

9.11 Powers, Recurrences, and Markov Chains

Section 4 promised that diagonalization is a tool for computation, and the Fibonacci numbers gave a first taste. This section collects three situations in which a quantity evolves in discrete steps by a fixed linear rule: a vector recurrence $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$, a scalar recurrence such as $a_{k+3} = 2a_{k+2} + a_{k+1} - 2a_k$, and a system that moves at random between finitely many states. In each case the state after k steps is a matrix power applied to the starting state, and the eigenvalues decide what happens in the long run. We use only the explicit formula $\mathbf{A}^k = \mathbf{P}\mathbf{D}^k\mathbf{P}^{-1}$ and limits of sequences of numbers, one entry at a time.

9.11.1 Vector recurrences

A **linear recurrence** of first order is a rule $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$ for $k \in \mathbb{N}$, with $\mathbf{A} \in M_n(F)$ fixed and a starting vector $\mathbf{x}_0 \in F^n$. By induction on k , the unique sequence obeying the rule is $\mathbf{x}_k = \mathbf{A}^k\mathbf{x}_0$. The question is how to see this power, and diagonalization answers it by changing to coordinates in which $\mathbf{A}$ acts coordinate by coordinate.

Theorem 9.11.1. Solving a Diagonalizable Recurrence

Let $\mathbf{A} \in M_n(F)$ be diagonalizable, with $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$, $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$, and let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be the columns of $\mathbf{P}$. Let $\mathbf{x}_0 \in F^n$, and let $(c_1, \dots, c_n) = \mathbf{P}^{-1}\mathbf{x}_0$, so that $\mathbf{x}_0 = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n$. Then the sequence defined by $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$ is

$$\mathbf{x}_k = \mathbf{A}^k\mathbf{x}_0 = c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 + \dots + c_n\lambda_n^k\mathbf{v}_n \quad (k \in \mathbb{N}).$$

Proof. By induction on k , $\mathbf{x}_k = \mathbf{A}^k\mathbf{x}_0$: this holds for $k = 0$ since $\mathbf{A}^0 = \mathbf{I}$, and $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k = \mathbf{A} \cdot \mathbf{A}^k\mathbf{x}_0 = \mathbf{A}^{k+1}\mathbf{x}_0$. By Theorem 9.4.3, $\mathbf{A}^k\mathbf{x}_0 = \mathbf{P}\mathbf{D}^k\mathbf{P}^{-1}\mathbf{x}_0 = \mathbf{P}\mathbf{D}^k\mathbf{c}$ with $\mathbf{c} = (c_1, \dots, c_n)$. Now $\mathbf{D}^k\mathbf{c} = (\lambda_1^k c_1, \dots, \lambda_n^k c_n)$, and by Theorem 0.9.3, $\mathbf{P}$ times this vector is $\sum_i \lambda_i^k c_i \mathbf{v}_i$. Finally, $\mathbf{P}\mathbf{c} = \mathbf{x}_0$ says $\mathbf{x}_0 = \sum_i c_i \mathbf{v}_i$, by the same result. ■

In words: expand the starting vector in a basis of eigenvectors; each coordinate is then multiplied by its own eigenvalue at every step, independently of the others. Over $\mathbb{R}$ or $\mathbb{C}$, where absolute values are available, the long-run behavior is visible at once. Terms with $|\lambda_i| < 1$ die out, terms with $|\lambda_i| > 1$ grow, and the largest $|\lambda_i|$ with $c_i \neq 0$ wins. The worked Fibonacci example (Example (A Closed Formula for the Fibonacci Numbers)) is the case $n = 2$.

9.11.2 Scalar recurrences and companion matrices

A recurrence such as $a_{k+3} = 2a_{k+2} + a_{k+1} - 2a_k$ involves one number sequence but several previous terms. The trick of Example (A Closed Formula for the Fibonacci Numbers) turns it into a vector recurrence: remember the last few terms as one vector.

Let $n \geq 1$ and $c_0, \dots, c_{n-1} \in F$, and consider sequences $(a_k)_{k \in \mathbb{N}}$ in F with

$$a_{k+n} + c_{n-1}a_{k+n-1} + \dots + c_1a_{k+1} + c_0a_k = 0 \quad (*)$$

for every $k \in \mathbb{N}$.

Such a sequence is determined by its **initial values** $a_0, \dots, a_{n-1}$, because $(*)$ computes each later term from the n before it. The polynomial

$$p(x) = x^n + c_{n-1}x^{n-1} + \dots + c_1x + c_0$$

is called the **characteristic polynomial of the recurrence**; the name is justified below. Put $\mathbf{x}_k = (a_k, a_{k+1}, \dots, a_{k+n-1}) \in F^n$. Then $(*)$ says exactly $\mathbf{x}_{k+1} = \mathbf{C}\mathbf{x}_k$, where

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -c_0 & -c_1 & -c_2 & \dots & -c_{n-1} \end{pmatrix} :$$

the first $n-1$ rows shift the window by one, and the last row is $(*)$ solved for a_{k+n} . For $n \geq 2$, $\mathbf{C}$ is the transpose of the companion matrix $\mathbf{C}(p)$ of Exercise (C1: Companion matrices), so $p_{\mathbf{C}} = p_{\mathbf{C}(p)} = p$ by Exercise (B2: Transpose) and Exercise (C1: Companion matrices) (for $n=1$, $\mathbf{C} = (-c_0)$ and $p_{\mathbf{C}} = x + c_0 = p$). The eigenvalues of $\mathbf{C}$ are the roots of p .

Theorem 9.11.2. Scalar Recurrences with Distinct Roots

Let $n \geq 1$, $c_0, \dots, c_{n-1} \in F$, and suppose $p = x^n + c_{n-1}x^{n-1} + \dots + c_0$ has n **distinct** roots $\lambda_1, \dots, \lambda_n \in F$. Then a sequence $(a_k)_{k \in \mathbb{N}}$ in F satisfies $(*)$ if and only if there are $b_1, \dots, b_n \in F$ with

$$a_k = b_1\lambda_1^k + b_2\lambda_2^k + \dots + b_n\lambda_n^k \quad \text{for every } k \in \mathbb{N}.$$

In that case $b_1, \dots, b_n$ are uniquely determined by the initial values $a_0, \dots, a_{n-1}$.

Idea. Why powers of the roots? A geometric sequence $a_k = \lambda^k$ satisfies $(*)$ exactly when $\lambda^k p(\lambda) = 0$, so the roots of p give n solutions. To see that there are no others, use the matrix $\mathbf{C}$: its eigenvector for λ_i is the window $(1, \lambda_i, \dots, \lambda_i^{n-1})$ of the geometric sequence, these eigenvectors form an invertible Vandermonde matrix, and [Theorem 9.11.1](#) does the rest.

Proof. $(\Leftarrow)$ For each i and k , $\lambda_i^{k+n} + c_{n-1}\lambda_i^{k+n-1} + \dots + c_0\lambda_i^k = \lambda_i^k p(\lambda_i) = 0$. If $a_k = \sum_i b_i \lambda_i^k$, then the left side of $(*)$ is $\sum_i b_i \cdot 0 = 0$, after regrouping the finite sums.

