Chapter 18
Convergence of Markov Chains

We consider a Markov chain X with invariant distribution π and investigate conditions under which the distribution of X_n converges to π for $n \to \infty$. Essentially it is necessary and sufficient that the state space of the chain cannot be decomposed into subspaces

- that the chain does not leave
- or that are visited by the chain periodically; e.g., only for odd n or only for even n.

In the first case, the chain would be called *reducible*, and in the second case, it would be *periodic*.

We study periodicity of Markov chains in the first section. In the second section, we prove the convergence theorem. The third section is devoted to applications of the convergence theorem to computer simulations with the so-called Monte Carlo method. In the last section, we describe the speed of convergence to the equilibrium by means of the spectrum of the transition matrix.

18.1 Periodicity of Markov Chains

We study the conditions under which a positive recurrent Markov chain X on the countable space E (and with transition matrix p), started in an arbitrary $\mu \in \mathcal{M}_1(E)$, converges in distribution to an invariant distribution π; that is, $\mu p^n \overset{n\to\infty}{\longrightarrow} \pi$. Clearly, it is necessary that π be the *unique* invariant distribution; that is, up to a factor π it is the unique left eigenvector of p for the eigenvalue 1. As shown in Theorem 17.49, for this uniqueness it is sufficient that the chain be irreducible.

In order for $\mu p^n \overset{n\to\infty}{\longrightarrow} \pi$ to hold for every $\mu \in \mathcal{M}_1(E)$, a certain contraction property of p is necessary. Manifestly, 1 is the largest (absolute value of an) eigenvalue of p. However, p is sufficiently contractive only if the multiplicity of the eigenvalue 1 is exactly 1 and if there are no further (possibly complex-valued) eigenvalues of modulus 1.

For the latter property, it is not sufficient that the chain be irreducible. For example, consider on $E = \{0, \ldots, N-1\}$ the Markov chain with transition matrix

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_18,

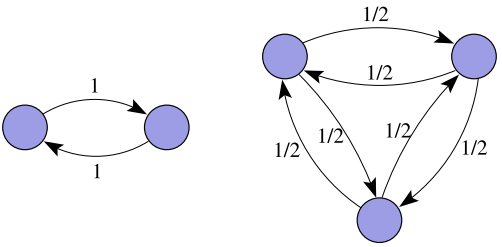

Fig. 18.1 The left Markov chain is periodic with period 2, and the right Markov chain is aperiodic

$p(x, y) = \mathbb{1}_{\{y=x+1 \pmod N\}}$. The eigenvalue 1 has the multiplicity 1. However, all complex Nth roots of unity $e^{2\pi ik/N}$, $k = 0, \ldots, N-1$, are eigenvalues of modulus 1. Clearly, the uniform distribution on E is invariant but $\lim_{n\to\infty} \delta_x p^n$ does not exist for any $x \in E$. In fact, every point is visited periodically after N steps. In order to obtain criteria for the convergence of Markov chains, we thus have to understand periodicity first. Thereafter, for irreducible *aperiodic* chains, we state the convergence theorem.

If $m, n \in \mathbb{N}$, then write $m|n$ if m is a divisor of n; that is, if $\frac{n}{m} \in \mathbb{N}$. If $M \subset \mathbb{N}$, then denote by $\gcd(M)$ the greatest common divisor of all $n \in M$. In the following, let X be a Markov chain on the countable space E with transition matrix p.

Definition 18.1

(i) For $x, y \in E$, define

$$N(x, y) := \{n \in \mathbb{N}_0 : p^n(x, y) > 0\}.$$

For any $x \in E$, $d_x := \gcd(N(x, x))$ is called the *period* of the state x.
(ii) If $d_x = d_y$ for all $x, y \in E$, then $d := d_x$ is called the period of X.
(iii) If $d_x = 1$ for all $x \in E$, then X is called *aperiodic*.

See Figs. 18.1 and 18.2 for illustrations of aperiodic and periodic Markov chains.

Lemma 18.2 *For any $x \in E$, there exists an $n_x \in \mathbb{N}$ with*

$$p^{nd_x}(x, x) > 0 \quad \text{for all } n \geq n_x. \tag{18.1}$$

Proof Let $k_1, \ldots, k_r \in N(x, x)$ with $\gcd(\{k_1, \ldots, k_r\}) = d_x$. Then, for all $m_1, \ldots, m_r \in \mathbb{N}_0$, we also have $\sum_{i=1}^r k_i m_i \in N(x, x)$. Basic number theory then yields that, for every $n \geq n_x := r \cdot \prod_{i=1}^r (k_i/d_x)$, there are numbers $m_1, \ldots, m_r \in \mathbb{N}_0$ with $nd_x = \sum_{i=1}^r k_i m_i$. Hence (18.1) holds. □

The problem of finding the *smallest* number N such that any nd_x, $n \geq N$ can be written as a nonnegative integer linear combination of $k_1, \ldots, k_r$ is called the *Frobenius problem*. The general solution is unknown; however, for the case $r = 2$,

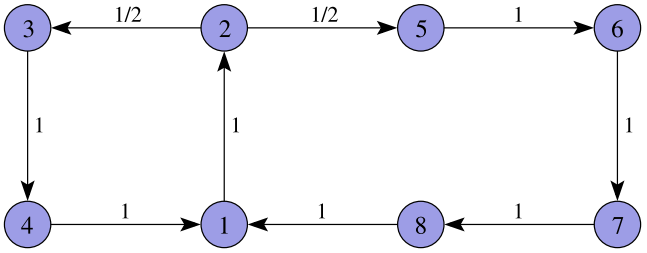

Fig. 18.2 Here $N(8,8) = \{6, 10, 12, 14, 16, \ldots\}$; hence $d_8 := \gcd(\{6, 10, 12, \ldots\}) = 2$ and $n_8 = 5$. The chain thus has period 2. However, $n_1 = 2$ and $n_4 = 4$

Sylvester [162] showed that $N = (k_1/d_x - 1)(k_2/d_x - 1)$ is minimal. In the general case, for N, the upper bound $2\max\{k_i : i = 1, \ldots, r\}^2/(rd_x^2)$ is known; see, e.g., [45].

Lemma 18.3 *Let X be irreducible. Then the following statements hold.*

(i) $d := d_x = d_y$ *for all* $x, y \in E$.
(ii) *For all* $x, y \in E$, *there exist* $n_{x,y} \in \mathbb{N}$ *and* $L_{x,y} \in \{0, \ldots, d-1\}$ *such that*

$$nd + L_{x,y} \in N(x, y) \quad \text{for all } n \geq n_{x,y}. \tag{18.2}$$

$L_{x,y}$ *is uniquely determined, and we have*

$$L_{x,y} + L_{y,z} + L_{z,x} = 0 \pmod{d} \quad \text{for all } x, y, z \in E. \tag{18.3}$$

Proof (i) Let $m, n \in \mathbb{N}_0$ with $p^m(x, y) > 0$ and $p^n(y, z) > 0$. Then

$$p^{m+n}(x, z) \geq p^m(x, y)p^n(y, z) > 0.$$

Hence we have

$$N(x, y) + N(y, z) := \{m + n : m \in N(x, y), n \in N(y, z)\} \subset N(x, z). \tag{18.4}$$

If, in particular, $m \in N(x, y)$, $n \in N(y, x)$ and $k \geq n_y$, then $kd_y \in N(y, y)$; hence $m + kd_y \in N(x, y)$ and $m + n + kd_y \in N(x, x)$. Therefore, $d_x | (m + n + kd_y)$ for every $k \geq n_y$; hence $d_x | d_y$. Similarly, we get $d_y | d_x$; hence $d_x = d_y$.

