

19.6 Markov Chains

Chapter 9 §11 proved that a Markov chain settles down to a unique steady state, but only under three hypotheses on the eigenvalues of its transition matrix, one of them diagonalizability. It closed with a promise: the Perron–Frobenius theorem would give the same conclusions “for every stochastic matrix with some power having all entries positive, without assuming diagonalizability and without computing eigenvalues”. This section keeps that promise, using the limit theorem of §05. It then asks what happens when no power is positive. Irreducible chains still have a unique steady state, and they approach it on average even when they oscillate. Reducible chains can have many steady states, and among them the absorbing chains, which end in a trap, have an exact theory built on the Neumann series.

Conventions. Stochastic means **column**-stochastic, as in Definition 9.11.1: $\mathbf{A} \geq \mathbf{0}$ and every column sums to 1, that is, $\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top$. The entry a_{ij} is the probability of moving **to** state i **from** state j , and a Markov chain is $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$ (Definition 9.11.2). Many books use rows and the transposed matrix; we never do. The order $\geq$ is entrywise.

19.6.1 Steady states and the Perron root

We first place stochastic matrices inside the theory of this chapter. Since $\mathbf{x}_m = \mathbf{A}^m \mathbf{x}_0$, the entry $(\mathbf{A}^m)_{ij}$ is the probability of being in state i after m steps when starting from state j . By Lemma 19.2.1 (a) it is positive exactly when the graph $G(\mathbf{A})$ has a walk of length m from j to i : the edges of the graph are the moves the chain can make in one step, and the walks are the routes it can take.

Proposition 19.6.1. The Perron Root of a Stochastic Matrix

Let $\mathbf{A} \in M_n(\mathbb{R})$ be stochastic. Then:

- (a) $\rho(\mathbf{A}) = 1$, and $\mathbf{1}$ is a positive left eigenvector for it;
- (b) the steady states of $\mathbf{A}$ are exactly the nonnegative eigenvectors for $\rho(\mathbf{A})$ whose entries sum to 1;
- (c) if $\mathbf{A}$ is irreducible, then $\mathbf{A}$ has **exactly one** steady state $\mathbf{v}$, and $\mathbf{v} > \mathbf{0}$.

Idea. Example (An irreducible stochastic matrix) already found these facts for one matrix, and the argument there is general. The column sums make $\mathbf{1}^\top$ a left eigenvector for 1, and no eigenvalue is larger in modulus; so a steady state is a nonnegative eigenvector for the Perron root; when $\mathbf{A}$ is irreducible, Perron–Frobenius with the normalization $\mathbf{1}^\top \mathbf{v} = 1$ leaves room for only one.

Proof.

- (a) By Proposition 9.11.3 (c), $1 \in \text{spec}(\mathbf{A})$ and $\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top$. By Corollary 10.10.5 (a), every eigenvalue has modulus at most 1. So $\rho(\mathbf{A}) = 1$.
- (b) This is Definition 9.11.2 read with (a): a steady state is a probability vector $\mathbf{v}$ with $\mathbf{A}\mathbf{v} = \mathbf{v} = \rho(\mathbf{A})\mathbf{v}$.
- (c) By Theorem 19.3.2 (a), there is $\mathbf{u} > \mathbf{0}$ with $\mathbf{A}\mathbf{u} = \mathbf{u}$; then $\mathbf{v} = \mathbf{u}/(\mathbf{1}^\top \mathbf{u})$ is a steady state with $\mathbf{v} > \mathbf{0}$. If $\mathbf{v}'$ is any steady state, it is a nonnegative eigenvector, so $\mathbf{v}' = c\mathbf{v}$

with $c > 0$ by Theorem 19.3.2 (c), and $1 = \mathbf{1}^\top \mathbf{v}' = c \mathbf{1}^\top \mathbf{v} = c$. Hence $\mathbf{v}' = \mathbf{v}$. ■

In the notation of §03, the Perron root of a stochastic matrix is 1, the all-ones vector is a positive left eigenvector, and for an irreducible stochastic matrix the steady state is the right Perron vector.

Warning. **A unique positive steady state does not make the chain converge.** The swap $\mathbf{S} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is stochastic and irreducible, so by (c) it has exactly one steady state, $(\frac{1}{2}, \frac{1}{2})$, and it is positive. Yet the chain started at $\mathbf{e}_1$ alternates $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1, \dots$ and never approaches it. Existence and uniqueness of the steady state come from irreducibility; convergence needs more.

19.6.2 The long-run theorem

The “more” is primitivity. Here are the three conclusions of Theorem 9.11.4, in the same words, under the hypothesis that Chapter 9 §11 promised.

Theorem 19.6.2. Long-Run Behavior of a Markov Chain

Let $\mathbf{A} \in M_n(\mathbb{R})$ be stochastic, and suppose that $\mathbf{A}^k > \mathbf{O}$ for some $k \geq 1$. Then:

- (a) $\mathbf{A}$ has exactly one steady state $\mathbf{v}$, and $\mathbf{v} > \mathbf{0}$;
- (b) $\mathbf{A}^m \rightarrow \mathbf{v} \mathbf{1}^\top$ entrywise as $m \rightarrow \infty$;
- (c) for every probability vector $\mathbf{x}_0$, the Markov chain $\mathbf{x}_m = \mathbf{A}^m \mathbf{x}_0$ converges entrywise to $\mathbf{v}$.

Idea. For a stochastic matrix, Theorem 19.5.2 has $\rho = 1$, and the left eigenvector can be taken to be $\mathbf{1}$, so the limit $\mathbf{v} \mathbf{w}^\top / (\mathbf{w}^\top \mathbf{v})$ becomes $\mathbf{v} \mathbf{1}^\top$ once the steady state is normalized by $\mathbf{1}^\top \mathbf{v} = 1$. The chain inherits the limit column by column.

Proof. By hypothesis $\mathbf{A}$ is primitive (Definition 19.5.1), hence irreducible (Proposition 19.5.1 (b)). So (a) is Proposition 19.6.1 (c).

- (b) By Proposition 19.6.1 (a), $\rho(\mathbf{A}) = 1$ and $\mathbf{w} = \mathbf{1}$ is positive with $\mathbf{w}^\top \mathbf{A} = \mathbf{w}^\top$; and $\mathbf{v} > \mathbf{0}$ satisfies $\mathbf{A} \mathbf{v} = \mathbf{v}$. By Theorem 19.5.2,

$$\mathbf{A}^m = \left(\frac{1}{\rho(\mathbf{A})} \mathbf{A} \right)^m \rightarrow \frac{\mathbf{v} \mathbf{1}^\top}{\mathbf{1}^\top \mathbf{v}} = \mathbf{v} \mathbf{1}^\top,$$

since $\mathbf{1}^\top \mathbf{v} = 1$.