$(\Rightarrow)$ Let (a_k) satisfy $(*)$, and let $\mathbf{x}_k$ and $\mathbf{C}$ be as above, so $\mathbf{x}_{k+1} = \mathbf{C}\mathbf{x}_k$. For each i , let $\mathbf{w}_i = (1, \lambda_i, \lambda_i^2, \dots, \lambda_i^{n-1})$. Then $\mathbf{C}\mathbf{w}_i = \lambda_i \mathbf{w}_i$: for $r \leq n-1$, entry r of $\mathbf{C}\mathbf{w}_i$ is entry $r+1$ of $\mathbf{w}_i$, which is $\lambda_i^r = \lambda_i \cdot \lambda_i^{r-1}$; and the last entry is $-(c_0 + c_1\lambda_i + \dots + c_{n-1}\lambda_i^{n-1}) = \lambda_i^n - p(\lambda_i) = \lambda_i^n = \lambda_i \cdot \lambda_i^{n-1}$. Let $\mathbf{W}$ be the matrix with columns $\mathbf{w}_1, \dots, \mathbf{w}_n$. Its (r, i) -entry is λ_i^{r-1} , so $\mathbf{W} = \mathbf{V}(\lambda_1, \dots, \lambda_n)^\top$ in the notation of Theorem 7.8.1, and $\mathbf{W}$ is invertible, since the λ_i are distinct: $\det \mathbf{W} = \det \mathbf{V}(\lambda_1, \dots, \lambda_n) \neq 0$ by Theorem 7.4.5 and Theorem 7.8.1, and Theorem 7.5.2 applies. The columns of $\mathbf{C}\mathbf{W}$ are $\lambda_i \mathbf{w}_i$, which are the columns of $\mathbf{W} \text{diag}(\lambda_1, \dots, \lambda_n)$, so $\mathbf{C} = \mathbf{W} \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{W}^{-1}$.

By [Theorem 9.11.1](#), with $(b_1, \dots, b_n) = \mathbf{W}^{-1}\mathbf{x}_0$,

$$\mathbf{x}_k = b_1\lambda_1^k \mathbf{w}_1 + \dots + b_n\lambda_n^k \mathbf{w}_n.$$

The first entry of $\mathbf{x}_k$ is a_k , and the first entry of each $\mathbf{w}_i$ is 1, so $a_k = \sum_i b_i \lambda_i^k$.

For uniqueness, suppose $a_k = \sum_i b_i \lambda_i^k$ for all k . Taking $k = 0, \dots, n-1$ gives $\mathbf{W}\mathbf{b} = (a_0, \dots, a_{n-1}) = \mathbf{x}_0$, so $\mathbf{b} = \mathbf{W}^{-1}\mathbf{x}_0$ is determined by the initial values. ■

So solving a recurrence with distinct roots comes down to three steps: find the roots of p , write the general form $\sum b_i \lambda_i^k$, and fit $b_1, \dots, b_n$ to the initial values by solving a Vandermonde system.

Example. A Third-Order Recurrence

Find a formula for the sequence with $a_0 = 3$, $a_1 = 2$, $a_2 = 6$ and

$$a_{k+3} = 2a_{k+2} + a_{k+1} - 2a_k \quad (k \in \mathbb{N}).$$

Solution. In the form (*), $a_{k+3} - 2a_{k+2} - a_{k+1} + 2a_k = 0$, so $p(x) = x^3 - 2x^2 - x + 2$. Grouping, $p = x^2(x-2) - (x-2) = (x-2)(x-1)(x+1)$, with the distinct roots $1, -1, 2$. The window matrix is

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -2 & 1 & 2 \end{pmatrix},$$

with eigenvectors $(1, 1, 1)$, $(1, -1, 1)$, $(1, 2, 4)$; for instance $\mathbf{C}(1, 2, 4) = (2, 4, -2 + 2 + 8) = 2(1, 2, 4)$.

By Theorem 9.11.2, $a_k = b_1 + b_2(-1)^k + b_3 2^k$. The initial values give the Vandermonde system

$$b_1 + b_2 + b_3 = 3, \quad b_1 - b_2 + 2b_3 = 2, \quad b_1 + b_2 + 4b_3 = 6.$$

Subtracting the first equation from the third gives $3b_3 = 3$, so $b_3 = 1$. Then $b_1 + b_2 = 2$ and $b_1 - b_2 = 0$, so $b_1 = b_2 = 1$. Hence

$$a_k = 1 + (-1)^k + 2^k.$$

Check. The recurrence gives $a_3 = 2 \cdot 6 + 2 - 2 \cdot 3 = 8$ and $a_4 = 2 \cdot 8 + 6 - 2 \cdot 2 = 18$; the formula gives $1 - 1 + 8 = 8$ and $1 + 1 + 16 = 18$. For large k , $a_k \approx 2^k$: the root of largest absolute value dominates.

Repeated roots. When p has a repeated root, $\mathbf{C}$ is not diagonalizable: for a root λ , every eigenvector $(w_1, \dots, w_n)$ of $\mathbf{C}$ satisfies $w_{r+1} = \lambda w_r$, so it is a multiple of $(1, \lambda, \dots, \lambda^{n-1})$, and $g_{\mathbf{C}}(\lambda) = 1$ whatever $a_{\mathbf{C}}(\lambda)$ is. The general answer involves terms $k\lambda^k, k^2\lambda^k, \dots$, and it follows from the Jordan canonical form of Chapter 10. The 2×2 case can be done by hand.

Example. A Repeated Root

Over $\mathbb{R}$, find all sequences with $a_{k+2} = 4a_{k+1} - 4a_k$ for every $k \in \mathbb{N}$.

Solution. Here $p = x^2 - 4x + 4 = (x-2)^2$, and the window matrix is $\mathbf{C} = \begin{pmatrix} 0 & 1 \\ -4 & 4 \end{pmatrix}$. It

is not diagonalizable, since $\mathbf{C} - 2\mathbf{I} = \begin{pmatrix} -2 & 1 \\ -4 & 2 \end{pmatrix}$ has rank 1, so $g(2) = 1 < 2 = a(2)$. But $\mathbf{N} = \mathbf{C} - 2\mathbf{I}$ satisfies

$$\mathbf{N}^2 = \begin{pmatrix} 4-4 & -2+2 \\ 8-8 & -4+4 \end{pmatrix} = \mathbf{O}.$$

Claim: $\mathbf{C}^k = 2^k\mathbf{I} + k2^{k-1}\mathbf{N}$ for all $k \in \mathbb{N}$. For $k = 0$ both sides are $\mathbf{I}$. If it holds for k , then, since $\mathbf{C} = 2\mathbf{I} + \mathbf{N}$ and $\mathbf{N}^2 = \mathbf{O}$,

$$\begin{aligned} \mathbf{C}^{k+1} &= (2^k\mathbf{I} + k2^{k-1}\mathbf{N})(2\mathbf{I} + \mathbf{N}) \\ &= 2^{k+1}\mathbf{I} + 2^k\mathbf{N} + k2^k\mathbf{N} + k2^{k-1}\mathbf{N}^2 = 2^{k+1}\mathbf{I} + (k+1)2^k\mathbf{N}. \end{aligned}$$

Now $\mathbf{x}_k = \mathbf{C}^k\mathbf{x}_0$ with $\mathbf{x}_0 = (a_0, a_1)$, and a_k is the first entry. Row 1 of $\mathbf{C}^k$ is $(2^k - 2k2^{k-1}, k2^{k-1}) = ((1-k)2^k, k2^{k-1})$, so

$$a_k = (1-k)2^k a_0 + k2^{k-1}a_1 = \left(a_0 + \left(\frac{a_1}{2} - a_0\right)k\right)2^k.$$

So the solutions are exactly the sequences $a_k = (b_1 + b_2k)2^k$ with $b_1, b_2 \in \mathbb{R}$: every solution has this form with $b_1 = a_0$, $b_2 = a_1/2 - a_0$, and every such sequence is the solution with $a_0 = b_1$, $a_1 = 2b_1 + 2b_2$. For example, $a_0 = 1$, $a_1 = 4$ gives $a_k = (1+k)2^k$, and indeed $a_2 = 4 \cdot 4 - 4 \cdot 1 = 12 = 3 \cdot 4$. The new term $k2^k$ is the footprint of the missing eigenvector, just as the entry k was for $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^k$ in Section 4.

