) to permutation \(t\) is now obtained as follows:

\(s = s_{0} \rightarrow s_{1} \rightarrow s_{2} \rightarrow \cdots \rightarrow s_{n-1} = t.\)
where at each \( s_k, s_k(k) = t(k) \). (Thus, each \( s_k \) matches \( t \) up to element \( k, s_k(1 \ldots k) = t(1 \ldots k) \).)

Thus, e.g. the canonical path connecting \( s = (1234) \) to \( t = (3142) \) is as follows:

\[
(1234) \rightarrow (3|214) \tau \rightarrow (31|24) \rightarrow (314|2).
\]

Now let us denote the set of canonical paths containing a given transition \( \tau : \omega \rightarrow \omega' \) by \( \Gamma(\tau) \). We shall upper bound the size of \( \Gamma(t) \) by constructing an injective mapping \( \sigma_\tau : \Gamma(\tau) \rightarrow S_n \). Obviously, the existence of such a mapping implies that \( |\Gamma(\tau)| \leq n! \).

Suppose \( \tau \) transposes locations \( k + 1 \) and \( l, k + 1 < l \), of permutation \( \omega \). Then for any \( (s,t) \in \Gamma(\tau) \), define the permutation \( z = \sigma_\tau(s,t) \) as follows:

1. Place the elements in \( \omega(1 \ldots k) \) in the locations they appear in \( s \). (Note that permutation \( \omega \) is given and fixed as part of \( \tau \).)

2. Place the remaining elements in the remaining locations in the order they appear in \( t \)

Thus, for example in the above example case:

\[
\omega = (3|214), \quad k = 1
\]

Why is this mapping an injection, i.e. how do we recover \( s \) and \( t \) from a knowledge of \( \tau \) and \( z = \sigma_\tau(s,t) \)? The reasoning goes as follows:

1. \( t = \omega(1 \ldots k) + \text{“other elements in same order as in } z\text{”} \)

2. \( s = \text{“elements in } \omega(1 \ldots k) \text{ at locations indicated in } z\text{” + “other elements in locations deducible from the transposition path } s = s_0 \rightarrow s_1 \rightarrow \cdots \rightarrow s_k = \omega\} “\)

This is somewhat tricky, so let us consider an example. Say \( \omega = (3 \ 1|2 \ 4), k = 2, z = (1 \ 4 \ 3 \ 2) \). Then:

1. \( t = (3 \ 1| \_ \ \_ \_ \_\_4 \ 2) = (3 \ 1|4 \ 2) \)
Thus, we know that for each transition $\tau$,

$$|\Gamma(\tau)| \leq n!$$

We can now obtain a bound on the unweighted maximum edge loading induced by our collection of canonical paths:

$$\rho = \max_{\tau \in E} \frac{1}{q_\tau} \sum_{(s,k) \in \Gamma(\tau)} \pi_\tau \pi_s \leq \left( \frac{1}{n!} \cdot \frac{1}{2} \cdot \left( \frac{n}{2} \right)^{n-1} \right)^{-1} \cdot n! \cdot \left( \frac{1}{n!} \right)^2 = 2n! \left( \frac{n}{2} \right) \cdot n! \cdot \left( \frac{1}{n!} \right)^2 = 2 \cdot \left( \frac{n}{2} \right) = n(n-1).$$

By Theorem 3.9, the conductance of this chain is thus $\Phi \geq \frac{1}{2n(n-1)}$, and by Corollary 3.8, its mixing time is thus bounded by

$$\tau_n(\epsilon) \leq \frac{2}{\Phi^2} \left( \ln \frac{1}{\epsilon} + \ln \frac{1}{\pi_{\text{min}}} \right) \leq 2(2n(n-1))^2 \left( \ln \frac{1}{\epsilon} + \ln n! \right) = O \left( n^4 \left( n \ln n + \ln \frac{1}{\epsilon} \right) \right).$$

### 3.2 Coupling

An important “classical” approach to obtaining convergence results for Markov chains is the coupling method. As a simple case, let $\mathcal{M} = (X_0, X_1, \ldots)$ and $\mathcal{N} = (Y_0, Y_1, \ldots)$ be two independent Markov chains with the same state space $S = \{1, \ldots, n\}$ and the same regular transition probability matrix $P = (p_{ij})$, and consequently the same stationary distribution $\pi$.

