

10.10 Convergent and Power-Bounded Matrices

Chapter 9 asked what happens to $\mathbf{A}^m$ as m grows, and could answer only for diagonalizable matrices: the powers of $\mathbf{A}$ behave like the powers of its eigenvalues. For a Markov chain that left a gap, because a transition matrix need not be diagonalizable. The Jordan form closes the gap completely. This section characterizes, in terms of the eigenvalues and the block sizes alone, when $\mathbf{A}^m$ tends to zero, when it stays bounded, when it converges, and when the geometric series $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots$ converges.

Convergence here means what it meant in Chapter 9: **entrywise**. No notion of the size of a matrix is used. Chapter 16 introduces norms and re-proves several of these results in a way that also gives numerical bounds; the price of doing without them is that we must look at one entry at a time, and the Jordan form is what makes that possible.

10.10.1 Entrywise limits and bounds

Definition 10.10.1. Entrywise Convergence and Boundedness

Let $\mathbf{M}_1, \mathbf{M}_2, \dots$ be matrices in $M_{p \times q}(\mathbb{C})$ and $\mathbf{M} \in M_{p \times q}(\mathbb{C})$.

- (a) The sequence $(\mathbf{M}_m)$ **converges** to $\mathbf{M}$, written $\mathbf{M}_m \rightarrow \mathbf{M}$, if for **every** position (r, s) the sequence of complex numbers $(\mathbf{M}_m)_{rs}$ converges to $\mathbf{M}_{rs}$. We then write $\mathbf{M} = \lim_m \mathbf{M}_m$.
- (b) The sequence $(\mathbf{M}_m)$ is **bounded** if there is a real $c \geq 0$ with $|(\mathbf{M}_m)_{rs}| \leq c$ for **all** m and **all** positions (r, s) .
- (c) A series $\sum_{k \geq 0} \mathbf{M}_k$ **converges** to $\mathbf{M}$ if its sequence of partial sums $\sum_{k=0}^K \mathbf{M}_k$ converges to $\mathbf{M}$.

In words: nothing new is happening. Each of the pq positions is watched separately, and the words mean for matrices exactly what they mean for the finitely many number sequences at those positions. Since a finite set of bounds can be replaced by their maximum, (b) is the same as asking each entry sequence to be bounded.

Remark. What we assume from analysis. Beyond the four facts quoted in Chapter 9 §11, this section uses exactly the following.

- (B1) **Polynomial against geometric decay.** For every integer $j \geq 0$ and every real r with $0 \leq r < 1$, $m^j r^m \rightarrow 0$ as $m \rightarrow \infty$.
- (B2) **Growth.** For real $r > 1$, the sequence (r^m) is unbounded.
- (B3) **Limits.** Limits of convergent sequences of complex numbers respect sums, scalar multiples, and products; a convergent sequence is bounded; every subsequence of a convergent sequence converges to the same limit; and $z_m \rightarrow z$ implies $|z_m| \rightarrow |z|$.
- (B4) **Terms of a convergent series.** If $\sum_k z_k$ converges, then $z_k \rightarrow 0$.

Two reductions make every proof below short. The first moves the question along a similarity, so that we may assume $\mathbf{A}$ is a Jordan matrix. The second computes the powers of a single Jordan block.

Lemma 10.10.1. Convergence and Boundedness Survive a Similarity

Let $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{C})$ with $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ for an invertible $\mathbf{P} \in M_n(\mathbb{C})$, and let $\mathbf{L} \in M_n(\mathbb{C})$. Then:

- (a) $\mathbf{A}^m \rightarrow \mathbf{L}$ if and only if $\mathbf{B}^m \rightarrow \mathbf{P}^{-1}\mathbf{L}\mathbf{P}$;
- (b) $(\mathbf{A}^m)$ is bounded if and only if $(\mathbf{B}^m)$ is bounded.