9.11.3 Stochastic matrices

Now a system that moves at random. A shared bicycle is parked each night at one of a few stations; a machine is either working or under repair. We do not know the state for sure, but we know the probability of each state, and a fixed table of probabilities for moving from one state to another in one step.

Take states $1, \dots, n$. Record the probabilities of the states at step k as a column $\mathbf{x}_k \in \mathbb{R}^n$, and let a_{ij} be the probability of moving to state i in one step, given that the system is in state j . The probability of being in state i at step $k+1$ is the sum over the previous state j of “was in j ” times “moved from j to i ”, that is, $\sum_j a_{ij}(\mathbf{x}_k)_j$. So $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$: a linear recurrence. The matrix $\mathbf{A}$ has two features: its entries are probabilities, and from each state j the system must go **somewhere**, so column j adds up to 1.

A stochastic matrix is a table of transition probabilities, one column for each starting state.

Definition 9.11.1. Stochastic Matrix, Probability Vector

Let $n \geq 1$.

- (a) A vector $\mathbf{x} \in \mathbb{R}^n$ is a **probability vector** if $x_i \geq 0$ for every i and $x_1 + \dots + x_n = 1$.
- (b) A matrix $\mathbf{A} \in M_n(\mathbb{R})$ is **stochastic** if every column of $\mathbf{A}$ is a probability vector: $a_{ij} \geq 0$ for **all** i, j , and $\sum_{i=1}^n a_{ij} = 1$ for **every** j .

In words: the entries are nonnegative and each **column** sums to 1. With $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^n$, the column condition reads $\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top$, and $\mathbf{1}^\top \mathbf{x} = 1$ says that the entries of $\mathbf{x}$ add up to 1.

Examples.

- $\begin{pmatrix} 9/10 & 1/5 \\ 1/10 & 4/5 \end{pmatrix}$: columns $(9/10, 1/10)$ and $(1/5, 4/5)$ are probability vectors.
- **Degenerate cases.** $\mathbf{I}_n$ is stochastic: every state stays put. The 1×1 matrix (1) is the only 1×1 stochastic matrix. Every **permutation matrix** is stochastic, since each column is a standard basis vector; it moves the system deterministically.
- $\frac{1}{4} \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$: the columns sum to $\frac{4}{4}$, and all entries are nonnegative.

Nonexample by minimal change. Transpose the first example: $\begin{pmatrix} 9/10 & 1/10 \\ 1/5 & 4/5 \end{pmatrix}$ still has nonnegative entries, and its **rows** sum to 1, but its first column sums to $11/10$. It fails the column condition. Likewise $\begin{pmatrix} 3/2 & 0 \\ -1/2 & 1 \end{pmatrix}$ has columns summing to 1 but a negative entry.

Why this convention. Many books write probability vectors as **rows** and multiply on the left, $\mathbf{x}_{k+1}^\top = \mathbf{x}_k^\top \mathbf{M}$; their “stochastic” matrices have rows summing to 1, and they are the transposes of ours. We keep column vectors and $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$, in line with the rest of the book, so our stochastic matrices are **column-stochastic**. Translating between the conventions is a transpose, which changes neither eigenvalues nor their algebraic multiplicities (Proposition 9.10.5).

The first properties follow from $\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top$.

Proposition 9.11.3. Basic Properties of Stochastic Matrices

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{R})$ be stochastic and $\mathbf{x} \in \mathbb{R}^n$ a probability vector. Then:

- $\mathbf{A}\mathbf{x}$ is a probability vector;
- $\mathbf{A}\mathbf{B}$ is stochastic, and so is $\mathbf{A}^k$ for every $k \in \mathbb{N}$;
- $\mathbf{1}$ is a left eigenvector of $\mathbf{A}$ for the eigenvalue 1; in particular $1 \in \text{spec}(\mathbf{A})$.

Proof.

- Each entry $(\mathbf{A}\mathbf{x})_i = \sum_j a_{ij}x_j$ is a sum of products of nonnegative numbers, so it is nonnegative. By associativity, $\mathbf{1}^\top(\mathbf{A}\mathbf{x}) = (\mathbf{1}^\top \mathbf{A})\mathbf{x} = \mathbf{1}^\top \mathbf{x} = 1$.
- By Theorem 0.9.2, column j of $\mathbf{A}\mathbf{B}$ is $\mathbf{A}\mathbf{b}_j$, where $\mathbf{b}_j$ is column j of $\mathbf{B}$, a probability vector. By (a), $\mathbf{A}\mathbf{b}_j$ is a probability vector, so $\mathbf{A}\mathbf{B}$ is stochastic. By induction on k , $\mathbf{A}^k$ is stochastic: $\mathbf{A}^0 = \mathbf{I}$ is, and $\mathbf{A}^{k+1} = \mathbf{A}^k \mathbf{A}$.
- $\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top$ with $\mathbf{1} \neq \mathbf{0}$ is Definition 9.10.1 with $\lambda = 1$, and Proposition 9.10.5 gives $1 \in \text{spec}(\mathbf{A})$. ■

Warning. The all-ones vector is a left eigenvector, not a right one. For $\mathbf{A} = \begin{pmatrix} 9/10 & 1/5 \\ 1/10 & 4/5 \end{pmatrix}$, $\mathbf{A}\mathbf{1} = (11/10, 9/10) \neq \mathbf{1}$. The right eigenvectors for 1 are the multiples of $(2, 1)$, and they carry the interesting information: the long-run probabilities. Only when the

rows also sum to 1 is $\mathbf{1}$ a right eigenvector.

9.11.4 Markov chains

Definition 9.11.2. Markov Chain, Steady State

Let $\mathbf{A} \in M_n(\mathbb{R})$ be stochastic. The **Markov chain** with **transition matrix** $\mathbf{A}$ and initial probability vector $\mathbf{x}_0$ is the sequence of probability vectors $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$, that is, $\mathbf{x}_k = \mathbf{A}^k \mathbf{x}_0$. A **steady state** of $\mathbf{A}$ is a probability vector $\mathbf{v}$ with $\mathbf{A}\mathbf{v} = \mathbf{v}$.

Each $\mathbf{x}_k$ is a probability vector by [Proposition 9.11.3](#). A steady state is a right eigenvector for the eigenvalue 1, scaled so that its entries add up to 1 and required to have nonnegative entries; if the chain starts there, it stays there.