- (c) Entry r of $\mathbf{A}^m \mathbf{x}_0$ is $\sum_s (\mathbf{A}^m)_{rs} (\mathbf{x}_0)_s$, which by (b) converges to $\sum_s v_r (\mathbf{x}_0)_s = v_r (\mathbf{1}^\top \mathbf{x}_0) = v_r$. ■

Compare the hypotheses with those of Theorem 9.11.4. That theorem asked for three facts about the eigenvalues: $\mathbf{A}$ diagonalizable over $\mathbb{C}$, $a_{\mathbf{A}}(1) = 1$, and $|\lambda| < 1$ for every other eigenvalue. Chapter 10 §10 (Corollary 10.10.5) removed the first, but still asked to locate every eigenvalue on the unit circle. The new hypothesis is a statement about the **zero pattern** of one power of $\mathbf{A}$, checked by multiplying matrices; no eigenvalue is computed, and none needs to be. The last two eigenvalue conditions become conclusions, by Proposition 19.5.1 (c) and Theorem 19.3.2 (b), and diagonalizability is simply irrelevant. By Corollary 19.5.5 the hypothesis can often be checked without multiplying at all: an irreducible stochastic matrix with one positive diagonal entry, a state that can stay put, qualifies. In the vocabulary of probability, a stochastic matrix with a positive power is called **regular**.

Example. A Chain that Cannot Be Diagonalized

A token moves among three positions. From position 1 it always moves to 3; from 2 it moves to 1 or 3 with probability $\frac{1}{2}$ each; from 3 it moves to 1 or 2 with probability $\frac{1}{2}$ each. So

$$\mathbf{A} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 \\ 1 & 1/2 & 0 \end{pmatrix}.$$

Show that Theorem 9.11.4 does not apply but [Theorem 19.6.2](#) does, and find the long-run distribution.

Solution. Diagonalization fails. Expanding $\det(x\mathbf{I} - \mathbf{A})$ along the first column,

$$\begin{aligned} p_{\mathbf{A}} &= x \left(x^2 - \frac{1}{4} \right) - \left(\frac{1}{4} + \frac{x}{2} \right) \\ &= x^3 - \frac{3}{4}x - \frac{1}{4} = (x-1) \left(x + \frac{1}{2} \right)^2. \end{aligned}$$

For the double eigenvalue, $\mathbf{A} + \frac{1}{2}\mathbf{I} = \begin{pmatrix} 1/2 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 1 & 1/2 & 1/2 \end{pmatrix}$ has rank 2: its first two rows are

independent, and the third is twice the first minus the second. So $g_{\mathbf{A}}(-\frac{1}{2}) = 1 < 2 = a_{\mathbf{A}}(-\frac{1}{2})$, and $\mathbf{A}$ is not diagonalizable (Theorem 9.4.1). Theorem 9.11.4 does not apply.

A positive power. Squaring,

$$\mathbf{A}^2 = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 0 \\ 0 & 1/2 & 3/4 \end{pmatrix}, \quad \mathbf{A}^4 = (\mathbf{A}^2)^2 = \begin{pmatrix} 3/8 & 5/16 & 5/16 \\ 3/8 & 3/16 & 1/8 \\ 1/4 & 1/2 & 9/16 \end{pmatrix} > \mathbf{O}.$$

So [Theorem 19.6.2](#) applies with $k = 4$. ($\mathbf{A}^3$ still has a zero at position (2, 1): three steps cannot take the token from 1 to 2.)

The steady state. $(\mathbf{A} - \mathbf{I})\mathbf{v} = \mathbf{0}$ reads $-v_1 + \frac{1}{2}v_2 + \frac{1}{2}v_3 = 0$ and $-v_2 + \frac{1}{2}v_3 = 0$. So $v_3 = 2v_2$, then $v_1 = \frac{1}{2}v_2 + v_2 = \frac{3}{2}v_2$, and $\mathbf{v}$ is a multiple of $(3, 2, 4)$. Scaling to sum 1,

$$\mathbf{v} = \left(\frac{1}{3}, \frac{2}{9}, \frac{4}{9} \right), \quad \mathbf{A}^m \longrightarrow \mathbf{v}\mathbf{1}^\top.$$

Check: $\mathbf{A}\mathbf{v} = (\frac{1}{9} + \frac{2}{9}, \frac{2}{9}, \frac{1}{3} + \frac{1}{9}) = (\frac{1}{3}, \frac{2}{9}, \frac{4}{9})$. In the long run the token is at position 3 four times out of nine, wherever it starts.

What the Jordan block does. Solving the recurrence exactly (for instance from the Jordan form, as in Chapter 10 §10) gives

$$\begin{aligned} (\mathbf{A}^m)_{11} &= \frac{1}{3} \left(1 + 2 \left(-\frac{1}{2} \right)^m \right), \\ (\mathbf{A}^m)_{21} &= \frac{2}{9} \left(1 + (3m-1) \left(-\frac{1}{2} \right)^m \right), \end{aligned}$$

which one can confirm for $m = 0, 1, 2$ against $\mathbf{I}, \mathbf{A}, \mathbf{A}^2$. The factor m in the second formula comes from the 2×2 Jordan block for $-\frac{1}{2}$, whose powers carry the entry $m\lambda^{m-1}$ (Chapter 10 §10): the error decays like $m 2^{-m}$ rather than 2^{-m} . It still decays, and the theorem never needed to know.

Quick check. Does [Theorem 19.6.2](#) apply to $\mathbf{A} = \begin{pmatrix} 1/2 & 1 \\ 1/2 & 0 \end{pmatrix}$? If so, what is $\lim_m \mathbf{A}^m$?

Solution. Yes: $\mathbf{A}$ is stochastic and $\mathbf{A}^2 = \begin{pmatrix} 3/4 & 1/2 \\ 1/4 & 1/2 \end{pmatrix} > \mathbf{O}$. The steady state solves $-\frac{1}{2}v_1 + v_2 = 0$, so $\mathbf{v} = (\frac{2}{3}, \frac{1}{3})$, and $\mathbf{A}^m \rightarrow \mathbf{v}\mathbf{1}^\top = \begin{pmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{pmatrix}$.

The hypothesis of [Theorem 19.6.2](#) is also necessary for the theorem as stated: if $\mathbf{A}^m \rightarrow \mathbf{v}\mathbf{1}^\top$ with $\mathbf{v} > \mathbf{0}$, then $\mathbf{A}$ is primitive (exercise C2). Without the positivity of $\mathbf{v}$ it is not needed for convergence: $\begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix}$ has no positive power, yet its powers converge to $\mathbf{e}_1\mathbf{1}^\top$, whose steady state has a zero entry.

19.6.3 Irreducible chains: convergence on average

Now drop primitivity but keep irreducibility. By [Proposition 19.6.1](#) (c) the steady state $\mathbf{v}$ is still unique and positive, but by Theorem 19.5.3 and Theorem 19.5.4 the chain has $h \geq 2$ eigenvalues on the unit circle, and its powers do not converge (Corollary 10.10.5 (b)). What survives is convergence of **averages**. The average $\frac{1}{m}(\mathbf{x}_0 + \mathbf{x}_1 + \cdots + \mathbf{x}_{m-1})$ is, entry by entry, the average over the first m steps of the probability of being in each state; read through linearity of expectation (see the end of this section), it is the expected fraction of those steps spent there. It is this that settles down.