Proof. By Proposition 4.2.6 (c), $\mathbf{B}^m = \mathbf{P}^{-1}\mathbf{A}^m\mathbf{P}$ for every $m \geq 0$. Writing out the products (Definition 0.9.9), the (r, s) -entry of $\mathbf{B}^m$ is

$$(\mathbf{B}^m)_{rs} = \sum_{u,v} (\mathbf{P}^{-1})_{ru} (\mathbf{A}^m)_{uv} \mathbf{P}_{vs},$$

a linear combination of the entries of $\mathbf{A}^m$ with coefficients that do not depend on m . So if $\mathbf{A}^m \rightarrow \mathbf{L}$, then by (B3) each entry of $\mathbf{B}^m$ converges to the matching combination of the entries of $\mathbf{L}$, that is, $\mathbf{B}^m \rightarrow \mathbf{P}^{-1}\mathbf{L}\mathbf{P}$; and if $|(\mathbf{A}^m)_{uv}| \leq c$ for all m, u, v , then $|(\mathbf{B}^m)_{rs}| \leq c \sum_{u,v} |(\mathbf{P}^{-1})_{ru}| |\mathbf{P}_{vs}|$, a bound independent of m , so $(\mathbf{B}^m)$ is bounded. The converses follow by exchanging the roles of $\mathbf{A}$ and $\mathbf{B}$, using $\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}$. ■

For the second reduction, recall from Lemma 10.9.1, applied to the polynomial x^m , that

$$(\mathbf{J}_k(\lambda)^m)_{r,r+j} = \binom{m}{j} \lambda^{m-j} \quad (0 \leq j \leq k-1),$$

with zeros below the diagonal. Everything in this section is read off this one formula. Two remarks about it: the binomial coefficient is bounded by m^j , since

$$\binom{m}{j} = \frac{m(m-1) \cdots (m-j+1)}{j!} \leq \frac{m^j}{j!} \leq m^j \quad (m \geq j \geq 1),$$

and $\binom{m}{1} = m$ grows without bound, while $\binom{m}{0} = 1$ is constant. So the only question is whether the geometric factor λ^m beats the polynomial factor m^j .

Quick check. Compute $|(\mathbf{J}_3(\lambda)^m)_{1,3}|$ for $\lambda = \frac{1}{2}$ and for $\lambda = 1$, and say in each case whether it tends to 0, stays bounded, or grows without bound.

Solution. Here $j = 2$, so the entry is $\binom{m}{2} \lambda^{m-2}$, of modulus $\frac{m(m-1)}{2} |\lambda|^{m-2}$. For $\lambda = \frac{1}{2}$ this is $2 \cdot \frac{m(m-1)}{2} 2^{-m} \leq 2m^2 2^{-m}$, which tends to 0 by (B1) with $j = 2$ and $r = \frac{1}{2}$. For $\lambda = 1$ it is $\frac{m(m-1)}{2}$, which grows without bound. The modulus of λ decides, and when $|\lambda| = 1$ the block size decides.

10.10.2 Powers tending to zero

The powers of $\mathbf{A}$ die out exactly when every eigenvalue has modulus less than one.

Theorem 10.10.2. Powers Tending to Zero

Let $\mathbf{A} \in M_n(\mathbb{C})$, $n \geq 1$. Then $\mathbf{A}^m \rightarrow \mathbf{O}$ if and only if every eigenvalue λ of $\mathbf{A}$ satisfies $|\lambda| < 1$.

Idea. ($\Leftarrow$) Pass to the Jordan form and look at one entry: it is $\binom{m}{j}\lambda^{m-j}$, a polynomial in m times a geometric sequence that decays, and geometric decay wins. ($\Rightarrow$) The other direction needs no Jordan form at all: an eigenvector turns $\mathbf{A}^m$ into λ^m .