(ii) Let $m \in N(x, y)$. Then $m + kd \in N(x, y)$ for every $k \geq n_x$. Hence (18.2) holds with

$$n_{x,y} := n_x + \left\lfloor \frac{m}{d} \right\rfloor \quad \text{and} \quad L_{x,y} := m - d\left\lfloor \frac{m}{d} \right\rfloor.$$

Owing to (18.4), we have

$$(n_{x,y} + n_{y,z})d + L_{x,y} + L_{y,z} \in N(x, z).$$

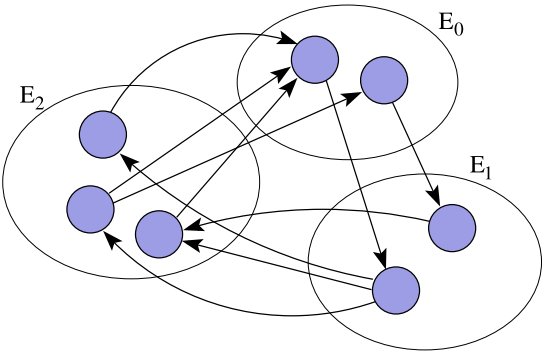

Fig. 18.3 State space decomposition of a Markov chain with period $d = 3$

Together with $z = x$, it follows that $d|(L_{x,y} + L_{y,x})$. Hence the value of $L_{x,y}$ is unique in $\{0, \dots, d-1\}$ and $L_{x,y} = -L_{y,x} \pmod d$. For general z, we infer that $d|(L_{x,y} + L_{y,z} + L_{z,x})$; hence (18.3). □

Theorem 18.4 *Let X be irreducible with period d. Then there exists a disjoint decomposition of the state space*

$$E = \biguplus_{i=0}^{d-1} E_i \tag{18.5}$$

with the property

$$p(x, y) > 0 \quad \textit{and} \quad x \in E_i \quad \Longrightarrow \quad y \in E_{i+1 \pmod d}. \tag{18.6}$$

This decomposition is unique up to cyclic permutations.

See Fig. 18.3 for an illustration of the state space decomposition of a periodic Markov chain.

Property (18.6) says that X visits the E_i one after the other (see Fig. 18.3 or Fig. 18.2, where $d = 2$, $E_0 = \{1, 3, 5, 7\}$ and $E_1 = \{2, 4, 6, 8\}$). Somewhat more formally, we could write: If $x \in E_i$ for some i, then $\mathbf{P}_x[X_n \in E_{i+n \pmod d}] = 1$.

Proof Existence. Fix an arbitrary $x_0 \in E$ and let

$$E_i := \{y \in E : L_{x_0,y} = i\} \quad \text{for } i = 0, \dots, d-1.$$

Clearly, (18.5) holds. Let $i \in \{0, \dots, d-1\}$ and $x \in E_i$. If $y \in E$ with $p(x, y) > 0$, then $L_{x,y} = 1$ and hence $L_{x_0,y} = L_{x_0,x} + L_{x,y} = i + 1 \pmod d$.

Uniqueness. Let $(\widetilde{E}_i,\ i = 0, \dots, d-1)$ be another decomposition that satisfies (18.5) and (18.6). Without loss of generality, assume $E_0 \cap \widetilde{E}_0 \neq \emptyset$ (otherwise permute the $\widetilde{E}_i$ cyclically until this holds). Fix an arbitrary $x_0 \in E_0 \cap \widetilde{E}_0$. By assumption, $p(x_0, y) > 0$ now implies $y \in E_1$ and $y \in \widetilde{E}_1$; hence $y \in E_1 \cap \widetilde{E}_1$. Inductively, we get that $p^{nd+i}(x, y) > 0$ implies $y \in E_i \cap \widetilde{E}_i$ (for all $n \in \mathbb{N}$ and $i = 0, \dots, d-1$).

However, since the chain is irreducible, for every $y \in E$, there exist numbers $n(y)$ and $i(y)$ such that $p^{n(y)d+i(y)}(x_0, y) > 0$; hence $y \in E_{i(y)} \cap \widetilde{E}_{i(y)}$. Therefore, we have $E_i = \widetilde{E}_i$ for every $i = 0, \dots, d-1$. □

18.2 Coupling and Convergence Theorem

Our goal is to use a coupling of two discrete Markov chains that are started in different distributions μ and ν in order to show the convergence theorem for Markov chains.

In the following, let E be a countable space and let p be a stochastic matrix on E. Recall the definition of a general coupling of two probability measures from Definition 17.53.

Definition 18.5 A bivariate process $((X_n, Y_n))_{n\in\mathbb{N}_0}$ with values in $E \times E$ is called a coupling if $(X_n)_{n\in\mathbb{N}_0}$ and $(Y_n)_{n\in\mathbb{N}_0}$ are Markov chains, each with transition matrix p.

A coupling is called successful if $\mathbf{P}_{(x,y)}[\bigcup_{m\geq n}\{X_m \neq Y_m\}] \overset{n\to\infty}{\longrightarrow} 0$ for all $x, y \in E$.

Of course, two independent chains form a coupling, though maybe not the most interesting one.

Example 18.6 (Independent coalescence) The most important coupling is Markov chains that run independently until they coalesce: Let X and Y be independent chains with transition matrix p until they first meet. After that, the chains run together. We call this coupling the *independent coalescent*. The transition matrix is

$$\bar{p}\big((x_1, y_1), (x_2, y_2)\big) = \begin{cases} p(x_1, x_2) \cdot p(y_1, y_2), & \text{if } x_1 \neq y_1, \\ p(x_1, x_2), & \text{if } x_1 = y_1, x_2 = y_2, \\ 0, & \text{if } x_1 = y_1, x_2 \neq y_2. \end{cases}$$

Denote by $\tau := \inf\{n \in \mathbb{N}_0 : X_n = Y_n\}$ the time of coalescence. We can construct the coupling using two independent chains $\tilde{X}$ and $\tilde{Y}$ by defining $X := \tilde{X}$, $\tilde{\tau} := \inf\{n \in \mathbb{N}_0 : \tilde{X}_n = \tilde{Y}_n\}$ and

$$Y_n := \begin{cases} \tilde{Y}_n, & \text{if } n < \tilde{\tau}, \\ X_n, & \text{if } n \geq \tilde{\tau}. \end{cases}$$

Instead of checking by a direct computation that this process (X, Y) is indeed a coupling with transition matrix $\bar{p}$, consider the construction of Markov chains from Theorem 17.17: Let $(R_n(x) : n \in \mathbb{N}_0, x \in E)$ be independent random variables with distribution $\mathbf{P}[R_n(x_1) = x_2] = p(x_1, x_2)$, and let $\tilde{R}_n((x_1, y_1)) = (R_n(x_1), R_n(y_1))$. Then $(\tilde{R}_n)_{n\in\mathbb{N}_0}$ is independent and we have

$$\mathbf{P}\big[\tilde{R}_n\big((x_1, y_1)\big) = (x_2, y_2)\big] = \bar{p}\big((x_1, y_1), (x_2, y_2)\big).$$

As we saw in Theorem 17.17, by $X_{n+1} := R_n(X_n)$ and $Y_{n+1} := R_n(Y_n)$, two Markov chains X and Y are defined with transition matrix p. On the other hand, we have $(X_{n+1}, Y_{n+1}) = \tilde{R}_n((X_n, Y_n))$. Hence the bivariate process is indeed a coupling with transition matrix $\bar{p}$. ◊

Example 18.7 Let $E = \mathbb{Z}$ and $p(x, y) = 1/3$ if $|x - y| \le 1$ and 0 otherwise. Clearly, p is the transition matrix of an aperiodic recurrent random walk on $\mathbb{Z}$. We will show that we can obtain a successful coupling by coalescing independent chains.

Accordingly, let $\tilde{X}$ and $\tilde{Y}$ be independent random walks with transition matrix p. Then the difference chain $(Z_n)_{n\in\mathbb{N}_0} := (\tilde{X}_n - \tilde{Y}_n)_{n\in\mathbb{N}_0}$ is a symmetric random walk with finite expectation and hence recurrent. Furthermore, Z is irreducible. For any two points $x, y \in \mathbb{Z}$, we thus have

$$\mathbf{P}_{(x,y)}[\tilde{\tau} < \infty] = \mathbf{P}_{x-y}[Z_n = 0 \text{ for some } n \in \mathbb{N}_0] = 1.$$

Therefore, X and Y coalesce almost surely. ◊

Recurrence, irreducibility and aperiodicity alone are not sufficient for the independent coalescence coupling to be successful. In Exercise 18.2.4, an example is studied that shows that spacial homogeneity cannot easily be dropped if we want to have a successful coupling. Dropping the assumption of recurrence is easier, as the following theorem shows.