Theorem 19.6.3. Averages of an Irreducible Chain

Let $\mathbf{A} \in M_n(\mathbb{R})$ be stochastic and irreducible, with steady state $\mathbf{v}$. Then:

- (a) $\frac{1}{m}(\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots + \mathbf{A}^{m-1}) \rightarrow \mathbf{v}\mathbf{1}^\top$ as $m \rightarrow \infty$;
- (b) for every probability vector $\mathbf{x}_0$, the averages $\frac{1}{m}(\mathbf{x}_0 + \mathbf{x}_1 + \cdots + \mathbf{x}_{m-1})$ of the chain converge to $\mathbf{v}$;

(c) $\mathbf{A}^m$ converges if and only if $\mathbf{A}$ is primitive.

Idea. Split every vector into its component along $\mathbf{v}$ and a remainder whose entries sum to 0. The remainder lies in the image of $\mathbf{I} - \mathbf{A}$, by a dimension count, and on that image the sum $\sum_{k < m} \mathbf{A}^k$ telescopes: $\sum_{k < m} \mathbf{A}^k (\mathbf{I} - \mathbf{A}) \mathbf{y} = \mathbf{y} - \mathbf{A}^m \mathbf{y}$, which stays bounded, so dividing by m kills it. The component along $\mathbf{v}$ is fixed by every power.

Proof.

- (a) Let $U = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{1}^\top \mathbf{x} = 0\}$, the kernel of the nonzero linear functional $\mathbf{x} \mapsto \mathbf{1}^\top \mathbf{x}$, so $\dim U = n - 1$ by Rank–Nullity (Theorem 2.3.1). Since $\mathbf{1}^\top (\mathbf{I} - \mathbf{A}) = \mathbf{1}^\top - \mathbf{1}^\top = \mathbf{0}^\top$, the image of $\mathbf{I} - \mathbf{A}$ lies in U . The kernel of $\mathbf{I} - \mathbf{A}$ on $\mathbb{R}^n$ is the real eigenspace of $\mathbf{A}$ for 1; it contains $\mathbf{v} \neq \mathbf{0}$, and it has dimension at most $g_{\mathbf{A}}(1) \leq a_{\mathbf{A}}(1) = 1$ by Theorem 9.3.1 and Theorem 19.3.2 (b). (The real and the complex kernel of the real matrix $\mathbf{I} - \mathbf{A}$ have the same dimension, n minus its rank, since row reduction uses only real arithmetic.) So $\text{rank}(\mathbf{I} - \mathbf{A}) = n - 1$ by Rank–Nullity, and an inclusion of subspaces of equal dimension is an equality: $\text{im}(\mathbf{I} - \mathbf{A}) = U$.

Now fix $\mathbf{x} \in \mathbb{R}^n$. Since $\mathbf{1}^\top \mathbf{v} = 1$, the vector $\mathbf{x} - (\mathbf{1}^\top \mathbf{x})\mathbf{v}$ lies in U , so $\mathbf{x} = (\mathbf{1}^\top \mathbf{x})\mathbf{v} + (\mathbf{I} - \mathbf{A})\mathbf{y}$ for some $\mathbf{y} \in \mathbb{R}^n$. As $\mathbf{A}^k \mathbf{v} = \mathbf{v}$ for every k ,

$$\begin{aligned} \sum_{k=0}^{m-1} \mathbf{A}^k \mathbf{x} &= m (\mathbf{1}^\top \mathbf{x}) \mathbf{v} + \sum_{k=0}^{m-1} (\mathbf{A}^k \mathbf{y} - \mathbf{A}^{k+1} \mathbf{y}) \\ &= m (\mathbf{1}^\top \mathbf{x}) \mathbf{v} + \mathbf{y} - \mathbf{A}^m \mathbf{y}. \end{aligned}$$

Every $\mathbf{A}^m$ is stochastic (Proposition 9.11.3 (b)), so its entries lie in $[0, 1]$ and $|(\mathbf{A}^m \mathbf{y})_i| \leq \sum_j |y_j|$. Hence each entry of $\frac{1}{m}(\mathbf{y} - \mathbf{A}^m \mathbf{y})$ has modulus at most $\frac{2}{m} \sum_j |y_j|$, which tends to 0 because $\frac{1}{m} \rightarrow 0$, by the Archimedean property, a consequence of completeness (fact (A1) in the introduction to Chapter 16). Dividing the display by m , $\frac{1}{m} \sum_{k < m} \mathbf{A}^k \mathbf{x} \rightarrow (\mathbf{1}^\top \mathbf{x})\mathbf{v} = \mathbf{v} \mathbf{1}^\top \mathbf{x}$. Taking $\mathbf{x} = \mathbf{e}_j$ gives the convergence of column j , for each j .

- (b) $\frac{1}{m} \sum_{k < m} \mathbf{x}_k = (\frac{1}{m} \sum_{k < m} \mathbf{A}^k) \mathbf{x}_0 \rightarrow \mathbf{v} \mathbf{1}^\top \mathbf{x}_0 = \mathbf{v}$ by (a), entry by entry as in the proof of Theorem 19.6.2 (c).
- (c) ($\Leftarrow$) is Theorem 19.6.2 (b). ($\Rightarrow$) Suppose $\mathbf{A}^m \rightarrow \mathbf{L}$. Then $\mathbf{A}^{m+1} = \mathbf{A} \mathbf{A}^m$ tends both to $\mathbf{L}$ and to $\mathbf{A} \mathbf{L}$, so $\mathbf{A} \mathbf{L} = \mathbf{L}$, and every column of $\mathbf{L}$ lies in the real eigenspace for 1, which is $\text{span}(\mathbf{v})$ by the proof of (a). Each column of $\mathbf{A}^m$ has entries ≥ 0 summing to 1, and both properties survive the limit, by the elementary algebra of limits recalled in the introduction to Chapter 16 (limits respect sums, and a nonstrict inequality survives a limit), so each column of $\mathbf{L}$ is a probability vector on the line $\text{span}(\mathbf{v})$, namely $\mathbf{v}$ itself. Thus $\mathbf{L} = \mathbf{v} \mathbf{1}^\top > \mathbf{O}$. As in the proof of Theorem 19.5.3, the n^2 entries of $\mathbf{A}^m$ are then positive for all large m , so $\mathbf{A}$ is primitive. ■

The proof of (a) used irreducibility only to know that the eigenspace for 1 is a line through a probability vector $\mathbf{v}$; nothing about the period or the other eigenvalues entered.

Example. A Walk that Oscillates

A walker on three sites 1, 2, 3 in a row steps to a neighboring site at every step, choosing each neighbor with probability $\frac{1}{2}$ when there are two. So

$$\mathbf{A} = \begin{pmatrix} 0 & 1/2 & 0 \\ 1 & 0 & 1 \\ 0 & 1/2 & 0 \end{pmatrix}.$$

Show that $\mathbf{A}^m$ does not converge, and compute the averages of [Theorem 19.6.3](#) exactly.