Proof. ($\Leftarrow$) Suppose $|\lambda| < 1$ for every $\lambda \in \text{spec}(\mathbf{A})$. By Theorem 10.3.1, $\mathbf{A} = \mathbf{PJP}^{-1}$ with $\mathbf{J} = \mathbf{J}_{k_1}(\lambda_1) \oplus \cdots \oplus \mathbf{J}_{k_d}(\lambda_d)$ a Jordan matrix, whose diagonal entries λ_i are the eigenvalues of $\mathbf{A}$ (Corollary 10.3.2 (a)). By Theorem 8.1.4 (b), $\mathbf{J}^m$ is the direct sum of the blocks $\mathbf{J}_{k_i}(\lambda_i)^m$, so each entry of $\mathbf{J}^m$ is either 0 for every m or of the form $\binom{m}{j}\lambda_i^{m-j}$ with $0 \leq j \leq k_i - 1$. For the latter, with $r = |\lambda_i| < 1$, if $r = 0$ the entry is 0 for $m > j$; otherwise, for $m \geq j$,

$$\left| \binom{m}{j} \lambda_i^{m-j} \right| \leq m^j r^{m-j} = r^{-j} m^j r^m \rightarrow 0$$

by (B1), the factor r^{-j} being a constant. So $\mathbf{J}^m \rightarrow \mathbf{O}$, and $\mathbf{A}^m \rightarrow \mathbf{POP}^{-1} = \mathbf{O}$ by Lemma 10.10.1 (a).

($\Rightarrow$) Suppose $\mathbf{A}^m \rightarrow \mathbf{O}$ and let $\lambda \in \text{spec}(\mathbf{A})$ with eigenvector $\mathbf{v} \neq \mathbf{0}$. Then $\mathbf{A}^m \mathbf{v} = \lambda^m \mathbf{v}$ for every m (induction on m). Pick a coordinate r with $v_r \neq 0$. The r -th entry of $\mathbf{A}^m \mathbf{v}$ is a linear combination of entries of $\mathbf{A}^m$ with coefficients $v_1, \dots, v_n$, so it tends to 0 by (B3); that entry is $\lambda^m v_r$, so $\lambda^m \rightarrow 0$ and hence $|\lambda|^m = |\lambda^m| \rightarrow 0$ by (B3). If $|\lambda| \geq 1$, then $|\lambda|^m \geq 1$ for every m and the limit cannot be 0. Hence $|\lambda| < 1$. This proves the theorem. ■

Note the asymmetry: the hard direction needs the Jordan form, the easy one needs only an eigenvector. That pattern repeats in the next two theorems.

Warning. **The condition is on the eigenvalues, not on the entries.** Every entry of $\mathbf{A} = \begin{pmatrix} 0.9 & 0.9 \\ 0.9 & 0.9 \end{pmatrix}$ has modulus less than 1, yet $\mathbf{A}\mathbf{1} = 1.8\mathbf{1}$, so 1.8 is an eigenvalue and $\mathbf{A}^m \mathbf{1} = 1.8^m \mathbf{1}$ grows without bound. Conversely, a matrix with huge entries can have all eigenvalues 0: the matrix $\begin{pmatrix} 0 & 10^6 \\ 0 & 0 \end{pmatrix}$ squares to $\mathbf{O}$. Nothing can be concluded from the size of the entries alone.

10.10.3 Powers that stay bounded

Dropping “tends to zero” to “stays bounded” allows eigenvalues on the unit circle, but only if they carry no chains: a block of size ≥ 2 contributes an entry $\binom{m}{1}\lambda^{m-1}$ of modulus m .

Theorem 10.10.3. Power-Bounded Matrices

Let $\mathbf{A} \in M_n(\mathbb{C})$, $n \geq 1$, with Jordan form $\mathbf{J}$. The sequence $(\mathbf{A}^m)_{m \geq 0}$ is bounded if and only if

- (i) every eigenvalue λ of $\mathbf{A}$ satisfies $|\lambda| \leq 1$, and
- (ii) every eigenvalue λ with $|\lambda| = 1$ has **all** of its blocks in $\mathbf{J}$ of size 1, equivalently $a_{\mathbf{A}}(\lambda) = g_{\mathbf{A}}(\lambda)$, equivalently $x - \lambda$ occurs to the first power in $m_{\mathbf{A}}$.