To talk about the long run we need limits, and we use the simplest kind. A sequence of complex matrices $\mathbf{M}_k \in M_{m \times n}(\mathbb{C})$ **converges entrywise** to $\mathbf{M}$ if, for every position (r, s) , the number sequence $(\mathbf{M}_k)_{rs}$ converges to $\mathbf{M}_{rs}$. We use four facts about sequences of complex numbers from calculus: sums and constant multiples of convergent sequences converge to the sums and multiples of the limits; $z_k \rightarrow 0$ if and only if $|z_k| \rightarrow 0$; $r^k \rightarrow 0$ for real $0 \leq r < 1$; and a convergent sequence of real numbers ≥ 0 has a real limit ≥ 0 . In particular, if $|\lambda| < 1$, then $|\lambda^k| = |\lambda|^k \rightarrow 0$ (Theorem 0.6.2), so $\lambda^k \rightarrow 0$.

Example. A Two-State Chain

A machine is either working (state 1) or under repair (state 2) each day. A working machine is still working the next day with probability 9/10; a machine under repair is back to work the next day with probability 1/5. So the transition matrix is

$$\mathbf{A} = \begin{pmatrix} 9/10 & 1/5 \\ 1/10 & 4/5 \end{pmatrix}.$$

Find $\mathbf{A}^k$, its entrywise limit, and the long-run probability that the machine is working, if it is working on day 0.

Solution. Eigenvalues. $\text{tr } \mathbf{A} = 17/10$ and $\det \mathbf{A} = 36/50 - 1/50 = 7/10$, so, by the 2×2 formula $p_{\mathbf{A}} = x^2 - (\text{tr } \mathbf{A})x + \det \mathbf{A}$ (the examples after Definition 7.10.1), $p_{\mathbf{A}} = x^2 - \frac{17}{10}x + \frac{7}{10} = (x-1)(x-\frac{7}{10})$. The eigenvalue 1 was guaranteed by [Proposition 9.11.3](#).

Eigenvectors. $\mathbf{A} - \mathbf{I} = \begin{pmatrix} -1/10 & 1/5 \\ 1/10 & -1/5 \end{pmatrix}$ gives $x = 2y$, eigenvector $(2, 1)$. $\mathbf{A} - \frac{7}{10}\mathbf{I} = \begin{pmatrix} 1/5 & 1/5 \\ 1/10 & 1/10 \end{pmatrix}$ gives $y = -x$, eigenvector $(1, -1)$. The eigenvalues are distinct, so

$$\mathbf{P} = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{D} = \text{diag}(1, \frac{7}{10}), \quad \mathbf{P}^{-1} = \frac{1}{3} \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix},$$

using Theorem 0.9.8 with $\det \mathbf{P} = -3$.

Powers. By Theorem 9.4.3, writing $t = (7/10)^k$,

$$\mathbf{A}^k = \mathbf{P} \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix} \mathbf{P}^{-1} = \frac{1}{3} \begin{pmatrix} 2 & t \\ 1 & -t \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2+t & 2-2t \\ 1-t & 1+2t \end{pmatrix}.$$

Check: for $k = 1$, $t = 7/10$ and $\frac{1}{3} \begin{pmatrix} 27/10 & 6/10 \\ 3/10 & 24/10 \end{pmatrix} = \mathbf{A}$; each column of $\mathbf{A}^k$ sums to 1, as Proposition 9.11.3 requires.

The limit. Since $0 \leq 7/10 < 1$, $t \rightarrow 0$, so entrywise

$$\mathbf{A}^k \rightarrow \frac{1}{3} \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

Starting from $\mathbf{x}_0 = \mathbf{e}_1$, $\mathbf{x}_k = \mathbf{A}^k \mathbf{e}_1 = (\frac{2+t}{3}, \frac{1-t}{3}) \rightarrow (\frac{2}{3}, \frac{1}{3})$. In the long run the machine works on two days out of three. The limit does not depend on the start: for any probability vector $\mathbf{x}_0 = (s, 1-s)$, $\mathbf{A}^k \mathbf{x}_0 \rightarrow \frac{1}{3}(2s + 2(1-s), s + (1-s)) = (\frac{2}{3}, \frac{1}{3})$. The vector $\mathbf{v} = (\frac{2}{3}, \frac{1}{3})$ is the steady state, the eigenvector $(2, 1)$ scaled to sum 1.

The example shows the general mechanism. The eigenvalue 1 survives in $\mathbf{D}^k$, all others fade, and the surviving rank-one piece is (right eigenvector for 1) times (left eigenvector for 1), which for a stochastic matrix is the steady state times $\mathbf{1}^\top$.

Theorem 9.11.4. Long-Run Behavior of a Diagonalizable Markov Chain

Let $\mathbf{A} \in M_n(\mathbb{R})$, $n \geq 1$, be stochastic. Suppose that $\mathbf{A}$ is diagonalizable over $\mathbb{C}$, that $a_{\mathbf{A}}(1) = 1$, and that every other eigenvalue $\lambda \in \mathbb{C}$ of $\mathbf{A}$ has $|\lambda| < 1$. Then:

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

Idea. Use the expansion $\mathbf{A}^k = \sum_i \lambda_i^k \mathbf{v}_i \mathbf{y}_i^\top$ of Section 10. The terms with $|\lambda_i| < 1$ tend to $\mathbf{O}$, which leaves the single term for $\lambda_1 = 1$. Its left eigenvector $\mathbf{y}_1$ must be a multiple of $\mathbf{1}$, because the left eigenvectors for 1 form a line and $\mathbf{1}$ is one of them. Rescaling turns $\mathbf{v}_1 \mathbf{y}_1^\top$ into $\mathbf{v}\mathbf{1}^\top$, and the nonnegativity of $\mathbf{v}$ comes for free as a limit of nonnegative entries.

Proof. Since $\mathbf{A}$ is diagonalizable over $\mathbb{C}$, there are an invertible $\mathbf{P} \in M_n(\mathbb{C})$ and $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$ with $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. A diagonal matrix is upper triangular, so each eigenvalue occurs on the diagonal of $\mathbf{D}$ as often as its algebraic multiplicity (Theorem 9.7.4 (b)); in particular 1 occurs exactly once, because $a_{\mathbf{A}}(1) = 1$. Permuting the columns of $\mathbf{P}$ and the diagonal entries of $\mathbf{D}$ together leaves $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ intact, so we may assume $\lambda_1 = 1$, and then $|\lambda_i| < 1$ for $i \geq 2$ by hypothesis. Let $\mathbf{v}_i$ be the columns of $\mathbf{P}$ and $\mathbf{y}_i^\top$ the rows of $\mathbf{P}^{-1}$. By Proposition 9.10.7, for every k ,

$$\mathbf{A}^k = \mathbf{v}_1 \mathbf{y}_1^\top + \sum_{i=2}^n \lambda_i^k \mathbf{v}_i \mathbf{y}_i^\top.$$

Each entry of the sum is $\sum_{i \geq 2} \lambda_i^k z_i$ for fixed complex numbers z_i , and $\lambda_i^k \rightarrow 0$ for $i \geq 2$. Hence $\mathbf{A}^k \rightarrow \mathbf{L} = \mathbf{v}_1 \mathbf{y}_1^\top$ entrywise.