Solution. The edges of $G(\mathbf{A})$ are $1 \rightarrow 2$, $2 \rightarrow 1$, $2 \rightarrow 3$ and $3 \rightarrow 2$. Every site reaches every other through site 2, so $\mathbf{A}$ is irreducible (Theorem 19.2.5); and every edge joins site 2 to $\{1, 3\}$, so closed walks have even length and the period is 2. The steady state solves $v_1 = \frac{1}{2}v_2$, $v_3 = \frac{1}{2}v_2$, so $\mathbf{v} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Multiplying out,

$$\mathbf{A}^2 = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}, \quad \mathbf{A}^3 = \mathbf{A}\mathbf{A}^2 = \mathbf{A}.$$

So $\mathbf{A}^m = \mathbf{A}$ for odd m and $\mathbf{A}^m = \mathbf{A}^2$ for even $m \geq 2$, and since $\mathbf{A} \neq \mathbf{A}^2$ the powers do not converge. A walker who starts at site 2 is at site 2 at every even step and never at an odd one.

For the averages, note that $\frac{1}{2}(\mathbf{A} + \mathbf{A}^2) = \begin{pmatrix} 1/4 & 1/4 & 1/4 \\ 1/2 & 1/2 & 1/2 \\ 1/4 & 1/4 & 1/4 \end{pmatrix} = \mathbf{v}\mathbf{1}^\top$. For $m = 2p$, the sum $\sum_{k < 2p} \mathbf{A}^k$ contains $\mathbf{I}$ once, $\mathbf{A}$ p times, and $\mathbf{A}^2$ $p - 1$ times, so

$$\begin{aligned} \frac{1}{2p} \sum_{k=0}^{2p-1} \mathbf{A}^k &= \frac{\mathbf{I} - \mathbf{A}^2 + p(\mathbf{A} + \mathbf{A}^2)}{2p} \\ &= \mathbf{v}\mathbf{1}^\top + \frac{1}{2p}(\mathbf{I} - \mathbf{A}^2) \longrightarrow \mathbf{v}\mathbf{1}^\top. \end{aligned}$$

For $m = 2p + 1$ there is one more term, $\mathbf{A}^{2p} = \mathbf{A}^2$, which changes the average by at most $\frac{1}{m}$ in each entry. So the walker spends, in the long run, half its time at the middle site and a quarter at each end, although at no single time is its distribution close to $\mathbf{v}$ if it starts at a site.

The oscillation can also be removed rather than averaged. The **lazy** chain $\frac{1}{2}(\mathbf{I} + \mathbf{A})$, which at each step stays put with probability $\frac{1}{2}$ and otherwise moves by $\mathbf{A}$, is stochastic, irreducible, and has positive diagonal, so it is primitive by Corollary 19.5.5, and it has the same steady state; exercise C1 works this out.

19.6.4 Reducible chains: many steady states

Without irreducibility, uniqueness fails too. The chain can split into parts that never communicate, and each can carry its own equilibrium.

Example. Two Traps

States 1 and 2 exchange with each other and are never left; state 3 is never left either; state 4 stays with probability $\frac{1}{4}$ and otherwise moves to 1, 2 or 3 with probability $\frac{1}{4}$ each:

$$\mathbf{A} = \begin{pmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 1/2 & 3/4 & 0 & 1/4 \\ 0 & 0 & 1 & 1/4 \\ 0 & 0 & 0 & 1/4 \end{pmatrix}.$$

Find all steady states, and the limit of the chain from each starting state.

Solution. Steady states. Let $\mathbf{A}\mathbf{u} = \mathbf{u}$ with $\mathbf{u}$ a probability vector. The last row gives $\frac{1}{4}u_4 = u_4$, so $u_4 = 0$. The first row then gives $\frac{1}{2}u_1 + \frac{1}{4}u_2 = u_1$, so $u_2 = 2u_1$; the second and third rows then hold automatically. So $\mathbf{u} = (s, 2s, t, 0)$ with $s, t \geq 0$ and $3s + t = 1$. Writing $\mathbf{v}_1 = (\frac{1}{3}, \frac{2}{3}, 0, 0)$ and $\mathbf{v}_2 = \mathbf{e}_3$, the steady states are exactly

$$\mathbf{u} = c\mathbf{v}_1 + (1 - c)\mathbf{v}_2, \quad 0 \leq c \leq 1,$$

the whole segment $[\mathbf{v}_2, \mathbf{v}_1]$ of Chapter 18. There are infinitely many, and $E_1(\mathbf{A}) = \text{span}(\mathbf{v}_1, \mathbf{v}_2)$ is a plane: $a_{\mathbf{A}}(1) = 2$, since $p_{\mathbf{A}} = (x - 1)^2(x - \frac{1}{4})^2$, the first factor from the block of states 1, 2 and from state 3, the second from the block's other eigenvalue $\frac{1}{4}$ and from state 4. In particular $\mathbf{A}$ is not irreducible, by Proposition 19.6.1 (c).

Limits. The only eigenvalue of modulus 1 is 1, so $\mathbf{A}^m$ converges by Corollary 10.10.5 (b). Its limit $\mathbf{\Pi}$ satisfies $\mathbf{A}\mathbf{\Pi} = \mathbf{\Pi}$, as in the proof of Theorem 19.6.3 (c), and its columns are limits of probability vectors, hence probability vectors; so each column of $\mathbf{\Pi}$ is a steady state. Columns 1 and 2 of every $\mathbf{A}^m$ have zeros in rows 3 and 4, by induction on m : $\mathbf{A}$ maps a vector with zeros in those rows to a combination of its own first two columns, which again has zeros there. So columns 1 and 2 of $\mathbf{\Pi}$ are steady states with $u_3 = 0$, that is, $\mathbf{v}_1$. Column 3 of $\mathbf{A}$ is $\mathbf{e}_3$, so column 3 of every power, and of $\mathbf{\Pi}$, is $\mathbf{e}_3 = \mathbf{v}_2$. For column 4, write it as $c\mathbf{v}_1 + (1 - c)\mathbf{v}_2$ and find c from a conserved quantity: the row vector $\mathbf{y}^\top = (1, 1, 0, \frac{2}{3})$ satisfies $\mathbf{y}^\top \mathbf{A} = \mathbf{y}^\top$, as a column-by-column check shows (column 4: $\frac{1}{4} + \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{4} = \frac{2}{3}$). Then $\mathbf{y}^\top \mathbf{A}^m = \mathbf{y}^\top$ for all m , hence $\mathbf{y}^\top \mathbf{\Pi} = \mathbf{y}^\top$, and entry 4 gives $\mathbf{y}^\top (c\mathbf{v}_1 + (1 - c)\mathbf{v}_2) = c = \frac{2}{3}$. Thus

$$\mathbf{\Pi} = \begin{pmatrix} 1/3 & 1/3 & 0 & 2/9 \\ 2/3 & 2/3 & 0 & 4/9 \\ 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

with last column $\frac{2}{3}\mathbf{v}_1 + \frac{1}{3}\mathbf{v}_2$. The long run depends on where the chain starts: from state 4, the chain ends up in the pair $\{1, 2\}$ with probability $\frac{2}{3}$ and in state 3 with probability $\frac{1}{3}$, the ratio $\frac{1}{2} : \frac{1}{4}$ of the one-step probabilities of leaving 4 for each.