Theorem 18.8 *Let X be an arbitrary aperiodic and irreducible random walk on $\mathbb{Z}^d$ with transition matrix p. Then there exists a successful coupling (X, Y).*

Proof Step 1. First, consider the case where $p(0, x) = 3^{-d}$ for all $x \in \{-1, 0, 1\}^d$. The individual coordinates $X^{(1)}, \dots, X^{(d)}$ of X are independent random walks on $\mathbb{Z}$ with transition probabilities $\mathbf{P}_0[X_1^{(i)} = x_i] = 1/3$ for $x_i = -1, 0, 1$. By Example 18.7, we can construct independent successful couplings $(X^{(i)}, Y^{(i)})$, $i = 1, \dots, d$, with merging times $\tau^{(i)}$. Define $Y = (Y^{(1)}, \dots, Y^{(d)})$ and $\tau = \max\{\tau^{(1)}, \dots, \tau^{(d)}\} < \infty$. Then (X, Y) is a successful coupling and $X_n = Y_n$ for $n \ge \tau$.

Step 2. Now, consider the case where

$$\lambda := 3^d \min\bigl\{p(0, x) : x \in \{-1, 0, 1\}^d\bigr\} > 0.$$

If $\lambda = 1$, then the condition of Step 1 is fulfilled and we are done. Hence, we assume that $\lambda \in (0, 1)$. We define the transition matrix $\hat{p}$ on $\mathbb{Z}^d$ by $\hat{p}(x, y) = 3^{-d}$ for $y - x \in \{-1, 0, 1\}^d$. Note that also $\check{p} := (p - \lambda\hat{p})/(1 - \lambda)$ is the transition matrix of a random walk on $\mathbb{Z}^d$ and that

$$p = \lambda\hat{p} + (1 - \lambda)\check{p}.$$

Let $\hat{X}$ and $\check{X}$ be independent random walks with transition matrices $\hat{p}$ and $\check{p}$, respectively. Assume that $\hat{X}_0 = X_0$ and $\check{X}_0 = 0$. Furthermore, let $Z_1, Z_2, \dots$ be i.i.d.

Bernoulli random variables with parameter λ that are independent of $\hat{X}$ and $\check{X}$. Define $S_n := Z_1 + \ldots + Z_n$ for $n \in \mathbb{N}$ and

$$X_n := \hat{X}_{S_n} + \check{X}_{n-S_n}.$$

That is, in each time step, a coin flip decides whether X makes a jump according to the matrix $\hat{p}$ or $\check{p}$. Hence X is a random walk with transition matrix p.

By Step 1, there exists a successful coupling $(\hat{X}, \hat{Y})$ such that $\hat{Y}$ is independent of $\check{X}$ and $Z_1, Z_2, \ldots$. Consequently,

$$Y_n := \hat{Y}_{S_n} + \check{X}_{n-S_n}, \quad n \in \mathbb{N},$$

is also a random walk with transition matrix p. Since we have $S_n \to \infty$ almost surely, the coupling (X, Y) is successful.

Step 3. Finally, we consider the general situation. Since X is irreducible and aperiodic, by Lemma 18.3(ii), there exists an $N \in \mathbb{N}$, such that the N-step transition matrix fulfills

$$p^N(0, x) > 0 \quad \text{for all } x \in \{-1, 0, 1\}^d.$$

Hence, the random walk $X' = (X'_n)_{n\in\mathbb{N}} := (X_{nN})_{n\in\mathbb{N}}$ fulfills the condition from Step 2. Let (X', Y') be the coupling that was constructed in Step 2 and let

$$\tau := \inf\bigl\{n \in \mathbb{N}_0 : X'_m = Y'_m \text{ for all } m \geq n\bigr\}.$$

Then Y' is a random walk with transition matrix p^N. For $n \in \mathbb{N}_0$, define $Y_{nN} := Y'_n$. It remains to close the gaps between the points $\{0, N, 2N, \ldots\}$ in such a way that Y is a random walk and (X, Y) is a successful coupling.

Let $(U^{x,y,n} : x, y \in \mathbb{Z}^d, n \in \mathbb{N}_0)$ be an independent family of $(\mathbb{Z}^d)^{N-1}$-valued random variables $U^{x,y,n} = (U_1^{x,y,n}, \ldots, U_{N-1}^{x,y,n})$ such that

$$\mathbf{P}\bigl[(X_1, \ldots, X_{N-1}) \in \cdot \mid X_0 = x, X_N = y\bigr] = \mathbf{P}_{U^{x,y,n}}$$

for all $x, y \in \mathbb{Z}^d$ with $p^N(x, y) > 0$ and for all $n \in \mathbb{N}_0$. We further assume that the $U^{x,y,n}$ are independent of X and Y'. For $k \in \{nN + 1, \ldots, (n + 1)N - 1\}$, define

$$Y_k := \begin{cases} U_{k-nN}^{Y'_n, Y'_{n+1}, n}, & \text{if } n < \tau, \\ X_k, & \text{else.} \end{cases}$$

It is easy to check that Y is indeed a random walk with transition matrix p. By construction, the coupling (X, Y) is successful. □

Theorem 18.9 *Let X be a Markov chain on E with transition matrix p. If there exists a successful coupling, then every bounded harmonic function is constant.*

Proof Let $f : E \to \mathbb{R}$ be bounded and harmonic; hence $pf = f$. Let $x, y \in E$, and let (X, Y) be a successful coupling. By Lemma 17.45, $(f(X_n))_{n\in\mathbb{N}_0}$ and

$(f(Y_n))_{n\in\mathbb{N}_0}$ are martingales; hence we have

$$\big|f(x)-f(y)\big| = \big|\mathbf{E}_{(x,y)}\big[f(X_n)-f(Y_n)\big]\big| \le 2\|f\|_\infty \mathbf{P}_{(x,y)}[X_n \neq Y_n] \overset{n\to\infty}{\longrightarrow} 0. \ \square$$

Corollary 18.10 *If X is an irreducible random walk on $\mathbb{Z}^d$, then every bounded harmonic function is constant.*

This statement holds more generally if we replace $\mathbb{Z}^d$ by a locally compact Abelian group. In that form, the theorem goes back to Choquet and Deny [24], see also [143].

Proof Let p be the transition matrix of X. Let $\bar{X}$ be a Markov chain with transition matrix $\bar{p}(x,y) = \frac{1}{2}p(x,y) + \frac{1}{2}\mathbb{1}_{\{x\}}(y)$. Clearly, X and $\bar{X}$ have the same harmonic functions. Now $\bar{X}$ is an aperiodic irreducible random walk; hence, by Theorem 18.8, there is a successful coupling for all initial states. $\square$

Theorem 18.11 *Let p be the transition matrix of an irreducible, positive recurrent, aperiodic Markov chain on E. Then the independent coalescent chain is a successful coupling.*

Proof Let $\tilde{X}$ and $\tilde{Y}$ be two independent Markov chains on E, each with transition matrix p. Then the bivariate Markov chain $Z := ((\tilde{X}_n, \tilde{Y}_n))_{n\in\mathbb{N}_0}$ has the transition matrix $\widetilde{p}$ defined by

$$\widetilde{p}\big((x_1,y_1),(x_2,y_2)\big) = p(x_1,x_2)\cdot p(y_1,y_2).$$

We first show that the matrix $\widetilde{p}$ is irreducible. Only here do we need aperiodicity of p. Accordingly, fix $(x_1,y_1),(x_2,y_2)\in E\times E$. Then, by Lemma 18.2, there exists an $m_0\in\mathbb{N}$ such that

$$p^n(x_1,x_2)>0 \quad \text{and} \quad p^n(y_1,y_2)>0 \quad \text{for all } n\ge m_0.$$

For $n\ge m_0$, we thus have $\widetilde{p}^n((x_1,y_1),(x_2,y_2))>0$. Hence $\widetilde{p}$ is irreducible.