The shape of $\mathbf{L}$. By Proposition 9.10.7, $\mathbf{y}_1$ is a left eigenvector for 1, and so is $\mathbf{1}$ (Proposition 9.11.3). Both lie in $E_1(\mathbf{A}^\top) \subseteq \mathbb{C}^n$ (Proposition 9.10.5 (a)), which has dimension $g_{\mathbf{A}^\top}(1) = g_{\mathbf{A}}(1) \leq a_{\mathbf{A}}(1) = 1$ by Proposition 9.10.5 (b) and Theorem 9.3.1. Since $\mathbf{1} \neq \mathbf{0}$, $\mathbf{y}_1 = c\mathbf{1}$ for some $c \in \mathbb{C}$. Put $\mathbf{v} = c\mathbf{v}_1$. Then $\mathbf{L} = \mathbf{v}\mathbf{1}^\top$, $\mathbf{A}\mathbf{v} = c\mathbf{A}\mathbf{v}_1 = \mathbf{v}$, and $\mathbf{1}^\top \mathbf{v} = \mathbf{y}_1^\top \mathbf{v}_1 = 1$ by Proposition 9.10.7 (b).

$\mathbf{v}$ is a steady state. Column 1 of $\mathbf{L} = \mathbf{v}\mathbf{1}^\top$ is $\mathbf{v}$. It is the limit of column 1 of $\mathbf{A}^k$, and $\mathbf{A}^k$ is stochastic (Proposition 9.11.3), so every entry of $\mathbf{v}$ is a limit of real numbers ≥ 0 , hence real and ≥ 0 . With $\mathbf{1}^\top \mathbf{v} = 1$, $\mathbf{v}$ is a probability vector, and $\mathbf{A}\mathbf{v} = \mathbf{v}$.

Uniqueness. If $\mathbf{w}$ is a steady state, then $\mathbf{w} \in E_1(\mathbf{A})$, a subspace of $\mathbb{C}^n$ of dimension $g_{\mathbf{A}}(1) \leq 1$ that contains $\mathbf{v} \neq \mathbf{0}$. So $\mathbf{w} = s\mathbf{v}$ for some $s \in \mathbb{C}$, and $1 = \mathbf{1}^\top \mathbf{w} = s\mathbf{1}^\top \mathbf{v} = s$. Hence $\mathbf{w} = \mathbf{v}$. This proves (a) and (b).

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

■

In practice: find the eigenvalues, check the hypotheses, and compute $\mathbf{v}$ by solving $(\mathbf{A} - \mathbf{I})\mathbf{v} = \mathbf{0}$ and scaling. Then (c) gives the long run **without** computing $\mathbf{P}^{-1}$ at all.

Example. Bicycles at Three Stations

A shared bicycle is parked each night at station 1, 2, or 3. The transition matrix is

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

for instance a bicycle at station 3 stays there with probability 3/4 and never goes to station 1 in one night. Find $\mathbf{A}^k$, its limit, and the long-run distribution of the bicycle.

Solution. Eigenvectors. We check three candidates:

$$\mathbf{A} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix}, \quad \mathbf{A} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix}, \quad \mathbf{A} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

So $1, \frac{1}{2}, \frac{1}{4}$ are eigenvalues, with eigenvectors $\mathbf{v}_1 = (1, 2, 3)$, $\mathbf{v}_2 = (-1, 0, 1)$, $\mathbf{v}_3 = (-1, 1, 0)$. A 3×3 matrix has at most three eigenvalues (Corollary 9.2.3), so these are all; they are distinct, so $\mathbf{A}$ is diagonalizable (Corollary 9.4.2) and $a_{\mathbf{A}}(1) = 1$. As a check, $\text{tr } \mathbf{A} = \frac{7}{4} = 1 + \frac{1}{2} + \frac{1}{4}$. The hypotheses of Theorem 9.11.4 hold.

The expansion. With $\mathbf{P} = \begin{pmatrix} 1 & -1 & -1 \\ 2 & 0 & 1 \\ 3 & 1 & 0 \end{pmatrix}$, whose determinant is -6 ,

$$\mathbf{P}^{-1} = \begin{pmatrix} 1/6 & 1/6 & 1/6 \\ -1/2 & -1/2 & 1/2 \\ -1/3 & 2/3 & -1/3 \end{pmatrix},$$

as multiplying out $\mathbf{P}^{-1}\mathbf{P} = \mathbf{I}$ confirms. The first row is $\frac{1}{6}\mathbf{1}^\top$, a multiple of $\mathbf{1}^\top$ as the proof of the theorem predicts. By Proposition 9.10.7,

$$\mathbf{A}^k = \frac{1}{6} \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{pmatrix} + \frac{1}{2^k} \cdot \frac{1}{2} \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ -1 & -1 & 1 \end{pmatrix} + \frac{1}{4^k} \cdot \frac{1}{3} \begin{pmatrix} 1 & -2 & 1 \\ -1 & 2 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

For example, the $(1, 1)$ -entry of $\mathbf{A}^k$ is $\frac{1}{6} + \frac{1}{2} \cdot 2^{-k} + \frac{1}{3} \cdot 4^{-k}$, which is 1 for $k = 0$ and $\frac{1}{6} + \frac{1}{4} + \frac{1}{12} = \frac{1}{2}$ for $k = 1$, matching $a_{11} = 2/4$.

The limit. As $k \rightarrow \infty$, $2^{-k} \rightarrow 0$ and $4^{-k} \rightarrow 0$, so

$$\mathbf{A}^k \rightarrow \frac{1}{6} \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{pmatrix} = \mathbf{v}\mathbf{1}^\top, \quad \mathbf{v} = \left(\frac{1}{6}, \frac{1}{3}, \frac{1}{2}\right).$$

Wherever the bicycle starts, in the long run it is at station 1, 2, 3 with probabilities $\frac{1}{6}, \frac{1}{3}, \frac{1}{2}$. Directly, $\mathbf{v} = \frac{1}{6}\mathbf{v}_1$ and $\mathbf{A}\mathbf{v} = \mathbf{v}$. The slowest-fading term is the one with $\frac{1}{2}$: each entry's gap to its limit roughly halves every night.

Warning. A stochastic matrix need not have convergent powers. The swap $\mathbf{S} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is stochastic, and $\mathbf{S}^k$ is $\mathbf{I}$ for even k and $\mathbf{S}$ for odd k , so $\mathbf{S}^k$ does not converge. Starting from $\mathbf{x}_0 = \mathbf{e}_1$, the chain alternates $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1, \dots$ forever, although $\mathbf{S}$ has the steady state $(\frac{1}{2}, \frac{1}{2})$. The theorem does not apply: $\mathbf{S}$ has the eigenvalue -1 , with $|-1| = 1$. At the other extreme, $\mathbf{I}_2$ has convergent powers, but $a(1) = 2$, every probability vector is a steady state, and the limit of $\mathbf{x}_k$ depends on $\mathbf{x}_0$.

Chapter 19 proves the Perron–Frobenius theorem, which gives the conclusions of [Theorem 9.11.4](#) for every stochastic matrix with some power having all entries positive, without assuming diagonalizability and without computing eigenvalues. The exercises prove the first ingredient: no eigenvalue of a stochastic matrix has $|\lambda| > 1$.

Quick check. Is $\mathbf{A} = \begin{pmatrix} 1/2 & 1/3 \\ 1/2 & 2/3 \end{pmatrix}$ stochastic? If so, find its steady state.