In general the steady states of a stochastic matrix form a convex set, since a convex combination of probability vectors fixed by $\mathbf{A}$ is again one; the example shows it can be a segment rather than a point. One steady state for each part of the chain that, once entered, is never left, and all convex combinations of these: that is the general picture, which we do not prove here.

19.6.5 Absorbing chains

The simplest traps are single states. A state j is **absorbing** if $a_{jj} = 1$, so that column j of $\mathbf{A}$ is $\mathbf{e}_j$: once there, the chain stays. Absorbing states model the end of a process: a game won or lost, a customer who has left, a molecule that has reacted.

Definition 19.6.1. Absorbing Chain

A stochastic matrix $\mathbf{A} \in M_n(\mathbb{R})$ is **absorbing** if it has at least one absorbing state, and **from every state some absorbing state can be reached**: for every j there are an absorbing state i and an $m \geq 0$ with $(\mathbf{A}^m)_{ij} > 0$. The states that are not absorbing are called **transient**.

Number the states so that the $t \geq 1$ transient states come first and the $r \geq 1$ absorbing states last. Column j for a transient j splits into the probabilities of moving to transient states and to absorbing states, and the column of an absorbing state is a standard basis vector. So $\mathbf{A}$ has the block form

$$\mathbf{A} = \begin{pmatrix} \mathbf{Q} & \mathbf{O} \\ \mathbf{R} & \mathbf{I}_r \end{pmatrix}, \quad \mathbf{Q} \in M_t(\mathbb{R}), \quad \mathbf{R} \in M_{r \times t}(\mathbb{R}),$$

with $\mathbf{Q}, \mathbf{R} \geq \mathbf{O}$ and $\mathbf{1}^\top \mathbf{Q} + \mathbf{1}^\top \mathbf{R} = \mathbf{1}^\top$. Multiplying blocks and inducting on m ,

$$\mathbf{A}^m = \begin{pmatrix} \mathbf{Q}^m & \mathbf{O} \\ \mathbf{R}_m & \mathbf{I}_r \end{pmatrix}, \quad \mathbf{R}_m = \mathbf{R}(\mathbf{I} + \mathbf{Q} + \cdots + \mathbf{Q}^{m-1}),$$

since $\mathbf{A}^{m+1} = \mathbf{A}\mathbf{A}^m$ has lower-left block $\mathbf{R}\mathbf{Q}^m + \mathbf{R}_m = \mathbf{R}_{m+1}$. The entry $(\mathbf{Q}^m)_{ij}$ is the probability of being at transient state i after m steps from transient state j ; the column sums of $\mathbf{Q}^m$ are the probabilities of not yet having been absorbed.

Lemma 19.6.4. Transient Mass Dies Out

If $\mathbf{A}$ is absorbing, with $\mathbf{Q}$ as above, then every eigenvalue of $\mathbf{Q}$ has modulus less than 1, and $\mathbf{Q}^m \rightarrow \mathbf{O}$.

Idea. Every column of $\mathbf{Q}$ sums to at most 1, which only gives $\rho(\mathbf{Q}) \leq 1$. But after enough steps every transient state has leaked some mass into an absorbing state, so every column of a suitable power $\mathbf{Q}^M$ sums to **strictly** less than 1, and the column-sum bound of Chapter 16 gives $\rho(\mathbf{Q}^M) < 1$.

Proof. For absorbing i and any j , $(\mathbf{A}^{m+1})_{ij} = \sum_l a_{il}(\mathbf{A}^m)_{lj} \geq a_{ii}(\mathbf{A}^m)_{ij} = (\mathbf{A}^m)_{ij}$, since all terms are nonnegative; so the probability of being at i never decreases. Fix a transient j . By Definition 19.6.1 there are an absorbing i and m_j with $(\mathbf{A}^{m_j})_{ij} > 0$, and then $(\mathbf{A}^m)_{ij} > 0$ for every $m \geq m_j$. Let M be the largest of the m_j over the transient j . Column j of $\mathbf{A}^M$ sums to 1 (Proposition 9.11.3 (b)) and has a positive entry $(\mathbf{A}^M)_{ij}$ in the absorbing rows, so column j of $\mathbf{Q}^M$ sums to less than 1. Since $\mathbf{Q}^M \geq \mathbf{O}$, its largest absolute column sum is some $c < 1$, and by Corollary 16.4.2, $\rho(\mathbf{Q}^M) \leq c < 1$. By Theorem 9.10.1 with $q = x^M$, the eigenvalues of $\mathbf{Q}^M$ are the M -th powers of those of $\mathbf{Q}$, so $\rho(\mathbf{Q})^M = \rho(\mathbf{Q}^M) < 1$ and $\rho(\mathbf{Q}) < 1$. Finally $\mathbf{Q}^m \rightarrow \mathbf{O}$ by Theorem 10.10.2.



The Neumann series now does all the work.

Theorem 19.6.5. Absorbing Chains

Let $\mathbf{A} = \begin{pmatrix} \mathbf{Q} & \mathbf{O} \\ \mathbf{R} & \mathbf{I}_r \end{pmatrix}$ be absorbing, in the block form above. Then:

- (a) $\mathbf{I} - \mathbf{Q}$ is invertible, and the **fundamental matrix** $\mathbf{N} := (\mathbf{I} - \mathbf{Q})^{-1} = \sum_{k \geq 0} \mathbf{Q}^k$ satisfies $\mathbf{N} \geq \mathbf{O}$;
- (b) $\mathbf{A}^m \rightarrow \begin{pmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{RN} & \mathbf{I}_r \end{pmatrix}$ entrywise;
- (c) every column of the $r \times t$ matrix $\mathbf{B} := \mathbf{RN}$ is a probability vector.

Idea. The lemma gives $\rho(\mathbf{Q}) < 1$, which is exactly what the Neumann series needs. In the block formula for $\mathbf{A}^m$, the lower-left block is $\mathbf{R}$ times a partial sum of that series, so it converges to $\mathbf{RN}$; and column sums equal to 1 pass to the limit.

Proof.

- (a) By Lemma 19.6.4 and Theorem 10.10.6, $\sum_k \mathbf{Q}^k$ converges and equals $(\mathbf{I} - \mathbf{Q})^{-1}$. Each partial sum is $\geq \mathbf{O}$, and a limit of nonnegative numbers is nonnegative, since a nonstrict inequality survives a limit (the elementary algebra of limits, recalled in the introduction to Chapter 16); so $\mathbf{N} \geq \mathbf{O}$.
- (b) By the block formula, the upper-left block of $\mathbf{A}^m$ is $\mathbf{Q}^m \rightarrow \mathbf{O}$ (Lemma 19.6.4), and the lower-left block is $\mathbf{R}_m = \mathbf{R} \sum_{k < m} \mathbf{Q}^k$. Each entry of $\mathbf{R}_m$ is a fixed linear combination of entries of the partial sum, so $\mathbf{R}_m \rightarrow \mathbf{RN}$ by the algebra of limits, (B3) of Chapter 10 §10. The other two blocks do not depend on m .
- (c) Every column of $\mathbf{A}^m$ sums to 1, and so does every column of the limit in (b). For a transient j , column j of the limit is $(\mathbf{0}, \mathbf{Be}_j)$, so the entries of $\mathbf{Be}_j$ sum to 1; they are nonnegative because $\mathbf{R}, \mathbf{N} \geq \mathbf{O}$.