Now define the stopping time τ of the first entrance of $(\tilde{X},\tilde{Y})$ into the diagonal $D := \{(x,x) : x\in E\}$ by $\tau := \inf\{n\in\mathbb{N}_0 : \tilde{X}_n = \tilde{Y}_n\}$. Let π be the invariant distribution of $\tilde{X}$. Then, clearly, the product measure $\pi\otimes\pi\in\mathcal{M}_1(E\times E)$ is an (and then *the*) invariant distribution of $(\tilde{X},\tilde{Y})$. Thus $(\tilde{X},\tilde{Y})$ is positive recurrent (hence, in particular, recurrent) by Theorem 17.51. Therefore, $\mathbf{P}_{(x,y)}[\tau<\infty]=1$ for all initial points $(x,y)\in E\times E$ of Z. $\square$

Theorem 18.12 *Let X be a Markov chain with transition matrix p such that there exists a successful coupling. Then $\|(\mu-\nu)p^n\|_{TV} \overset{n\to\infty}{\longrightarrow} 0$ for all $\mu,\nu\in\mathcal{M}_1(E)$.*

If X is aperiodic and positive recurrent with invariant distribution π, then we have $\|\mathcal{L}_\mu[X_n]-\pi\|_{TV} \overset{n\to\infty}{\longrightarrow} 0$ for all $\mu\in\mathcal{M}_1(E)$.

Proof It is enough to consider the case $\mu = \delta_x$, $\nu = \delta_y$ for some $x, y \in E$. Summation over x and y yields the general case. Let $(X_n, Y_n)_{n\in\mathbb{N}_0}$ be a successful coupling. Then

$$\left\| (\delta_x - \delta_y) p^n \right\|_{TV} \leq 2\mathbf{P}_{(x,y)}[X_n \neq Y_n] \overset{n\to\infty}{\longrightarrow} 0. \qquad \square$$

We summarize the connection between aperiodicity and convergence of distributions of X in the following theorem.

Theorem 18.13 (Convergence of Markov chains) *Let X be an irreducible, positive recurrent Markov chain on E with invariant distribution π. Then the following are equivalent*:

(i) *X is aperiodic.*
(ii) *For every $x \in E$, we have*

$$\left\| \mathcal{L}_x[X_n] - \pi \right\|_{TV} \overset{n\to\infty}{\longrightarrow} 0. \tag{18.7}$$

(iii) *Equation (18.7) holds for some $x \in E$.*
(iv) *For every $\mu \in \mathcal{M}_1(E)$, we have $\|\mu p^n - \pi\|_{TV} \overset{n\to\infty}{\longrightarrow} 0$.*

Proof The implications (iv) $\iff$ (ii) $\implies$ (iii) are evident. The implication (i) $\implies$ (ii) was shown in Theorem 18.12. Hence we only show (iii) $\implies$ (i).

"(iii) $\implies$ (i)" Assume that (i) does not hold. If X has period $d \geq 2$, and if $n \in \mathbb{N}$ is not a multiple of d, then, by Theorem 17.51,

$$\left\| \delta_x p^n - \pi \right\|_{TV} \geq \left| p^n(x,x) - \pi\big(\{x\}\big) \right| = \pi\big(\{x\}\big) > 0.$$

Thus, for every $x \in E$, we have $\limsup_{n\to\infty} \|\delta_x p^n - \pi\|_{TV} > 0$. Therefore, (iii) does not hold. $\square$

Exercise 18.2.1 Let d_P be the Prohorov metric (see (13.3) and Exercise 13.2.1). Show that $d_P(P, Q) \leq \sqrt{d_W(P, Q)}$ for all $P, Q \in \mathcal{M}_1(E)$. If E has a finite diameter $\operatorname{diam}(E)$, then $d_W(P, Q) \leq (\operatorname{diam}(E) + 1) d_P(P, Q)$ for all $P, Q \in \mathcal{M}_1(E)$.

Exercise 18.2.2 Consider the bivariate process (X, Y) that was constructed from $\tilde{X}$ and $\tilde{Y}$ in Example 18.6. Show that (X, Y) is a coupling with transition matrix $\bar{p}$.

Exercise 18.2.3 Let X be an arbitrary aperiodic irreducible recurrent random walk on $\mathbb{Z}^d$. Show that, for any two starting points, the independent coalescent coupling is successful.

Hint: Show that the difference of two independent recurrent random walks is a recurrent random walk.

Exercise 18.2.4 Let X be a Markov chain on $\mathbb{Z}^2$ with transition matrix

$$p\big((x_1,x_2),(y_1,y_2)\big) = \begin{cases} \frac{1}{4}, & \text{if } x_1 = 0, \|y-x\|_2 = 1, \\ \frac{1}{4}, & \text{if } x_1 \neq 0 \text{ and } y_1 = x_1 \pm 1, x_2 = y_2, \\ \frac{1}{2}, & \text{if } x_1 \neq 0 \text{ and } y_1 = x_1, x_2 = y_2, \\ 0, & \text{else.} \end{cases}$$

Intuitively, this is the symmetric simple random walk whose vertical transitions are all blocked away from the vertical axis. Show that X is null recurrent, irreducible and aperiodic and that independent coalescence does not give a successful coupling.

18.3 Markov Chain Monte Carlo Method

Let E be a finite set and let $\pi \in \mathcal{M}_1(E)$ with $\pi(x) := \pi(\{x\}) > 0$ for every $x \in E$. We consider the problem of sampling a random variable Y with distribution π on a computer. For example, this is a relevant problem if E is a very large set and if sums of the type $\sum_{x\in E} f(x)\pi(x)$ have to be approximated numerically by the estimator $n^{-1}\sum_{i=1}^{n} f(Y_i)$ (see Example 5.21).

Assume that our computer has a random number generator that provides realizations of i.i.d. random variables $U_1, U_2, \ldots$ that are uniformly distributed on $[0,1]$. In order for the problem to be interesting, assume also that the distribution π cannot be constructed directly too easily.

Metropolis Algorithm

We have seen already in Example 17.19 how to simulate a Markov chain on a computer. Now the idea is to construct a Markov chain X whose distribution converges to π in the long run. If we simulate such a chain and let it run long enough this should give a sample that is distributed approximately like π. The chain should be designed so that at each step, only a small number of transitions are possible in order to ensure that the procedure described in Example 17.19 works efficiently. (Of course, the chain with transition matrix $p(x,y) = \pi(y)$ converges to π, but this does not help a lot.) This method of producing (approximately) π-distributed samples and using them to estimate expected values of functions of interest is called the *Markov chain Monte Carlo method* or, briefly, *MCMC* (see [15, 112, 119]).

Let q be the transition matrix of an arbitrary irreducible Markov chain on E (with $q(x,y) = 0$ for most $y \in E$). We use this to construct the Metropolis matrix (see [70, 114]).

Definition 18.14 Define a stochastic matrix p on E by

$$p(x,y) = \begin{cases} q(x,y)\min\left(1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right), & \text{if } x \neq y, q(x,y) > 0, \\ 0, & \text{if } x \neq y, q(x,y) = 0, \\ 1 - \sum_{z \neq x} p(x,z), & \text{if } x = y. \end{cases}$$

p is called the *Metropolis matrix* of q and π.

Note that p is *reversible* (see Section 19.2); that is, for all $x, y \in E$, we have

$$\pi(x)p(x,y) = \pi(y)p(y,x). \tag{18.8}$$

In particular, π is invariant (check this!). We thus obtain the following theorem.

Theorem 18.15 *Assume that q is irreducible and that for any $x, y \in E$, we have $q(x,y) > 0$ if and only if $q(y,x) > 0$. Then the Metropolis matrix p of q and π is irreducible with unique invariant distribution π. If, in addition, q is aperiodic, or q is not reversible with respect to π, then p is aperiodic.*

In order to simulate a chain X that converges to π, we take a reference chain with transition matrix q and use the *Metropolis algorithm*: If the chain with transition matrix q proposes a transition from the present state x to state y, then we accept this proposal with probability

$$\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)} \wedge 1.$$

Otherwise the chain X stays at x.

In the definition of p, the distribution π appears only in terms of the quotients $\pi(y)/\pi(x)$. In many cases of interest, these quotients are easy to compute even though $\pi(x)$ and $\pi(y)$ are not. We illustrate this with an example.