Thus, if one considers the Markov chain $\mathcal{M} \times \mathcal{N}$ with random variables $Z_t = (X_t, Y_t)$, one obtains transition probabilities

$$p_{ij,kl}^Z = \Pr(Z_t = (k, l) \mid Z_{t-1} = (i, j)) = \Pr(X_t = k \mid X_{t-1} = i) \cdot \Pr(Y_t = l \mid Y_{t-1} = j) = p_{ik}p_{jl}. $$
Moreover, since $\mathcal{M}$ and $\mathcal{N}$ are regular with stationary distribution $\pi$, then so is $\mathcal{M} \times \mathcal{N}$ with stationary distribution $\pi^Z = \pi^T \pi$ (i.e. $\pi^Z_{ij} = \pi_i \pi_j$).

Note once more that “projected” (marginalised) to its first or the second component, $\mathcal{M} \times \mathcal{N}$ yields realisations of the same process, i.e.

$$\Pr(Z_t = (k, *)) \mid Z_0 = (k_0, l_0)) = \Pr(X_t = k \mid X_0 = k_0)$$

$$= p_{k0}^{(t)}, \text{ independent of } l_0;$$

$$\Pr(Z_t = (*, l) \mid Z_0 = (k_0, l_0)) = \Pr(Y_t = l \mid Y_0 = l_0)$$

$$= p_{l0}^{(t)}, \text{ independent of } k_0. \quad (7)$$

Now define a random variable $T$ that for any realisation of $\mathcal{M} \times \mathcal{N}$ indicates the first time at which $X_t$ and $Y_t$ have the same value, i.e. their coupling time:

$$T = \inf\{t \geq 0 \mid X_t = Y_t\}.$$ 

One can in fact modify the chain $\mathcal{M} \times \mathcal{N}$ so that after coupling the $X$- and $Y$-components not just have the same distributions, but in fact strictly the same values (i.e. $X_t = Y_t \forall t \geq T$), yet the marginal properties (7) stay the same. Simply define $X'_t = (X'_t, Y_t)$, where

$$X'_t = \begin{cases} 
X_t, & t < T, \\
Y_t, & t \geq T.
\end{cases}$$

Let us denote the resulting nonhomogeneous chain by $\mathcal{M} \lvert \mathcal{N}$. Now the projections of $\mathcal{M} \lvert \mathcal{N}$ to its $X$- and $Y$-components are surely not independent, but viewed in isolation, as marginals of $\mathcal{M} \lvert \mathcal{N}$, they have exactly the same stochastic properties.

In particular, in a coupled chain $\mathcal{M} \lvert \mathcal{N}$, let us choose an arbitrary initial state $X_0 = k_0$ for $\mathcal{M}$, and similarly $Y_0 = l_0$ for $\mathcal{N}$, and denote the respective time $t$ distributions as $p^{(t)} = (p_{k0}^{(t)})_k$ and $q^{(t)} = (p_{l0}^{(t)})_l$. Then for any $A \subseteq S$:

$$p^{(t)}(A) = \Pr(X_t \in A)$$

$$\geq \Pr(Y_t \in A \land X_t = Y_t)$$

$$= 1 - \Pr(Y_t \notin A \lor X_t \neq Y_t)$$

$$\geq 1 - \Pr(Y_t \notin A) - \Pr(X_t \neq Y_t)$$

$$= \Pr(Y_t \in A) - \Pr(t < T)$$

$$= q^{(t)}(A) - \Pr(t < T),$$
Part I. Markov Chains and Stochastic Sampling

Figure 13: A realisation of the \((D_t)\) chain.