Reading the answer. The theorem is pure algebra; its meaning comes from reading entries as probabilities. The (i, j) -entry of the limit in (b), for absorbing i and transient j , is the limit of the nondecreasing probabilities of being at i after m steps from j : the **probability of being absorbed at i** . So $\mathbf{B} = \mathbf{RN}$ is the matrix of absorption probabilities. Two further readings use a fact from probability that this book does not develop, **linearity of expectation**: the expected number of steps at which an event occurs is the sum over the steps of the probabilities of that event. With it, since $(\mathbf{Q}^k)_{ij}$ is the probability of being at transient i at step k from j ,

$$\begin{aligned} N_{ij} &= \sum_{k \geq 0} (\mathbf{Q}^k)_{ij} \\ &= \text{expected number of visits to } i \text{ starting from } j, \end{aligned}$$

counting the visit at step 0 when $i = j$; and summing over i , the column sums $\tau_j = (\mathbf{1}^\top \mathbf{N})_j$

are the **expected numbers of steps before absorption**. Nothing in the proofs depends on these readings.

Example. Gambler's Ruin

A gambler holds 1, 2, or 3 coins and bets one coin at a time on a fair toss, winning or losing a coin with probability $\frac{1}{2}$ each. Play stops when the gambler holds 0 coins (ruin) or 4 coins (target). Find the fundamental matrix, the expected length of play, and the probability of reaching the target, from each starting fortune.

Solution. The transient states are the fortunes 1, 2, 3, in that order, and the absorbing ones are 0 and 4, in that order. From fortune j the gambler moves to $j-1$ or $j+1$ with probability $\frac{1}{2}$ each, so

$$\mathbf{Q} = \begin{pmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 0 & 1/2 \end{pmatrix},$$

the first row of $\mathbf{R}$ recording ruin from fortune 1 and the second reaching the target from fortune 3. Each column of $\begin{pmatrix} \mathbf{Q} \\ \mathbf{R} \end{pmatrix}$ sums to 1, and every fortune reaches 0 by losing, so the chain is absorbing.

The fundamental matrix. Solve $(\mathbf{I} - \mathbf{Q})\mathbf{N} = \mathbf{I}$ with $\mathbf{I} - \mathbf{Q} = \begin{pmatrix} 1 & -1/2 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & -1/2 & 1 \end{pmatrix}$. The result is

$$\mathbf{N} = \begin{pmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{pmatrix};$$

check the first column: $(\mathbf{I} - \mathbf{Q})(\frac{3}{2}, 1, \frac{1}{2}) = (\frac{3}{2} - \frac{1}{2}, -\frac{3}{4} + 1 - \frac{1}{4}, -\frac{1}{2} + \frac{1}{2}) = (1, 0, 0)$, and the other columns likewise. A gambler starting with 2 coins holds exactly 2 coins at 2 steps on average, counting the start, and holds 1 or 3 coins at one step each on average.

Length of play. $\mathbf{1}^\top \mathbf{N} = (3, 4, 3)$: from fortunes 1, 2, 3 the game lasts 3, 4, 3 tosses on average.

Absorption.

$$\mathbf{B} = \mathbf{R}\mathbf{N} = \begin{pmatrix} 3/4 & 1/2 & 1/4 \\ 1/4 & 1/2 & 3/4 \end{pmatrix}.$$

From fortune j the gambler reaches the target with probability $j/4$ and is ruined with probability $1 - j/4$. Each column sums to 1, as [Theorem 19.6.5](#) (c) requires: the game ends with probability 1. Both answers match the classical formulas for a fair game with target $T = 4$: success probability j/T and expected length $j(T - j)$, which gives 3, 4, 3.

The matrix $\mathbf{Q}$ of this example has eigenvalues 0 and $\pm \frac{1}{\sqrt{2}}$, of modulus less than 1 as [Lemma 19.6.4](#) requires; but the lemma was proved without finding them, which is the pattern of this whole section.

19.6.6 Exercises

19.6.6.1 A. Check your understanding

Exercise. A1

- (a) State the long-run theorem [Theorem 19.6.2](#), and name the three hypotheses of Theorem 9.11.4 that it no longer needs.
- (b) True or false: every irreducible stochastic matrix has exactly one steady state. Justify your answer.
- (c) True or false: for every irreducible stochastic matrix, $\mathbf{A}^m$ converges. Justify your answer.
- (d) True or false: if a stochastic matrix has two different steady states, it has infinitely many. Justify your answer.
- (e) In an absorbing chain, what do the column sums of the fundamental matrix measure, and why do the columns of $\mathbf{RN}$ sum to 1?

Solution.

- (a) If $\mathbf{A}$ is stochastic and $\mathbf{A}^k > \mathbf{O}$ for some k , then $\mathbf{A}$ has exactly one steady state $\mathbf{v}$, it is positive, $\mathbf{A}^m \rightarrow \mathbf{v}\mathbf{1}^\top$, and every chain $\mathbf{A}^m \mathbf{x}_0$ converges to $\mathbf{v}$. It no longer assumes that $\mathbf{A}$ is diagonalizable, that $a_{\mathbf{A}}(1) = 1$, or that every other eigenvalue has modulus less than 1.
- (b) True, by [Proposition 19.6.1](#) (c); moreover the steady state is positive.
- (c) False. The swap $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is irreducible and stochastic, and its powers alternate between $\mathbf{I}$ and itself. By [Theorem 19.6.3](#) (c), the powers converge exactly when the matrix is primitive.
- (d) True. If $\mathbf{u} \neq \mathbf{u}'$ are steady states, then $c\mathbf{u} + (1-c)\mathbf{u}'$ for $0 \leq c \leq 1$ is nonnegative, has entries summing to 1, and is fixed by $\mathbf{A}$; different c give different vectors, since $\mathbf{u} - \mathbf{u}' \neq \mathbf{0}$.
- (e) The column sum $(\mathbf{1}^\top \mathbf{N})_j$ is the expected number of steps before absorption from transient state j . The columns of $\mathbf{RN}$ sum to 1 because the columns of $\mathbf{A}^m$ do and $\mathbf{Q}^m \rightarrow \mathbf{O}$ ([Theorem 19.6.5](#) (c)): absorption is certain.

19.6.6.2 B. Practice

Exercise. B1: When does the long-run theorem apply?

For each stochastic matrix, determine whether [Theorem 19.6.2](#) applies. Justify your answer, and when it applies, find the limit of $\mathbf{A}^m$.

(a) $\begin{pmatrix} 0 & 0 & 1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 1/2 \end{pmatrix}$.

- (b) $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$
- (c) $\begin{pmatrix} 1/2 & 0 \\ 1/2 & 1 \end{pmatrix}.$

Solution.