Example 18.16 (Ising model) The Ising model (pronounced like the English word "easing") is a thermodynamical (and quantum mechanical) model for ferromagnetism in crystals. It makes the following assumptions:

- Atoms are placed at the sites of a lattice Λ (for example, $\Lambda = \{0, \ldots, N-1\}^2$).
- Each atom $i \in \Lambda$ has a magnetic spin $x(i) \in \{-1, 1\}$ that either points upwards ($x(i) = +1$) or downwards ($x(i) = -1$).
- Neighboring atoms interact.
- Due to thermic fluctuations, the state of the system is random and distributed according to the so-called *Boltzmann distribution* π on the state space $E := \{-1, 1\}^\Lambda$. A parameter of this distribution is the inverse temperature $\beta = \frac{1}{T} \geq 0$ (with T the absolute temperature).

Define the local energy that describes the energy level of a single atom at $i \in \Lambda$ as a function H^i of the state x of the whole system,

$$H^i(x) = \frac{1}{2} \sum_{j \in \Lambda : i \sim j} \mathbb{1}_{\{x(i) \neq x(j)\}}.$$

Here $i \sim j$ indicates that i and j are neighbors in Λ (that is, coordinate-wise mod N, we also speak of *periodic boundary conditions*). The total energy (or Hamilton function) of the system in state x is the sum of the individual energies,

$$H(x) = \sum_{i \in \Lambda} H^i(x) = \sum_{i \sim j} \mathbb{1}_{\{x(i) \neq x(j)\}}.$$

The Boltzmann distribution π on $E := \{-1, 1\}^\Lambda$ for the inverse temperature $\beta \geq 0$ is defined by

$$\pi(x) = Z_\beta^{-1} \exp\bigl(-\beta H(x)\bigr),$$

where the *partition sum* $Z_\beta = \sum_{x \in E} \exp(-\beta H(x))$ is the normalising constant such that π is a probability measure.

Macroscopically, the individual spins cannot be observed but the average magnetization can; that is, the modulus of the average of all spins,

$$m_\Lambda(\beta) = \sum_{x \in E} \pi(x) \left| \frac{1}{\#\Lambda} \sum_{i \in \Lambda} x(i) \right|.$$

If we consider a very large system, then we are close to the so-called thermodynamic limit

$$m(\beta) := \lim_{\Lambda \uparrow \mathbb{Z}^d} m_\Lambda(\beta).$$

Using a contour argument, as for percolation (see [127]), one can show that (for $d \geq 2$) there exists a critical value $\beta_c = \beta_c(d) \in (0, \infty)$ such that

$$m(\beta) \begin{cases} > 0, & \text{if } \beta > \beta_c, \\ = 0, & \text{if } \beta < \beta_c. \end{cases} \tag{18.9}$$

See Fig. 18.4 for a computer simulation of the curve $\beta \mapsto m(\beta)$.

For a similar model, the Weiss ferromagnet, we will prove in Example 23.20 the existence of such a *phase transition*. In the physical literature, $T_c := 1/\beta_c$ is called the *Curie temperature* for spontaneous magnetization. This is a material-dependent constant (chromium bromide (CrBr) 37 Kelvin, nickel 645 K, iron 1017 K, cobalt 1404 K). Below the Curie temperature, these materials are magnetic, and above it they are not. Below the critical temperature, the magnetization increases with decreasing temperature. We will see in a computer simulation that the Ising model displays this critical temperature effect.

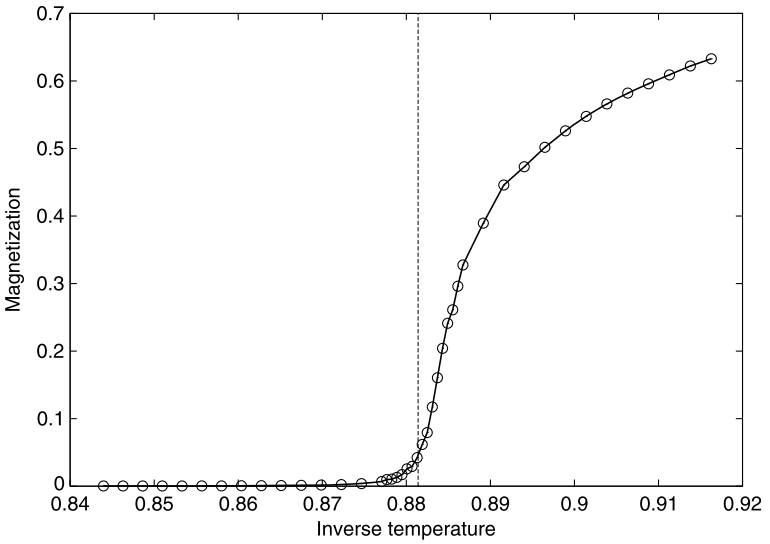

Fig. 18.4 Computer simulation of the magnetization curve of the Ising model on a 1000×1000 grid. The *dashed vertical line* indicates the critical inverse temperature

If $x \in E$, then denote by $x^{i,\sigma}$ the state in which at site i the spin is changed to $\sigma \in \{-1, +1\}$; that is,

$$x^{i,\sigma}(j) = \begin{cases} \sigma, & \text{if } j = i, \\ x(j), & \text{if } j \neq i. \end{cases}$$

Furthermore, define the state x^i in which the spin at i is reversed, $x^i := x^{i,-x(i)}$. As reference chain, we choose a chain with transition probabilities

$$q(x, y) = \begin{cases} \frac{1}{\#\Lambda}, & \text{if } y = x^i \text{ for some } i \in \Lambda, \\ 0, & \text{else.} \end{cases}$$

In words, we choose a random site $i \in \Lambda$ (uniformly on Λ) and invert the spin at that site. Clearly, q is irreducible.

The Metropolis algorithm for this chain accepts the proposal of the reference chain with probability 1 if $\pi(x^i) \geq \pi(x)$. Otherwise the proposal is accepted only with probability $\pi(x^i)/\pi(x)$. However, now

$$\begin{aligned} H\big(x^i\big) - H(x) &= \sum_{j:j\sim i} \mathbb{1}_{\{x(j)\neq -x(i)\}} - \sum_{j:j\sim i} \mathbb{1}_{\{x(j)\neq x(i)\}} \\ &= -2 \sum_{j:j\sim i} \left(\mathbb{1}_{\{x(j)\neq x(i)\}} - \frac{1}{2} \right). \end{aligned}$$

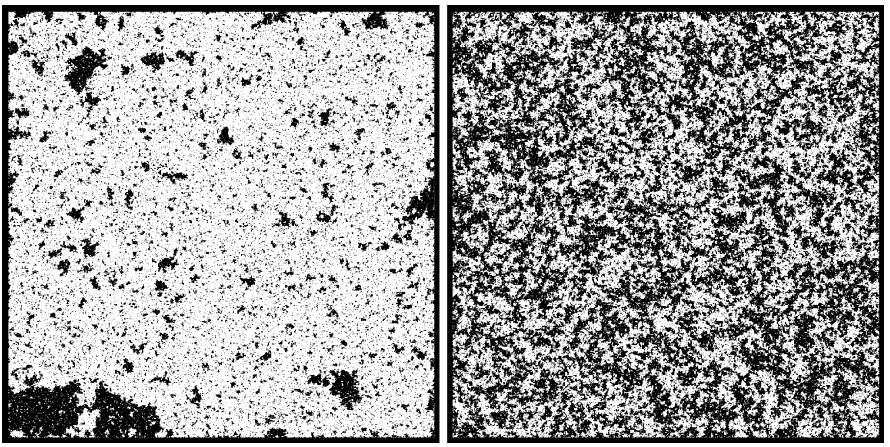

Fig. 18.5 Equilibrium states of the Ising model on an 800×800 grid (*black dot* = spin $+1$). *Left side*: below the critical temperature ($\beta > \beta_c$); *Right side*: above the critical temperature