Solution. Yes: the entries are nonnegative and the columns sum to 1. From $(\mathbf{A} - \mathbf{I})\mathbf{v} = \mathbf{0}$, the first row gives $-\frac{1}{2}v_1 + \frac{1}{3}v_2 = 0$, so $v_2 = \frac{3}{2}v_1$. Scaling to sum 1: $v_1 + \frac{3}{2}v_1 = 1$ gives $\mathbf{v} = (\frac{2}{5}, \frac{3}{5})$.

(The eigenvalues are 1 and $\text{tr } \mathbf{A} - 1 = \frac{1}{6}$; they are distinct, so $\mathbf{A}$ is diagonalizable with $a_{\mathbf{A}}(1) = 1$, and $\mathbf{v}$ is the only steady state and the long-run distribution, by [Theorem 9.11.4](#).)

9.11.5 A population model

Not every matrix of this kind is stochastic. In a population divided into age classes, individuals are born, age, and die, and the total need not be preserved. A **Leslie matrix** records, in its first row, the average number of offspring per individual of each age class, and just below the diagonal, the fraction surviving into the next class. The population vector evolves by $\mathbf{x}_{k+1} = \mathbf{L}\mathbf{x}_k$, and [Theorem 9.11.1](#) describes the long run.

Example. Three Age Classes

A species is counted once a year in three age classes: young, juvenile, and adult. Half of the young survive to become juveniles, and half of the juveniles survive to become adults; no adult survives another year. Each juvenile has on average $7/2$ offspring and each adult 3. The Leslie matrix is

$$\mathbf{L} = \begin{pmatrix} 0 & 7/2 & 3 \\ 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \end{pmatrix}.$$

Starting from $\mathbf{x}_0 = (20, 10, 5)$ (in hundreds of animals), find $\mathbf{x}_k$ and describe the long-run growth and age distribution.

Solution. Characteristic polynomial. Expanding $\det(x\mathbf{I} - \mathbf{L})$ along the first column,

$$\begin{aligned} p_{\mathbf{L}} &= x \det \begin{pmatrix} x & 0 \\ -1/2 & x \end{pmatrix} + \frac{1}{2} \det \begin{pmatrix} -7/2 & -3 \\ -1/2 & x \end{pmatrix} \\ &= x^3 + \frac{1}{2} \left(-\frac{7}{2}x - \frac{3}{2} \right) = x^3 - \frac{7}{4}x - \frac{3}{4}. \end{aligned}$$

Trying simple values, $p_{\mathbf{L}}(-1) = -1 + \frac{7}{4} - \frac{3}{4} = 0$, and dividing, $p_{\mathbf{L}} = (x + 1)(x^2 - x - \frac{3}{4}) = (x + 1)(x - \frac{3}{2})(x + \frac{1}{2})$. The eigenvalues $\frac{3}{2}, -1, -\frac{1}{2}$ are distinct, so $\mathbf{L}$ is diagonalizable.

Eigenvectors. $\mathbf{L}(w_1, w_2, w_3) = \lambda(w_1, w_2, w_3)$ forces $\frac{1}{2}w_1 = \lambda w_2$ and $\frac{1}{2}w_2 = \lambda w_3$, so taking $w_3 = 1$ gives $w_2 = 2\lambda$, $w_1 = 4\lambda^2$. Hence, after scaling,

$$\lambda = \frac{3}{2} : (9, 3, 1), \quad \lambda = -1 : (4, -2, 1), \quad \lambda = -\frac{1}{2} : (1, -1, 1).$$

For instance $\mathbf{L}(9, 3, 1) = (\frac{21}{2} + 3, \frac{9}{2}, \frac{3}{2}) = \frac{3}{2}(9, 3, 1)$.

The initial vector. Solving $c_1(9, 3, 1) + c_2(4, -2, 1) + c_3(1, -1, 1) = (20, 10, 5)$ gives $c_1 = 3$, $c_2 = -3$, $c_3 = 5$: indeed $(27 - 12 + 5, 9 + 6 - 5, 3 - 3 + 5) = (20, 10, 5)$. By [Theorem 9.11.1](#),

$$\mathbf{x}_k = 3 \left(\frac{3}{2} \right)^k \begin{pmatrix} 9 \\ 3 \\ 1 \end{pmatrix} - 3(-1)^k \begin{pmatrix} 4 \\ -2 \\ 1 \end{pmatrix} + 5 \left(-\frac{1}{2} \right)^k \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Check. For $k = 1$: $(\frac{81}{2} + 12 - \frac{5}{2}, \frac{27}{2} - 6 + \frac{5}{2}, \frac{9}{2} + 3 - \frac{5}{2}) = (50, 10, 5)$, and directly $\mathbf{L}\mathbf{x}_0 = (35 + 15, 10, 5)$.

The long run. Divide by $(3/2)^k$:

$$\left(\frac{2}{3}\right)^k \mathbf{x}_k = 3 \begin{pmatrix} 9 \\ 3 \\ 1 \end{pmatrix} - 3 \left(-\frac{2}{3}\right)^k \begin{pmatrix} 4 \\ -2 \\ 1 \end{pmatrix} + 5 \left(-\frac{1}{3}\right)^k \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 27 \\ 9 \\ 3 \end{pmatrix},$$

since $|\frac{2}{3}| < 1$ and $|\frac{1}{3}| < 1$. So the population eventually grows by the factor $\frac{3}{2}$ each year, and the age classes approach the proportions $9 : 3 : 1$, the entries of the eigenvector for the largest eigenvalue. The eigenvalue -1 makes the population oscillate from year to year, but its effect is bounded while the dominant term grows, so its share fades.

9.11.6 Exercises

9.11.6.1 A. Check your understanding

Exercise. A1

- Define a stochastic matrix, and say which convention (rows or columns) this book uses.
- True or false: for every stochastic $\mathbf{A}$, the powers $\mathbf{A}^k$ converge entrywise. Justify your answer.
- True or false: $\mathbf{1}$ is a right eigenvector of every stochastic matrix. Justify your answer.
- True or false: every stochastic matrix is invertible. Justify your answer.
- Write down the general solution of $a_{k+2} = 5a_{k+1} - 6a_k$ over $\mathbb{R}$.
- In [Theorem 9.11.1](#), what are the numbers $c_1, \dots, c_n$?

Solution.

- $\mathbf{A} \in M_n(\mathbb{R})$ is stochastic if all entries are nonnegative and every **column** sums to 1 ([Definition 9.11.1](#)). This book uses columns, with $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$.
- False. The swap $\mathbf{S} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ has $\mathbf{S}^k$ alternating between $\mathbf{I}$ and $\mathbf{S}$.
- False. $\mathbf{A} = \begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix}$ is stochastic, and $\mathbf{A}\mathbf{1} = (3/2, 1/2)$ is not a multiple of $\mathbf{1}$. It is a **left** eigenvector ([Proposition 9.11.3](#)).
- False. $\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ is stochastic with determinant 0.
- $p = x^2 - 5x + 6 = (x-2)(x-3)$, so by [Theorem 9.11.2](#) the solutions are $a_k = b_1 2^k + b_2 3^k$ with $b_1, b_2 \in \mathbb{R}$.
- The coordinates of $\mathbf{x}_0$ in the basis of eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$: $(c_1, \dots, c_n) = \mathbf{P}^{-1}\mathbf{x}_0$.