Proof. Write $\mathbf{A} = \mathbf{PJP}^{-1}$ with $\mathbf{J} = \mathbf{J}_{k_1}(\lambda_1) \oplus \cdots \oplus \mathbf{J}_{k_d}(\lambda_d)$ (Theorem 10.3.1). By Lemma 10.10.1 (b), $(\mathbf{A}^m)$ is bounded if and only if $(\mathbf{J}^m)$ is, and by Theorem 8.1.4 (b) the latter holds if and only if every block sequence $(\mathbf{J}_{k_i}(\lambda_i)^m)$ is bounded. The three reformulations in (ii) are Corollary 10.3.2 (a), (b), and (c): all λ -blocks have size 1 exactly when their number equals their total size, and exactly when the largest of their sizes is 1.

($\Leftarrow$) Assume (i) and (ii), and fix a block $\mathbf{J}_k(\lambda)$ of $\mathbf{J}$. If $|\lambda| < 1$, then $\mathbf{J}_k(\lambda)^m \rightarrow \mathbf{O}$ by Theorem 10.10.2 applied to the single block $\mathbf{J}_k(\lambda)$, whose only eigenvalue is λ , hence $(\mathbf{J}_k(\lambda)^m)$ is bounded by (B3). If $|\lambda| = 1$, then $k = 1$ by (ii), so $\mathbf{J}_k(\lambda)^m = (\lambda^m)$ is a 1×1 matrix of modulus $|\lambda|^m = 1$. Either way the block sequence is bounded, so $(\mathbf{J}^m)$ is bounded and so is $(\mathbf{A}^m)$.

($\Rightarrow$) Assume $(\mathbf{A}^m)$ is bounded, hence so is $(\mathbf{J}^m)$, hence so is every block sequence.

Proof of (i). Let λ be an eigenvalue and $\mathbf{J}_k(\lambda)$ one of its blocks. The $(1, 1)$ -entry of $\mathbf{J}_k(\lambda)^m$ is λ^m , so $(|\lambda|^m)$ is bounded; by (B2) this forces $|\lambda| \leq 1$.

Proof of (ii). Suppose some eigenvalue λ with $|\lambda| = 1$ had a block $\mathbf{J}_k(\lambda)$ with $k \geq 2$. Its $(1, 2)$ -entry at the m -th power is $\binom{m}{1}\lambda^{m-1} = m\lambda^{m-1}$, of modulus $m|\lambda|^{m-1} = m$, which is unbounded — contradicting the boundedness of that block sequence. So every such block has size 1. This proves the theorem. ■

10.10.4 Powers that converge

Boundedness is not convergence: $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^m$ alternates between two matrices forever. Among the eigenvalues of modulus 1, only $\lambda = 1$ has convergent powers, because λ^m has to sit still.

Theorem 10.10.4. Convergent Matrices

Let $\mathbf{A} \in M_n(\mathbb{C})$, $n \geq 1$, with Jordan form $\mathbf{J}$. Then $(\mathbf{A}^m)$ converges if and only if every eigenvalue λ of $\mathbf{A}$ satisfies

- (i) $|\lambda| < 1$, or
- (ii) $\lambda = 1$ and all of its blocks in $\mathbf{J}$ have size 1.

In that case

$$\lim_{m \rightarrow \infty} \mathbf{A}^m = \mathbf{\Pi},$$

where $\mathbf{\Pi} \in M_n(\mathbb{C})$ is the matrix, in the standard basis, of the projection onto $E_1(\mathbf{A})$ along

$\bigoplus_{\lambda \neq 1} G_\lambda(\mathbf{A})$ if $1 \in \text{spec}(\mathbf{A})$, and $\mathbf{\Pi} = \mathbf{O}$ otherwise.

Idea. For the shape of the answer, take the Jordan form and let $m \rightarrow \infty$ in each block: the blocks with $|\lambda| < 1$ go to $\mathbf{O}$, the blocks $\mathbf{J}_1(1) = (1)$ stay at 1. So $\mathbf{J}^m$ tends to the 0/1 diagonal matrix that marks the 1-blocks, which is the matrix of a projection; conjugating back identifies the two subspaces. For the converse, the $(1,1)$ -entry of a block is λ^m , and if that converges to z then so does $\lambda^{m+1} = \lambda \cdot \lambda^m$, giving $z = \lambda z$ and forcing $\lambda = 1$ or $z = 0$.