- (a) It applies. The edges of $G(\mathbf{A})$ are $1 \rightarrow 2$, $2 \rightarrow 3$, $3 \rightarrow 1$ and the loop $3 \rightarrow 3$. The cycle $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ reaches every state from every state, so $\mathbf{A}$ is irreducible (Theorem 19.2.5). It has the positive diagonal entry a_{33} , so it is primitive by Corollary 19.5.5. The steady state solves $v_1 = \frac{1}{2}v_3$, $v_2 = v_1$, so $\mathbf{v} = (1, 1, 2)/4$, and $\mathbf{A}^m \rightarrow \mathbf{v}\mathbf{1}^\top$, every column equal to $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$. Check: $\mathbf{A}\mathbf{v} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4} + \frac{1}{4}) = \mathbf{v}$.
- (b) It does not apply. This is a permutation matrix, so every power is a permutation matrix (Lemma 3.4.7) and has zero entries. Its powers cycle with period 3 (Example (The Cyclic Permutation)) and do not converge.
- (c) It does not apply. Every power is lower triangular with $(1, 2)$ -entry 0, as one checks by induction: $\mathbf{A}^m = \begin{pmatrix} 2^{-m} & 0 \\ 1 - 2^{-m} & 1 \end{pmatrix}$. The powers converge nonetheless, to $\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} = \mathbf{e}_2\mathbf{1}^\top$: state 2 is absorbing, and $\mathbf{e}_2$ is the unique steady state, with a zero entry.

Exercise. B2: A three-state chain

Let $\mathbf{A} = \begin{pmatrix} 1/2 & 1/4 & 0 \\ 1/2 & 1/2 & 1/2 \\ 0 & 1/4 & 1/2 \end{pmatrix}.$

- (a) Show that $\mathbf{A}$ is stochastic and that $\mathbf{A}^2 > \mathbf{O}$.
- (b) Find the steady state, and hence $\lim_m \mathbf{A}^m$ and the long-run distribution of every chain with this transition matrix.

Solution.

- (a) The entries are nonnegative and the columns sum to 1. Multiplying out,

$$\mathbf{A}^2 = \begin{pmatrix} 3/8 & 1/4 & 1/8 \\ 1/2 & 1/2 & 1/2 \\ 1/8 & 1/4 & 3/8 \end{pmatrix} > \mathbf{O}.$$

- (b) $(\mathbf{A} - \mathbf{I})\mathbf{v} = \mathbf{0}$: the first row gives $-\frac{1}{2}v_1 + \frac{1}{4}v_2 = 0$, so $v_2 = 2v_1$, and the third gives $\frac{1}{4}v_2 - \frac{1}{2}v_3 = 0$, so $v_3 = \frac{1}{2}v_2 = v_1$. So $\mathbf{v}$ is a multiple of $(1, 2, 1)$, and scaling to sum 1 gives $\mathbf{v} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. By Theorem 19.6.2, $\mathbf{A}^m \rightarrow \mathbf{v}\mathbf{1}^\top$, every column equal to $\mathbf{v}$, and every chain converges to $\mathbf{v}$ whatever its initial distribution.

Exercise. B3: An unfair game

A gambler holds 1 or 2 coins and wins each toss with probability $\frac{1}{3}$, losing with probability $\frac{2}{3}$. Play stops at 0 or 3 coins. Find the fundamental matrix, the expected length of play, and the probability of reaching 3 coins from each starting fortune.

Solution. Order the states as 1, 2 (transient) and 0, 3 (absorbing). From 1: to 2 with probability $\frac{1}{3}$, to 0 with $\frac{2}{3}$. From 2: to 3 with $\frac{1}{3}$, to 1 with $\frac{2}{3}$. So

$$\mathbf{Q} = \begin{pmatrix} 0 & 2/3 \\ 1/3 & 0 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 2/3 & 0 \\ 0 & 1/3 \end{pmatrix}.$$

The chain is absorbing, since from either fortune repeated losses reach 0. Then $\mathbf{I} - \mathbf{Q} = \begin{pmatrix} 1 & -2/3 \\ -1/3 & 1 \end{pmatrix}$ has determinant $1 - \frac{2}{9} = \frac{7}{9}$, so by Theorem 0.9.8

$$\mathbf{N} = \frac{9}{7} \begin{pmatrix} 1 & 2/3 \\ 1/3 & 1 \end{pmatrix} = \begin{pmatrix} 9/7 & 6/7 \\ 3/7 & 9/7 \end{pmatrix}.$$

The expected lengths of play are $\mathbf{1}^\top \mathbf{N} = (\frac{12}{7}, \frac{15}{7})$, and

$$\mathbf{R}\mathbf{N} = \begin{pmatrix} 6/7 & 4/7 \\ 1/7 & 3/7 \end{pmatrix}.$$

So the probability of reaching 3 coins is $\frac{1}{7}$ from 1 coin and $\frac{3}{7}$ from 2 coins, far below the fair values $\frac{1}{3}$ and $\frac{2}{3}$. The columns of $\mathbf{R}\mathbf{N}$ sum to 1, as they must.

19.6.6.3 C. Going deeper
Exercise. C1: The lazy chain

Let $\mathbf{A} \in M_n(\mathbb{R})$ be stochastic and irreducible, with steady state $\mathbf{v}$, and let $\mathbf{M} = \frac{1}{2}(\mathbf{I} + \mathbf{A})$.

- Prove that $\mathbf{M}$ is stochastic and primitive, and that $\mathbf{v}$ is its unique steady state. Deduce that $\mathbf{M}^m \rightarrow \mathbf{v}\mathbf{1}^\top$.
- Let $\mathbf{A}$ be the walk of [Example \(A Walk that Oscillates\)](#). Show that $\mathbf{M}$ is the matrix of exercise B2, and that $\mathbf{M}^m = \mathbf{v}\mathbf{1}^\top + 2^{-m}\mathbf{F}$ for every $m \geq 1$, where $\mathbf{F} = \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}$.

Hint: write $\mathbf{M}$ in terms of $\mathbf{v}\mathbf{1}^\top$ and $\mathbf{F}$.

Solution.

- $\mathbf{M} \geq \mathbf{O}$, and $\mathbf{1}^\top \mathbf{M} = \frac{1}{2}(\mathbf{1}^\top + \mathbf{1}^\top \mathbf{A}) = \mathbf{1}^\top$, so $\mathbf{M}$ is stochastic. By Exercise (C1: Shifting by the identity) (a) with $t = 1$, $\mathbf{A} + \mathbf{I}$ is primitive, say $(\mathbf{A} + \mathbf{I})^k > \mathbf{O}$; then $\mathbf{M}^k = 2^{-k}(\mathbf{A} + \mathbf{I})^k > \mathbf{O}$, so $\mathbf{M}$ is primitive. Also $\mathbf{M}\mathbf{v} = \frac{1}{2}(\mathbf{v} + \mathbf{v}) = \mathbf{v}$. By [Theorem 19.6.2 \(a\)](#), $\mathbf{M}$ has exactly one steady state, so it is $\mathbf{v}$, and by part (b) of that theorem $\mathbf{M}^m \rightarrow \mathbf{v}\mathbf{1}^\top$.