9.11.6.2 B. Practice

Exercise. B1: Two recurrences

- Solve $a_{k+2} = a_{k+1} + 2a_k$ with $a_0 = 0$, $a_1 = 3$.

- (b) Solve $a_{k+2} = 6a_{k+1} - 9a_k$ with $a_0 = 1$, $a_1 = 6$, by the method of [Example \(A Repeated Root\)](#).

Solution.

- (a) $p = x^2 - x - 2 = (x - 2)(x + 1)$ has distinct roots, so by [Theorem 9.11.2](#), $a_k = b_1 2^k + b_2 (-1)^k$. From $a_0 = b_1 + b_2 = 0$ and $a_1 = 2b_1 - b_2 = 3$, $3b_1 = 3$, so $b_1 = 1$, $b_2 = -1$. Hence $a_k = 2^k - (-1)^k$. Check: $a_2 = 3 + 0 = 3 = 4 - 1$, $a_3 = 3 + 6 = 9 = 8 + 1$.
- (b) The window matrix is $\mathbf{C} = \begin{pmatrix} 0 & 1 \\ -9 & 6 \end{pmatrix}$, with $p = (x - 3)^2$. Put $\mathbf{N} = \mathbf{C} - 3\mathbf{I} = \begin{pmatrix} -3 & 1 \\ -9 & 3 \end{pmatrix}$; then $\mathbf{N}^2 = \begin{pmatrix} 9 - 9 & -3 + 3 \\ 27 - 27 & -9 + 9 \end{pmatrix} = \mathbf{O}$. As in the example, by induction $\mathbf{C}^k = 3^k \mathbf{I} + k 3^{k-1} \mathbf{N}$: the step is $(3^k \mathbf{I} + k 3^{k-1} \mathbf{N})(3\mathbf{I} + \mathbf{N}) = 3^{k+1} \mathbf{I} + 3^k \mathbf{N} + k 3^k \mathbf{N}$. Row 1 of $\mathbf{C}^k$ is $(3^k - 3k 3^{k-1}, k 3^{k-1}) = ((1 - k)3^k, k 3^{k-1})$, so

$$a_k = (1 - k)3^k \cdot 1 + k 3^{k-1} \cdot 6 = (1 - k + 2k)3^k = (1 + k)3^k.$$

Check: $a_2 = 36 - 9 = 27 = 3 \cdot 9$, $a_3 = 6 \cdot 27 - 9 \cdot 6 = 108 = 4 \cdot 27$.

Exercise. B2: A three-state chain

Let

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

- (a) Verify that $\mathbf{A}$ is stochastic, and find its steady state.
- (b) Verify that $(-1, 0, 1)$ and $(1, -2, 1)$ are eigenvectors of $\mathbf{A}$, and find their eigenvalues.
- (c) Hence find the entrywise limit of $\mathbf{A}^k$, and the limit of the chain started at $\mathbf{x}_0 = \mathbf{e}_2$.

Solution.

- (a) All entries are nonnegative; the columns sum to $\frac{1}{2} + \frac{1}{4} + \frac{1}{4} = 1$, $\frac{1}{2} + 0 + \frac{1}{2} = 1$, and $\frac{1}{4} + \frac{1}{4} + \frac{1}{2} = 1$. For the steady state, solve $(\mathbf{A} - \mathbf{I})\mathbf{v} = \mathbf{0}$:

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

Multiplying the rows by 4: $-2v_1 + 2v_2 + v_3 = 0$, $v_1 - 4v_2 + v_3 = 0$, $v_1 + 2v_2 - 2v_3 = 0$. Subtracting the second from the third gives $6v_2 - 3v_3 = 0$, so $v_3 = 2v_2$; then the second gives $v_1 = 4v_2 - 2v_2 = 2v_2$, and the first holds: $-4v_2 + 2v_2 + 2v_2 = 0$. So $E_1(\mathbf{A}) = \text{span}((2, 1, 2))$, and scaling to sum 1, the steady state is $\mathbf{v} = (\frac{2}{5}, \frac{1}{5}, \frac{2}{5})$.

- (b) $\mathbf{A}(-1, 0, 1) = (-\frac{1}{2} + \frac{1}{4}, -\frac{1}{4} + \frac{1}{4}, -\frac{1}{4} + \frac{1}{2}) = (-\frac{1}{4}, 0, \frac{1}{4}) = \frac{1}{4}(-1, 0, 1)$, eigenvalue $\frac{1}{4}$.
 $\mathbf{A}(1, -2, 1) = (\frac{1}{2} - 1 + \frac{1}{4}, \frac{1}{4} + \frac{1}{4}, \frac{1}{4} - 1 + \frac{1}{2}) = (-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}) = -\frac{1}{4}(1, -2, 1)$, eigenvalue $-\frac{1}{4}$.
- (c) The eigenvalues $1, \frac{1}{4}, -\frac{1}{4}$ are distinct, so they are all the eigenvalues of $\mathbf{A}$, $\mathbf{A}$ is diag-

onalizable (Corollary 9.4.2), $a_{\mathbf{A}}(1) = 1$, and the other two have absolute value $\frac{1}{4} < 1$. By Theorem 9.11.4,

$$\mathbf{A}^k \rightarrow \mathbf{v}\mathbf{1}^\top = \frac{1}{5} \begin{pmatrix} 2 & 2 & 2 \\ 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix},$$

and every chain, in particular the one started at $\mathbf{e}_2$, converges to $\mathbf{v} = (\frac{2}{5}, \frac{1}{5}, \frac{2}{5})$.

Exercise. B3: The limit of a matrix power

Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1/2 & 3/2 \end{pmatrix} \in M_2(\mathbb{R})$.

- Diagonalize $\mathbf{A}$, and find a formula for $\mathbf{A}^k$.
- Find the entrywise limit of $\mathbf{A}^k$.
- Hence find $\lim_{k \rightarrow \infty} a_k$ for the sequence with $a_{k+2} = \frac{3}{2}a_{k+1} - \frac{1}{2}a_k$, in terms of a_0 and a_1 .

Solution.

- $\text{tr } \mathbf{A} = \frac{3}{2}$ and $\det \mathbf{A} = \frac{1}{2}$, so $p_{\mathbf{A}} = x^2 - \frac{3}{2}x + \frac{1}{2} = (x-1)(x-\frac{1}{2})$. $\mathbf{A} - \mathbf{I} = \begin{pmatrix} -1 & 1 \\ -1/2 & 1/2 \end{pmatrix}$ gives the eigenvector $(1, 1)$; $\mathbf{A} - \frac{1}{2}\mathbf{I} = \begin{pmatrix} -1/2 & 1 \\ -1/2 & 1 \end{pmatrix}$ gives $(2, 1)$. So $\mathbf{P} = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$, $\mathbf{D} = \text{diag}(1, \frac{1}{2})$, $\mathbf{P}^{-1} = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix}$ ($\det \mathbf{P} = -1$, Theorem 0.9.8). By Theorem 9.4.3, with $t = 2^{-k}$,

$$\mathbf{A}^k = \begin{pmatrix} 1 & 2t \\ 1 & t \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} -1+2t & 2-2t \\ -1+t & 2-t \end{pmatrix}.$$

For $k = 1$, $t = \frac{1}{2}$ gives $\begin{pmatrix} 0 & 1 \\ -1/2 & 3/2 \end{pmatrix} = \mathbf{A}$.