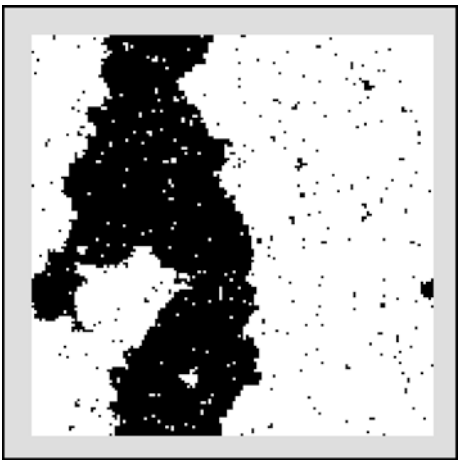

Fig. 18.6 Ising model (150×150 grid) below the critical temperature. Even after a long time, the computer simulation does not produce the equilibrium state but rather so-called metastable states, in which the Weiss domains are clearly visible

Hence $\pi(x^i)/\pi(x) = \exp(-2\beta \sum_{j\sim i}(\mathbb{1}_{\{x(j)=x(i)\}} - \frac{1}{2}))$, and this expression is easy to compute as it depends only on the $2d$ neighboring spins and, in particular, does not require knowledge of the value of Z_β. We thus obtain the Metropolis transition matrix

$$p(x,y) = \begin{cases} \frac{1}{\#\Lambda}\big(1 \wedge \exp\big[2\beta \sum_{j:j\sim i}\big(\mathbb{1}_{\{x(j)\neq x(i)\}} - \frac{1}{2}\big)\big]\big), & \text{if } y = x^i \text{ for some } i \in \Lambda, \\ 1 - \sum_{i\in\Lambda} p\big(x, x^i\big), & \text{if } x = y, \\ 0, & \text{else.} \end{cases}$$

For a practical simulation use the computer's random number generator to produce independent random variables $I_1, I_2, \dots$ and $U_1, U_2, \dots$ with $I_n \sim \mathcal{U}_\Lambda$ and $U_n \sim \mathcal{U}_{[0,1]}$. Then define

$$F_n(x) = \begin{cases} x^{I_n}, & \text{if } U_n \le \exp\big[2\beta \sum_{j:j\sim i} \big(\mathbb{1}_{\{x(j)\neq x(i)\}} - \frac{1}{2}\big)\big], \\ x, & \text{else,} \end{cases}$$

and define the Markov chain $(X_n)_{n\in\mathbb{N}}$ by $X_n = F_n(X_{n-1})$ for $n \in \mathbb{N}$. See Figs. 18.5 and 18.6 for computer simulations of equilibrium states and metastable states of the Ising model. ◊

Gibbs Sampler

We consider a situation where, as in the above example, a state consists of many components $x = (x_i)_{i\in\Lambda} \in E$ and where Λ is a finite set. As an alternative to the Metropolis chain, we consider a different procedure to establish a Markov chain with a given invariant distribution. For the so-called *Gibbs sampler* or *heat bath algorithm*, the idea is to adapt the state *locally* to the stationary distribution. If x is a state and $i \in \Lambda$, then define

$$x_{-i} := \big\{y \in E : y(j) = x(j) \text{ for } j \neq i\big\}.$$

Definition 18.17 (Gibbs sampler) Let $q \in \mathcal{M}_1(\Lambda)$ with $q(i) > 0$ for every $i \in \Lambda$. The transition matrix p on E with

$$p(x, y) = \begin{cases} q_i \frac{\pi(x^{i,\sigma})}{\pi(x_{-i})}, & \text{if } y = x^{i,\sigma} \text{ for some } i \in \Lambda, \\ 0, & \text{else,} \end{cases}$$

is called a *Gibbs sampler* for the invariant distribution π.

Verbally, each step of the chain with transition matrix p can be described by the following instructions.

(1) Choose a random coordinate I according to some distribution $(q_i)_{i\in\Lambda}$.
(2) With probability $\pi(x^{I,\sigma})/\pi(x_{-I})$, replace x by $x^{I,\sigma}$.

If $I = i$, then the new state has the distribution $\mathcal{L}(X|X_{-i} = x_{-i})$, where X is a random variable with distribution π. Note that, for the Gibbs sampler also it is enough to know the values of the distribution π only up to the normalising constant. (In a more general framework, the Gibbs sampler and the Metropolis algorithm can be understood as special cases of one and the same method.) For states x and y that differ only in the ith coordinate, we have (since $x_{-i} = y_{-i}$)

$$\pi(x)\,p(x, y) = \pi(x) q_i \frac{\pi(y)}{\pi(x_{-i})} = \pi(y) q_i \frac{\pi(x)}{\pi(y_{-i})} = \pi(y)\,p(y, x).$$

Thus the Gibbs sampler is a reversible Markov chain with invariant measure π. Irreducibility of the Gibbs sampler, however, has to be checked for each case.

Example 18.18 (Ising model) In the Ising model described above, we have $x_{-i} = \{x^{i,-1}, x^{i,+1}\}$. Hence, for $i \in \Lambda$ and $\sigma \in \{-1, +1\}$,

$$\begin{aligned}\pi\big(x^{i,\sigma}|x_{-i}\big) &= \frac{\pi(x^{i,\sigma})}{\pi(\{x^{i,-1}, x^{i,+1}\})} \\ &= \frac{e^{-\beta H(x^{i,\sigma})}}{e^{-\beta H(x^{i,-1})} + e^{-\beta H(x^{i,+1})}} \\ &= \big(1 + \exp\big[\beta\big(H\big(x^{i,\sigma}\big) - H\big(x^{i,-\sigma}\big)\big)\big]\big)^{-1} \\ &= \left(1 + \exp\left[2\beta \sum_{j:j\sim i} \left(\mathbb{1}_{\{x(j)\neq\sigma\}} - \frac{1}{2}\right)\right]\right)^{-1}.\end{aligned}$$

The Gibbs sampler for the Ising model is thus the Markov chain $(X_n)_{n\in\mathbb{N}_0}$ with values in $E = \{-1, 1\}^\Lambda$ and with transition matrix

$$p(x, y) = \begin{cases} \frac{1}{\#\Lambda}\big(1 + \exp\big[2\beta \sum_{j:j\sim i}\big(\mathbb{1}_{\{x(j)\neq x(i)\}} - \frac{1}{2}\big)\big]\big)^{-1}, \\ \quad \text{if } y = x^i \text{ for some } i \in \Lambda, \\ 0, \quad \text{otherwise.} \end{cases}$$ ◊

Perfect Sampling

The MCMC method as described above is based on hope: We let the chain run for a long time and hope that its distribution is close to the invariant distribution. Even if we can compute the speed of convergence (and in many cases, this is not trivial, we come back to this point in Section 18.4), the distribution will never be *exactly* the invariant distribution.

Although this flaw might seem inevitable in the MCMC method, it is in fact, at least theoretically, possible to use a very similar method that allows *perfect sampling* according to the invariant distribution π, even if we do not know anything about the speed of convergence. The idea is simple. Assume that $F_1, F_2, \ldots$ are i.i.d. random maps $E \to E$ with $\mathbf{P}[F(x) = y] = p(x, y)$ for all $x, y \in E$. We have seen how to construct the Markov chain X with initial value $X_0 = x$ by defining $X_n = F_n \circ F_{n-1} \circ \ldots \circ F_1(x)$.

Note that $F_1^n(x) := F_1 \circ \ldots \circ F_n(x) \stackrel{\mathcal{D}}{=} F_n \circ \ldots \circ F_1(x)$. Hence we have

$$\mathbf{P}\big[F_1^n(x) = y\big] \stackrel{n\to\infty}{\longrightarrow} \pi(y) \quad \text{for every } y.$$

However, if F_1^n turns out to be a constant map (e.g., $F_1^n \equiv x^*$ for some random x^*), then we will also have $F_1^m \equiv x^*$ for all $m \geq n$. If by some clever choice

of the distribution of F_n one can ensure that the stopping time $T := \inf\{n \in \mathbb{N} : F_1^n \text{ is constant}\}$ is almost surely finite (and this is always possible), then we will have $\mathbf{P}[F_1^T(x) = y] = \pi(y)$ for all $x, y \in E$. A simple algorithm for this method is the following.

(1) Let $F \leftarrow \mathrm{id}_E$ and $n \leftarrow 0$.
(2) Let $n \leftarrow n + 1$. Generate F_n and let $F \leftarrow F \circ F_n$.
(3) If F is not a constant map, then go to (2).
(4) Output $F(*)$.