Proof. Write $\mathbf{A} = \mathbf{PJP}^{-1}$ with $\mathbf{J} = \mathbf{J}_{k_1}(\lambda_1) \oplus \cdots \oplus \mathbf{J}_{k_d}(\lambda_d)$ (Theorem 10.3.1), and let $\mathcal{B} = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ be the corresponding Jordan basis, the columns of $\mathbf{P}$. For each i let $C_i \subseteq \mathbb{C}^n$ be the span of the k_i basis vectors belonging to block i , so that $\mathbb{C}^n = C_1 \oplus \cdots \oplus C_d$.

($\Leftarrow$) Suppose every eigenvalue satisfies (i) or (ii). Fix a block. If $|\lambda_i| < 1$, then $\mathbf{J}_{k_i}(\lambda_i)^m \rightarrow \mathbf{O}$ by Theorem 10.10.2 applied to the block $\mathbf{J}_{k_i}(\lambda_i)$, whose only eigenvalue is λ_i . If $\lambda_i = 1$, then $k_i = 1$ by (ii) and $\mathbf{J}_1(1)^m = (1)$ for every m . By Theorem 8.1.4 (b), $\mathbf{J}^m$ is the direct sum of these, so

$$\mathbf{J}^m \longrightarrow \mathbf{E},$$

where $\mathbf{E}$ is the diagonal matrix carrying 1 in the positions of the blocks with $\lambda_i = 1$ and 0 everywhere else. By Lemma 10.10.1 (a), $\mathbf{A}^m \rightarrow \mathbf{PEP}^{-1}$.

It remains to identify $\mathbf{PEP}^{-1}$. The matrix $\mathbf{E}$ is the matrix in the basis $\mathcal{B}$ of the projection onto

$$U := \bigoplus_{\lambda_i=1} C_i \quad \text{along} \quad W := \bigoplus_{\lambda_i \neq 1} C_i,$$

because $\mathbf{E}$ fixes each basis vector of the first family and kills each of the second (Theorem 4.5.1). Since the columns of $\mathbf{P}$ are the vectors of $\mathcal{B}$, $\mathbf{P}$ is the change-of-coordinates matrix from $\mathcal{B}$ to the standard basis, so $\mathbf{PEP}^{-1}$ is the matrix of that projection in the standard basis (Theorem 4.2.3). Now, for any $\mu \in \text{spec}(\mathbf{A})$,

$$\bigoplus_{\lambda_i=\mu} C_i = G_\mu(\mathbf{A}).$$

Indeed each C_i with $\lambda_i = \mu$ is killed by $(\mathbf{A} - \mu\mathbf{I})^{k_i}$, hence lies in $G_\mu(\mathbf{A})$ (Corollary 10.1.2 (a)); the sum on the left is direct and has dimension $\sum_{\lambda_i=\mu} k_i = a_{\mathbf{A}}(\mu)$ by Corollary 10.3.2 (a); and $\dim G_\mu(\mathbf{A}) = a_{\mathbf{A}}(\mu)$ by Theorem 10.1.3 (c). An inclusion of subspaces with equal dimensions is an equality (Theorem 1.6.8, applied inside $G_\mu(\mathbf{A})$). Hence $W = \bigoplus_{\lambda \neq 1} G_\lambda(\mathbf{A})$, and $U = G_1(\mathbf{A})$ if $1 \in \text{spec}(\mathbf{A})$, with $U = \{\mathbf{0}\}$ otherwise. Finally, when $1 \in \text{spec}(\mathbf{A})$, condition (ii) says all 1-blocks have size 1, so $g_{\mathbf{A}}(1) = a_{\mathbf{A}}(1)$ by Corollary 10.3.2 (a), (b); since $E_1(\mathbf{A}) \subseteq G_1(\mathbf{A})$ with $\dim E_1(\mathbf{A}) = g_{\mathbf{A}}(1) = a_{\mathbf{A}}(1) = \dim G_1(\mathbf{A})$, we get $G_1(\mathbf{A}) = E_1(\mathbf{A})$. So $\lim_m \mathbf{A}^m = \mathbf{\Pi}$ as stated.