(b) With $\mathbf{A}$ as in [Example \(A Walk that Oscillates\)](#),

$$\mathbf{M} = \frac{1}{2} \begin{pmatrix} 1 & 1/2 & 0 \\ 1 & 1 & 1 \\ 0 & 1/2 & 1 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/4 & 0 \\ 1/2 & 1/2 & 1/2 \\ 0 & 1/4 & 1/2 \end{pmatrix},$$

the matrix of [Exercise \(B2: A three-state chain\)](#), and $\mathbf{v} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Put $\mathbf{L} = \mathbf{v}\mathbf{1}^\top$, every column equal to $\mathbf{v}$. Then $\mathbf{L}^2 = \mathbf{v}(\mathbf{1}^\top \mathbf{v})\mathbf{1}^\top = \mathbf{L}$; $\mathbf{1}^\top \mathbf{F} = \mathbf{0}^\top$, since each column of $\mathbf{F}$ sums to 0, so $\mathbf{L}\mathbf{F} = \mathbf{v}(\mathbf{1}^\top \mathbf{F}) = \mathbf{0}$; and $\mathbf{F}\mathbf{v} = \frac{1}{2}(\frac{1}{4} - \frac{1}{4}, 0, -\frac{1}{4} + \frac{1}{4}) = \mathbf{0}$, so $\mathbf{F}\mathbf{L} = (\mathbf{F}\mathbf{v})\mathbf{1}^\top = \mathbf{0}$.

Also $\mathbf{F}^2 = \frac{1}{4} \begin{pmatrix} 2 & 0 & -2 \\ 0 & 0 & 0 \\ -2 & 0 & 2 \end{pmatrix} = \mathbf{F}$. Subtracting, $\mathbf{M} - \mathbf{L} = \frac{1}{4} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \frac{1}{2}\mathbf{F}$, so

$\mathbf{M} = \mathbf{L} + \frac{1}{2}\mathbf{F}$. If $\mathbf{M}^m = \mathbf{L} + 2^{-m}\mathbf{F}$, then

$$\begin{aligned} \mathbf{M}^{m+1} &= \left(\mathbf{L} + \frac{1}{2}\mathbf{F}\right)(\mathbf{L} + 2^{-m}\mathbf{F}) \\ &= \mathbf{L}^2 + 2^{-m}\mathbf{L}\mathbf{F} + \frac{1}{2}\mathbf{F}\mathbf{L} + 2^{-m-1}\mathbf{F}^2 = \mathbf{L} + 2^{-(m+1)}\mathbf{F}, \end{aligned}$$

so the formula holds for every $m \geq 1$ by induction, starting from $\mathbf{M} = \mathbf{L} + \frac{1}{2}\mathbf{F}$. The lazy walk converges although the walk itself oscillates, and its distance to the limit halves at every step.

Exercise. C2: The converse of the long-run theorem

Let $\mathbf{A} \in M_n(\mathbb{R})$ be stochastic.

- Prove that if $\mathbf{A}^m \rightarrow \mathbf{v}\mathbf{1}^\top$ for some $\mathbf{v} > \mathbf{0}$, then $\mathbf{A}$ is primitive.
- Give a stochastic $\mathbf{A}$ with $\mathbf{A}^m \rightarrow \mathbf{v}\mathbf{1}^\top$ for a probability vector $\mathbf{v}$, but no power of $\mathbf{A}$ positive. Which entry of $\mathbf{v}$ is responsible?

Solution.

- The limit $\mathbf{v}\mathbf{1}^\top$ has (i, j) -entry $v_i > 0$. Each of the n^2 entries of $\mathbf{A}^m$ converges to a positive number, so it is positive for all m beyond some m_{ij} ; for m larger than all the m_{ij} , $\mathbf{A}^m > \mathbf{0}$. So $\mathbf{A}$ is primitive.
- $\mathbf{A} = \begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix}$. By induction $\mathbf{A}^m = \begin{pmatrix} 1 & 1 - 2^{-m} \\ 0 & 2^{-m} \end{pmatrix}$: indeed $\mathbf{A}\mathbf{A}^m$ has first row $(1, 1 - 2^{-m} + 2^{-m-1}) = (1, 1 - 2^{-(m+1)})$ and second row $(0, 2^{-(m+1)})$. So $\mathbf{A}^m \rightarrow \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \mathbf{e}_1\mathbf{1}^\top$, while every power has $(2, 1)$ -entry 0. The responsible entry is $v_2 = 0$: (a) fails exactly because $\mathbf{v}$ is not positive.

Exercise. C3: First-step analysis

Let $\mathbf{A}$ be absorbing, with fundamental matrix $\mathbf{N}$ and $\boldsymbol{\tau}^\top = \mathbf{1}^\top \mathbf{N}$.

- Prove that $\mathbf{N} = \mathbf{I} + \mathbf{N}\mathbf{Q}$, and deduce that $\tau_j = 1 + \sum_i q_{ij}\tau_i$ for every transient j .

Interpret this equation.

- (b) Prove, by multiplying by $\mathbf{R}$, that $\mathbf{B} = \mathbf{R}\mathbf{N}$ satisfies $\mathbf{B} = \mathbf{R} + \mathbf{B}\mathbf{Q}$, and interpret it.
- (c) Use (a) to recompute the expected lengths of play (3, 4, 3) of [Example \(Gambler's Ruin\)](#) without inverting a matrix.

Solution.

- (a) $\mathbf{N}(\mathbf{I} - \mathbf{Q}) = \mathbf{I}$ because $\mathbf{N}$ is the inverse of $\mathbf{I} - \mathbf{Q}$, so $\mathbf{N} = \mathbf{I} + \mathbf{N}\mathbf{Q}$. Multiplying on the left by $\mathbf{1}^\top$, $\boldsymbol{\tau}^\top = \mathbf{1}^\top + \boldsymbol{\tau}^\top \mathbf{Q}$, whose j -th entry is $\tau_j = 1 + \sum_i \tau_i q_{ij}$. In words: from j , one step is taken, the chain lands at transient i with probability q_{ij} , and from there the expected remaining time is τ_i ; landing at an absorbing state contributes nothing more.
- (b) Multiply $\mathbf{N} = \mathbf{I} + \mathbf{N}\mathbf{Q}$ on the left by $\mathbf{R}$: $\mathbf{B} = \mathbf{R} + \mathbf{B}\mathbf{Q}$, that is, $b_{ij} = r_{ij} + \sum_l b_{il} q_{lj}$. The probability of absorption at i from j is the probability of stepping into i at once, plus the probability of stepping to a transient l and being absorbed at i from there.
- (c) With $\mathbf{Q}$ as in [Example \(Gambler's Ruin\)](#), (a) reads $\tau_1 = 1 + \frac{1}{2}\tau_2$, $\tau_2 = 1 + \frac{1}{2}\tau_1 + \frac{1}{2}\tau_3$, $\tau_3 = 1 + \frac{1}{2}\tau_2$. The first and third equations give $\tau_1 = \tau_3 = 1 + \frac{1}{2}\tau_2$, so $\tau_2 = 1 + \tau_1 = 1 + 1 + \frac{1}{2}\tau_2$, giving $\tau_2 = 4$ and $\tau_1 = \tau_3 = 3$. (The system has only one solution, since its matrix is $(\mathbf{I} - \mathbf{Q})^\top$, which is invertible by [Theorem 19.6.5](#) (a).)