This method is called *coupling from the past* and goes back to Propp and Wilson [138] (see also [55, 56, 92, 137, 139, 170]). David Wilson has nice simulations and a survey of the current research on his web site http://www.dbwilson.com/. A nice survey on MCMC methods including coupling from the past is [66].

For a practical implementation, there are two main problems: (1) The full map F_n has to be generated and has to be composed with F. The computer time needed for this is at least of the order of the size of the space E. (2) Checking if F is constant needs computer time of the same order of magnitude. Consequently, the method can be efficiently implemented only if there is more structure. For example, assume that E is partially ordered with a smallest element $\underline{0}$ and a largest element $\underline{1}$ (like the Ising model). Further, assume that the maps F_n can be chosen to be almost surely monotone increasing. In this case, it is enough to compute at each step $F(\underline{0})$ and $F(\underline{1})$ since F is constant if the values coincide.

18.4 Speed of Convergence

So far we have ignored the question of the speed of convergence of the distribution $\mathbf{P}_{X_n}$ to π. For practical purposes, however, this is often the most interesting question. We do not intend to go into the details and we only briefly touch upon the topic. Without loss of generality, assume $E = \{1, \ldots, N\}$. If p is reversible (Eq. (18.8)), then $f \mapsto pf$ defines a symmetric linear operator on $L^2(E, \pi)$ (exercise!). All eigenvalues $\lambda_1, \ldots, \lambda_N$ (listed according to the corresponding multiplicity) are real and have modulus at most 1 since p is stochastic. Thus we can arrange the eigenvalues by decreasing modulus $\lambda_1 = 1 \geq |\lambda_2| \geq \ldots \geq |\lambda_N|$. If p is irreducible and aperiodic, then $|\lambda_2| < 1$. Let $\mu_1 = \pi, \mu_2, \ldots, \mu_N$ be an orthonormal basis of left eigenvectors for the eigenvalues $\lambda_1, \ldots, \lambda_N$. Then, for every $\mu = \alpha_1\mu_1 + \ldots + \alpha_N\mu_N$, we have $\mu p^n = \sum_{i=1}^N \lambda_i^n \alpha_i \mu_i$ and hence

$$\left\| \mu p^n - \pi \right\|_{TV} \leq C|\lambda_2|^n \tag{18.10}$$

for a constant C (that does not depend on μ). A similar formula holds if p is not reversible; however, with a correction term of order at most n^{V-1}. Here, V is the size of the largest Jordan block square matrix for the eigenvalue λ_2 in the Jordan canonical form of p. In particular, V is no larger than the multiplicity of the eigenvalue with second largest modulus.

The speed of convergence is thus exponential with a rate that is determined by the *spectral gap* $1 - |\lambda_2|$ of the second largest eigenvalue of p. In practice, for a large space E, computing the spectral gap is often extremely difficult.

Example 18.19 Let $r \in (0,1)$ and $N \in \mathbb{N}$, $N \geq 2$. Further, let $E = \{0, \ldots, N-1\}$. We consider the transition matrix

$$p(i,j) = \begin{cases} r, & \text{if } j = i+1 \pmod N, \\ 1-r, & \text{if } j = i-1 \pmod N, \\ 0, & \text{else.} \end{cases}$$

p is the transition matrix of simple (asymmetric) random walk on the discrete torus $\mathbb{Z}/(N)$, which with probability r makes a jump to the right and with probability $1-r$ makes a jump to the left. Clearly, p is irreducible, and p is aperiodic if and only if N is odd. Furthermore, the uniform distribution $\mathcal{U}_E$ is the unique invariant distribution.

Case 1: *N odd*. Let $\theta_k = e^{2\pi i k/N}$, $k = 0, \ldots, N-1$, be the Nth roots of unity and let the corresponding (right) eigenvectors be

$$x^k := \left(\theta_k^0, \theta_k^1, \ldots, \theta_k^{N-1}\right).$$

It is easy to check that p has the eigenvalues

$$\lambda_k := r\theta_k + (1-r)\overline{\theta}_k = \cos\left(\frac{2\pi k}{N}\right) + (2r-1)i \sin\left(\frac{2\pi k}{N}\right),$$
$$k = 0, \ldots, N-1.$$

The moduli of the eigenvalues are given by $|\lambda_k| = f(2\pi k/N)$, where

$$f(\vartheta) = \sqrt{1 - 4r(1-r)\sin(\vartheta)^2} \quad \text{for } \vartheta \in \mathbb{R}.$$

Since N is odd, $|\lambda_k|$ is maximal (except for $k = 0$) for $k = \frac{N-1}{2}$ and for $k = \frac{N+1}{2}$. For these k, $|\lambda_k|$ equals $\gamma := \sqrt{1 - 4r(1-r)\sin(\pi/N)^2}$. Since all eigenvalues are different, every eigenvalue has multiplicity 1. Hence there is a constant $C < \infty$ such that

$$\left\|\mu p^n - \mathcal{U}_E\right\|_{TV} \leq C\gamma^n \quad \text{for all } n \in \mathbb{N}, \mu \in \mathcal{M}_1(E).$$

Case 2: *N even*. In this case, p is not aperiodic. Nevertheless, the eigenvalues and eigenvectors are of the same form as in Case 1. In order to get an aperiodic chain, for $\varepsilon > 0$, define the transition matrix

$$p_\varepsilon := (1-\varepsilon)p + \varepsilon I,$$

where I is the unit matrix on E. p_ε describes the random walk on E that with probability ε does not move and with probability $1-\varepsilon$ makes a jump according

to p. Clearly, p_ε is irreducible and aperiodic. The eigenvalues are

$$\lambda_{\varepsilon,k} = (1-\varepsilon)\lambda_k + \varepsilon, \quad k = 0, \ldots, N-1,$$

and the corresponding eigenvectors are the x^k from above. Evidently, $\lambda_{\varepsilon,0} = 1$, and if $\varepsilon > 0$ is very small, then $\lambda_{\varepsilon,N/2} = 2\varepsilon - 1$ is the eigenvalue with the second largest modulus. For larger values of ε, we have $|\lambda_{\varepsilon,1}| > |\lambda_{\varepsilon,N/2}|$. More precisely, if we let

$$\varepsilon_0 := \frac{(1-(2r-1)^2)\sin(2\pi/N)^2}{(1-(2r-1)^2)\sin(2\pi/N)^2 + 2\cos(2\pi/N)},$$

then the eigenvalue with the second largest modulus has modulus

$$\gamma_\varepsilon = |\lambda_{\varepsilon,N/2}| = 1 - 2\varepsilon, \quad \text{if } \varepsilon \le \varepsilon_0,$$

or

$$\gamma_\varepsilon = |\lambda_{\varepsilon,1}| = \sqrt{\left((1-\varepsilon)\cos\left(\frac{2\pi}{N}\right) + \varepsilon\right)^2 + \left((1-\varepsilon)(2r-1)\sin\left(\frac{2\pi}{N}\right)\right)^2}, \quad \text{if } \varepsilon \ge \varepsilon_0.$$

It is easy to check that $\varepsilon \mapsto |\lambda_{\varepsilon,N/2}|$ is monotone decreasing and that $\varepsilon \mapsto |\lambda_{\varepsilon,1}|$ is monotone increasing. Hence γ_ε is minimal for $\varepsilon = \varepsilon_0$.