- Since $t = 2^{-k} \rightarrow 0$, $\mathbf{A}^k \rightarrow \begin{pmatrix} -1 & 2 \\ -1 & 2 \end{pmatrix}$ entrywise.
- $\mathbf{A}$ is the window matrix of this recurrence, so $(a_k, a_{k+1}) = \mathbf{A}^k(a_0, a_1)$ and $a_k = (-1+2t)a_0 + (2-2t)a_1$. Hence $a_k \rightarrow 2a_1 - a_0$. (For instance $a_0 = 0$, $a_1 = 1$ gives $0, 1, \frac{3}{2}, \frac{7}{4}, \dots \rightarrow 2$.)

9.11.6.3 C. Going deeper

Exercise. C1: Eigenvalues of a stochastic matrix

Let $\mathbf{A} \in M_n(\mathbb{R})$ be stochastic, regarded as a complex matrix, and let $\lambda \in \mathbb{C}$ be an eigenvalue of $\mathbf{A}$.

- Prove that $|\lambda| \leq 1$.
- Refine the argument: prove that $|\lambda - a_{ii}| \leq 1 - a_{ii}$ for some index i .

- (c) Deduce that if every diagonal entry of $\mathbf{A}$ is positive, then -1 is not an eigenvalue of $\mathbf{A}$.
- (d) Show by an example that (c) fails if a diagonal entry is 0.

Hint: take a left eigenvector $\mathbf{y}$ for λ and an index i with $|y_i|$ as large as possible.

Solution.

- (a) By Proposition 9.10.5 (over $\mathbb{C}$), there is $\mathbf{y} \in \mathbb{C}^n$, $\mathbf{y} \neq \mathbf{0}$, with $\mathbf{A}^\top \mathbf{y} = \lambda \mathbf{y}$. Choose i with $|y_i| \geq |y_j|$ for all j ; then $|y_i| > 0$, since $\mathbf{y} \neq \mathbf{0}$. Entry i of $\mathbf{A}^\top \mathbf{y} = \lambda \mathbf{y}$ reads $\sum_j a_{ji} y_j = \lambda y_i$, because row i of $\mathbf{A}^\top$ is column i of $\mathbf{A}$. By Theorem 0.6.2 (d) and Theorem 0.6.4, applied to the n -term sum by induction on n , and since $a_{ji} \geq 0$ and column i sums to 1,

$$|\lambda| |y_i| = \left| \sum_j a_{ji} y_j \right| \leq \sum_j a_{ji} |y_j| \leq \sum_j a_{ji} |y_i| = |y_i|.$$

Dividing by $|y_i| > 0$ gives $|\lambda| \leq 1$.

- (b) With the same i , move the diagonal term across: $(\lambda - a_{ii})y_i = \sum_{j \neq i} a_{ji} y_j$. Hence, as in (a),

$$|\lambda - a_{ii}| |y_i| \leq \sum_{j \neq i} a_{ji} |y_j| \leq \left(\sum_{j \neq i} a_{ji} \right) |y_i| = (1 - a_{ii}) |y_i|,$$

and dividing by $|y_i|$ gives $|\lambda - a_{ii}| \leq 1 - a_{ii}$.

- (c) Suppose $\lambda = -1$. By (b), some i has $|-1 - a_{ii}| \leq 1 - a_{ii}$. Since $a_{ii} \geq 0$, $|-1 - a_{ii}| = 1 + a_{ii}$, so $1 + a_{ii} \leq 1 - a_{ii}$ and $a_{ii} \leq 0$. This contradicts $a_{ii} > 0$.
- (d) The swap $\mathbf{S} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is stochastic with zero diagonal, and $\mathbf{S}(1, -1) = (-1, 1)$, so -1 is an eigenvalue.

Exercise. C2: Two-state chains

Let $0 \leq p, q \leq 1$ and $\mathbf{A} = \begin{pmatrix} 1-p & q \\ p & 1-q \end{pmatrix}$.

- (a) Show that $\mathbf{A}$ is stochastic, with eigenvalues 1 and $1 - p - q$.
- (b) Suppose $0 < p + q < 2$. Prove that $\mathbf{A}^k$ converges entrywise, and find the limit and the steady state.
- (c) In particular, if $\mathbf{A}$ has no zero entry, then $\mathbf{A}^k$ converges. What happens when $p + q = 0$ and when $p + q = 2$?

Solution.

- (a) The entries lie in $[0, 1]$, and the columns sum to $(1 - p) + p = 1$ and $q + (1 - q) = 1$. Since $\text{tr } \mathbf{A} = 2 - p - q$ and $\det \mathbf{A} = (1 - p)(1 - q) - pq = 1 - p - q$, $p_{\mathbf{A}} = x^2 - (2 - p - q)x + (1 - p - q) = (x - 1)(x - (1 - p - q))$, as multiplying out confirms.
- (b) Put $r = 1 - p - q$. Since $0 < p + q < 2$, $-1 < r < 1$, so $r \neq 1$ and $|r| < 1$. The eigenvalues

$1 \neq r$ are distinct, so $\mathbf{A}$ is diagonalizable (Corollary 9.4.2). From $\mathbf{A} - \mathbf{I} = \begin{pmatrix} -p & q \\ p & -q \end{pmatrix}$, (q, p) is an eigenvector for 1, nonzero because $p + q > 0$; from $\mathbf{A} - r\mathbf{I} = \begin{pmatrix} q & q \\ p & p \end{pmatrix}$, $(1, -1)$ is one for r . So

$$\mathbf{P} = \begin{pmatrix} q & 1 \\ p & -1 \end{pmatrix}, \quad \mathbf{P}^{-1} = \frac{1}{p+q} \begin{pmatrix} 1 & 1 \\ p & -q \end{pmatrix},$$

with $\det \mathbf{P} = -(p+q) \neq 0$. By Theorem 9.4.3,

$$\mathbf{A}^k = \frac{1}{p+q} \begin{pmatrix} q & r^k \\ p & -r^k \end{pmatrix} \begin{pmatrix} 1 & 1 \\ p & -q \end{pmatrix} = \frac{1}{p+q} \begin{pmatrix} q + pr^k & q - qr^k \\ p - pr^k & p + qr^k \end{pmatrix}.$$

Since $|r| < 1$, $r^k \rightarrow 0$, and

$$\mathbf{A}^k \rightarrow \frac{1}{p+q} \begin{pmatrix} q & q \\ p & p \end{pmatrix}, \quad \mathbf{v} = \left(\frac{q}{p+q}, \frac{p}{p+q} \right).$$

The limit is $\mathbf{v}\mathbf{1}^\top$, and $\mathbf{v}$ is the steady state, in agreement with [Theorem 9.11.4](#).

- (c) If no entry is zero, then $1 - p > 0$ and $p > 0$, so $0 < p < 1$, and likewise $0 < q < 1$; hence $0 < p + q < 2$ and (b) applies. If $p + q = 0$, then $p = q = 0$ and $\mathbf{A} = \mathbf{I}$: the powers converge (to $\mathbf{I}$), but every probability vector is a steady state. If $p + q = 2$, then $p = q = 1$ and $\mathbf{A}$ is the swap, whose powers do not converge.