i.e. \(q^{(t)}(A) - p^{(t)}(A) \leq \Pr(t < T)\). A similar argument shows that also \(p^{(t)}(A) - q^{(t)}(A) \leq \Pr(t < T)\), and so for any \(A \subseteq S\), \(|p^{(t)}(A) - q^{(t)}(A)| \leq \Pr(T > t)\), implying that

\[
d_V(p^{(t)}, q^{(t)}) = \sup_{A \subseteq S} |p^{(t)}(A) - q^{(t)}(A)| \leq \Pr(T > t).
\] (8)

Since the coupling bound (8) holds for arbitrary pairs of initial states, it also holds for arbitrary initial distributions, when the bounding probability \(\Pr(T > t)\) is computed with respect to these distributions.

In particular, if the initial state of the chain \(Y\) is chosen according to the stationary distribution \(\pi\), then \(q^{(t)} = \pi\) for all \(t \geq 0\), and one obtains the convergence bound:

\[
d_V(p^{(t)}, \pi) = \frac{1}{2} \sum_i |p_i^{(t)} - \pi_i| \leq \Pr(T > t).
\] (9)

**Example 3.4** Cyclic random walk. Consider again the cyclic random walk of Figure 11 with \(n\) states, \(n\) even. To obtain an upper bound on the coupling probability \(\Pr(T > t)\), consider two independent copies \((X_t), (Y_t)\) of the walk, initiated at \(X_0 = 1\) and \(Y_0 = \frac{n}{2} + 1\).

Denote \(D_t = Y_t - X_t - \frac{n}{2}\). Then \(D_0 = 0\),

\[
D_{t+1} = \begin{cases} 
D_t - 2 & \text{with prob. } 1/16, \\
D_t - 1 & \text{with prob. } 1/4, \\
D_t & \text{with prob. } 3/8, \\
D_t + 1 & \text{with prob. } 1/4, \\
D_t + 2 & \text{with prob. } 1/16,
\end{cases}
\]

and \(T = \inf\{t | D_t = \pm \frac{n}{2}\}\) (cf. Figure 13). Thus,

\[
\Pr(T > t) = \Pr(|D_t| < \frac{n}{2} \quad \forall \ i = 0, 1, \ldots, t).
\]
To get a very rough upper bound on this probability one can observe that if from any initial state $D_k > -\frac{n}{2}$ there are $n$ consequent increases, then it must be the case that $D_{k+n} > \frac{n}{2}$. The probability of this event is

$$r = \Pr(D_{k+n} \geq D_k + n) \geq \frac{1}{4^n}.$$ 

Consequently,

$$\Pr(T > t) = \Pr\left(|D_i| < \frac{n}{2} \quad \forall \ i = 0, 1, \ldots, t\right) \leq (1 - r)^{\lfloor t/n \rfloor}.$$ 

Thus, we obtain a geometric bound on the convergence rate of this walk:

$$d_V(p^{(t)}, \pi) \leq (1 - 4^{-n})^{\lfloor t/n \rfloor}.$$ 

(However, the constants in the bound are not very good. A more careful analysis of the process $(D_t)$ would surely yield better bounds.)

More generally, a coupling of two Markov chains $(X_t)$ and $(Y_t)$ (or any stochastic processes) is a process $Z_t = (X'_t, Y'_t)$ that has $(X_t)$ and $(Y_t)$ as its marginal distributions.

In the case of finite Markov chains this means that:

$$\Pr(X'_{t+1} = k | X'_t = i, Y'_t = j) = \Pr(X_{t+1} = k | X_t = i) = p_{ik}^X,$$
$$\Pr(Y'_{t+1} = l | X'_t = i, Y'_t = j) = \Pr(Y_{t+1} = l | Y_t = j) = p_{jl}^Y.$$ 

The coupling conditions (10) are trivially satisfied by the independent coupling, where $p_{kl}^{Z,t} = p_{ik}^X p_{jl}^Y$, but the more interesting couplings are the non-independent ones.