Hence there is a $C < \infty$ with

$$\left\|\mu p_\varepsilon^n - \mathcal{U}_E\right\|_{TV} \le C\gamma_\varepsilon^n \quad \text{for all } n \in \mathbb{N}, \mu \in \mathcal{M}_1(E),$$

and the best speed of convergence (in this class of transition matrices) can be obtained by choosing $\varepsilon = \varepsilon_0$. ♢

Example 18.20 (Gambler's ruin) We consider the gambler's ruin problem from Example 10.19 with the probability of a gain $r \in (0,1)$. Here the state space is $E = \{0, \ldots, N\}$, and the transition matrix is of the form

$$p(i,j) = \begin{cases} r, & \text{if } j = i+1 \in \{2, \ldots, N\}, \\ 1-r, & \text{if } j = i-1 \in \{0, \ldots, N-2\}, \\ 1, & \text{if } j = i \in \{0, N\}, \\ 0, & \text{else.} \end{cases}$$

This transition matrix is not irreducible; rather it has two absorbing states 0 and N. In Example 10.19 (Eq. (10.5)) for the case $r \neq \frac{1}{2}$, and Example 10.16 for the case $r = \frac{1}{2}$, it was shown that, for every $\mu \in \mathcal{M}_1(E)$,

$$\mu p^n \overset{n\to\infty}{\longrightarrow} \big(1 - m(\mu)\big)\delta_0 + m(\mu)\delta_N. \tag{18.11}$$

Here $m(\mu) = \int p_N(x)\mu(dx)$, where the probability $p_N(x)$ that the chain, if started at x, hits N is given by

$$p_N(x) = \begin{cases} \frac{1-(\frac{1-r}{r})^x}{1-(\frac{1-r}{r})^N}, & \text{if } r \neq \frac{1}{2}, \\ \frac{x}{N}, & \text{if } r = \frac{1}{2}. \end{cases}$$

How quick is the convergence in (18.11)? Here also the convergence has exponential speed and the rate is determined by the second largest eigenvalue of p.

Hence we have to compute the spectrum of p. Clearly, $x^0 = (1, 0, \ldots, 0)$ and $x^N = (0, \ldots, 0, 1)$ are left eigenvectors for the eigenvalue 1. In order for $x = (x_0, \ldots, x_N)$ to be a left eigenvector for the eigenvalue λ, the following equations have to hold:

$$\lambda x_k = r x_{k-1} + (1-r)x_{k+1} \quad \text{for } k = 2, \ldots, N-2, \tag{18.12}$$

and

$$\lambda x_{N-1} = r x_{N-2}. \tag{18.13}$$

If (18.12) and (18.13) hold for $x_1, \ldots, x_{N-1}$, then we define $x_0 := \frac{1-p}{\lambda-1}x_1$ and $x_N := \frac{p}{\lambda-1}x_{N-1}$ and get that in fact $xp = \lambda x$. We make the ansatz

$$\lambda = (1-r)\rho(\theta + \overline{\theta}) \quad \text{and} \quad x_k = \varrho^k\big(\theta^k - \overline{\theta}^k\big) \quad \text{for } k = 1, \ldots, N-1,$$

where

$$\rho = \sqrt{r/(1-r)} \quad \text{and} \quad \theta \in \mathbb{C} \setminus \{-1, +1\} \quad \text{with } |\theta| = 1.$$

Thus we have $\theta\overline{\theta} = 1$ and $(1-r)\rho^{k+1} = r\rho^{k-1}$. Therefore, for every $k = 2, \ldots, N-1$,

$$\begin{aligned} \lambda x_k &= (1-r)\rho^{k+1}\big(\theta^k - \overline{\theta}^k\big)(\theta + \overline{\theta}) \\ &= (1-r)\rho^{k+1}\big[\big(\theta^{k+1} - \overline{\theta}^{k+1}\big) + \theta\overline{\theta}\big(\theta^{k-1} - \overline{\theta}^{k-1}\big)\big] \\ &= r\rho^{k-1}\big(\theta^{k-1} - \overline{\theta}^{k-1}\big) + (1-r)\rho^{k+1}\big(\theta^{k+1} - \overline{\theta}^{k+1}\big) \\ &= r x_{k-1} + (1-r)x_{k+1}. \end{aligned}$$

That is, (18.12) holds. The same computation with $k = N-1$ shows that (18.13) holds if and only if $\theta^N - \overline{\theta}^N = 0$; that is, if $\theta^{2N} = 1$. In all, then, for θ, we get $N-1$ different values (note that the complex conjugates of the values considered here lead to the same values λ_n),

$$\theta_n = e^{(n/N)\pi i} \quad \text{for } n = 1, \ldots, N-1.$$

The corresponding eigenvalues are

$$\lambda_n = \sigma \cos\left(\frac{n\pi}{N}\right) \quad \text{for } n = 1, \ldots, N-1.$$

Here the variance of the individual random walk step is

$$\sigma^2 := 4r(1-r). \tag{18.14}$$

As all eigenvalues are real, the corresponding eigenvectors are given by

$$x_k^n = 2\left(\frac{r}{1-r}\right)^{n/2} \sin\left(\frac{n\pi}{N}\right), \quad k = 1, \ldots, N-1.$$

The second largest modulus of an eigenvalue is $|\lambda_n| = \sigma\cos(\frac{\pi}{N})$ if $n = 1$ or $n = N-1$. Thus there exists a $C > 0$ such that, for every $\mu \in \mathcal{M}_1(E)$, we have

$$\mu p^n\big(\{1, \ldots, N-1\}\big) \le C\left(\sigma\cos\left(\frac{\pi}{N}\right)\right)^n \quad \text{for every } n \in \mathbb{N}.$$

In other words, the probability that the game has not finished up to the nth round is at most $C(\sigma\cos(\pi/N))^n$.

An alternative approach to the eigenvalues can be made via the roots of the characteristic polynomial

$$\chi_N(x) = \det(p - xI), \quad x \in \mathbb{R}.$$

Clearly, $\chi_1(x) = (1-x)^2$ and $\chi_2(x) = -x(1-x)^2$. Using Laplace's expansion formula for the determinant (elimination of rows and columns), we get the recursion

$$\chi_N(x) = -x\chi_{N-1}(x) - r(1-r)\chi_{N-2}(x). \tag{18.15}$$

The solution is (check this!)

$$\chi_N(x) = (-1)^{N-1}(\sigma/2)^{N-1}(1-x)^2 U_{N-1}(x/\sigma), \tag{18.16}$$

where

$$U_m(x) := \sum_{k=0}^{\lfloor m/2\rfloor} (-1)^k \binom{m-k}{k} (2x)^{m-2k}$$

denotes the so-called mth *Chebyshev polynomial* of the second kind.

Using de Moivre's formula, one can show that, for $x \in (-\sigma, \sigma)$,

$$\begin{aligned}\chi_N(x) &= (-1)^{N-1}(\sigma/2)^{N-1}(1-x)^2 \frac{\sin(N\arccos(x/\sigma))}{\sqrt{1-(x/\sigma)^2}} \\ &= (1-x)^2 \prod_{k=1}^{N-1}\left(\sigma\cos\left(\frac{\pi k}{N}\right) - x\right).\end{aligned} \tag{18.17}$$

Apart from the double zero at 1, we get the zeros

$$\sigma\cos(\pi k/N), \quad k = 1, \ldots, N-1. \qquad \Diamond$$

Exercise 18.4.1 Show (18.16).

Exercise 18.4.2 Show (18.17).

Exercise 18.4.3 Let $\nu(dx) = \frac{2}{\pi}\sqrt{1-x^2}\mathbb{1}_{[-1,1]}(x)\,dx$. Show that the Chebyshev polynomials of the second kind are orthonormal with respect to ν; that is,

$$\int U_m U_n \, d\nu = \mathbb{1}_{\{m=n\}}.$$

Exercise 18.4.4 Let

$$E = \{1, 2, 3\} \quad \text{and} \quad p = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/3 & 1/3 & 1/3 \\ 0 & 3/4 & 1/4 \end{pmatrix}.$$

Compute the invariant distribution and the exponential rate of convergence.

Exercise 18.4.5 Let $E = \{0, \dots, N-1\}$, $r \in (0, 1)$ and

$$p(i, j) = \begin{cases} r, & \text{if } j = i + 1 \pmod N, \\ 1 - r, & \text{if } j = i \pmod N, \\ 0, & \text{else.} \end{cases}$$

Show that p is the transition matrix of an irreducible, aperiodic random walk and compute the invariant distribution and the exponential rate of convergence.

Exercise 18.4.6 Let $N \in \mathbb{N}$ and let $E = \{0, 1\}^N$ denote the N-dimensional hypercube. That is, two points $x, y \in E$ are connected by an edge if they differ in exactly one coordinate. Let p be the transition matrix of the random walk on E that stays put with probability $\varepsilon > 0$ and that with probability $1 - \varepsilon$ makes a jump to a randomly (uniformly) chosen neighboring site.

Describe p formally and show that p is aperiodic and irreducible. Compute the invariant distribution and the exponential rate of convergence.