In the following Lemma, and also later in this section, mixing times are considered with respect to the total variation distance, i.e. for now

$$\tau(\varepsilon) = \tau^V(\varepsilon) = \min \left\{ t \mid d_V(p^{(i,s)}, \pi) \leq \varepsilon \quad \forall \ s \geq t \text{ and } \forall \text{ initial states } i \right\}.$$ 

**Lemma 3.12 (“Coupling lemma”)** Let $\mathcal{M}$ be a finite, regular Markov chain and $Z_t = (X_t, Y_t)$, $t \geq 0$, a coupling of two copies of $\mathcal{M}$ (i.e. $(Z_t)$ is a Markov chain whose $X$- and $Y$-marginals satisfy the coupling conditions (10) with respect to the transition probabilities of $\mathcal{M}$). Suppose further that $t : [0, 1] \rightarrow \mathbb{N}$ is a function such that given any $\varepsilon \in (0, 1]$, $\Pr(X_t \neq Y_t) \leq \varepsilon$ holds for all $t \geq t(\varepsilon)$, uniformly over the choice of the initial state $(X_0, Y_0)$. Then the mixing time $\tau(\varepsilon)$ of $\mathcal{M}$ is bounded above by $t(\varepsilon)$.
Proof. Let $X_0 = i$ be arbitrary, and choose $Y_0$ according to the stationary distribution $\pi$ of $M$. Fix $\varepsilon \in (0, 1]$ and let $t \geq t(\varepsilon)$. Then for any set of states $A$:

$$
p^{(i,t)}(A) = \Pr(X_t \in A) \\
\geq \Pr(Y_t \in A \land X_t = Y_t) \\
\geq 1 - \Pr(Y_t \notin A) - \Pr(X_t \neq Y_t) \\
\geq \Pr(Y_t \in A) - \varepsilon \\
= \pi(A) - \varepsilon,
$$

and similarly for the set $\bar{A} = S \setminus A$. Thus

$$
|p^{(i,t)}(A) - \pi(A)| \leq \varepsilon \quad \forall \, t \geq t(\varepsilon),
$$

and because $A$ was chosen arbitrarily, also

$$
d_V(p^{(i,t)}, \pi) = \max_{A \subseteq S} |p^{(i,t)}(A) - \pi(A)| \leq \varepsilon \quad \forall \, t \geq t(\varepsilon).
$$

Thus $\tau(\varepsilon) \leq t(\varepsilon)$. □

**Example 3.5** Gibbs sampler for graph colourings. Let $G = (V,E)$ be an undirected graph with maximum node degree $\Delta$. Without loss of generality assume that $V = \{1, \ldots, n\}$. A $q$-colouring of $G$ is a map $\sigma : V \to \{1, \ldots, q\} = Q$ such that

$$(i, j) \in E \implies \sigma(i) \neq \sigma(j).$$

According to so called Brooks’ Theorem, $G$ has a $q$-colouring for any $q \geq \Delta + 1$. (In fact, already for $q \geq \Delta$ unless $G$ contains a $(\Delta + 1)$-clique $K_{\Delta+1}$ as a component.)

For $q \geq \Delta + 2$, one can set up the following Gibbs sampler Markov chain $\mathcal{M}$ to sample $q$-colourings of $G$ asymptotically uniformly at random (cf. Example 2.2, p. 24):

Given a colouring $\sigma \in Q^V$:

(i) select a node $i \in V$ uniformly at random;

(ii) select a legal colour $c$ for $i$ uniformly at random ($c$ is legal for $i$ if $c \neq \sigma(j) \forall \, j \in \Gamma(i)$);

(iii) recolour $i$ with colour $c$ (i.e. move from $\sigma$ to $\sigma'$, where $\sigma'(i) = c$ and $\sigma'(j) = \sigma(j)$ for $j \neq i$).