($\Rightarrow$) Suppose $(\mathbf{A}^m)$ converges. By Lemma 10.10.1 (a) so does $(\mathbf{J}^m)$, hence so does each block sequence. Fix a block $\mathbf{J}_k(\lambda)$.

Case $|\lambda| > 1$. Impossible: a convergent sequence is bounded (B3), so this is excluded by Theorem 10.10.3.

Case $|\lambda| = 1$ and $\lambda \neq 1$. The $(1,1)$ -entries give a convergent sequence $\lambda^m \rightarrow z$. The sequence $(\lambda^{m+1})_{m \geq 0}$ is a tail of the same sequence, so it also converges to z ; on the other hand $\lambda^{m+1} = \lambda \cdot \lambda^m \rightarrow \lambda z$ by (B3). Hence $z = \lambda z$, that is, $(1 - \lambda)z = 0$, and $\lambda \neq 1$ forces

$z = 0$. But $|\lambda^m| = 1$ for every m , so $|z| = 1$ by (B3), a contradiction. So this case does not occur either.

Case $\lambda = 1$. If $k \geq 2$, the $(1, 2)$ -entry of $\mathbf{J}_k(1)^m$ is $\binom{m}{1} 1^{m-1} = m$, which does not converge. So $k = 1$.

The three cases leave exactly (i) and (ii). This proves the theorem. ■

The limit deserves a sentence. It is a projection, so $\mathbf{\Pi}^2 = \mathbf{\Pi}$ — which had to be true, since $\mathbf{\Pi}^2 = \lim_m \mathbf{A}^m \lim_m \mathbf{A}^m = \lim_m \mathbf{A}^{2m} = \mathbf{\Pi}$ by (B3) — and it acts as the identity on the eigenvectors for 1 and annihilates everything belonging to the other eigenvalues. In the long run, $\mathbf{A}$ forgets all of $\mathbb{C}^n$ except the fixed space $E_1(\mathbf{A})$.

Example. Deciding and Computing Limits of Powers

For each matrix, decide whether $\mathbf{A}^m$ tends to $\mathbf{O}$, stays bounded, converges, and compute $\lim \mathbf{A}^m$ when it exists.

- (a) $\mathbf{A}_1 = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}$.
- (b) $\mathbf{A}_2 = \begin{pmatrix} 0 & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$.
- (c) $\mathbf{A}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 1 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$.

Solution.

- (a) $p_{\mathbf{A}_1} = x(x-2)+1 = (x-1)^2$, so the only eigenvalue is 1. Put $\mathbf{N} = \mathbf{A}_1 - \mathbf{I} = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$;

then $\mathbf{N} \neq \mathbf{O}$ and $\mathbf{N}^2 = \begin{pmatrix} 1-1 & 1-1 \\ -1+1 & -1+1 \end{pmatrix} = \mathbf{O}$, so $g_{\mathbf{A}_1}(1) = 2 - \text{rank } \mathbf{N} = 1$ and the Jordan form is $\mathbf{J}_2(1)$. Condition (ii) of [Theorem 10.10.3](#) fails, so $(\mathbf{A}_1^m)$ is **unbounded**, and in particular does not converge. Explicitly, $\mathbf{I}$ and $\mathbf{N}$ commute and $\mathbf{N}^2 = \mathbf{O}$, so the binomial theorem gives

$$\mathbf{A}_1^m = (\mathbf{I} + \mathbf{N})^m = \mathbf{I} + m\mathbf{N} = \begin{pmatrix} 1+m & m \\ -m & 1-m \end{pmatrix},$$

which grows linearly.

- (b) $p_{\mathbf{A}_2} = x^2 + \frac{1}{4}$, with roots $\pm \frac{i}{2}$, both of modulus $\frac{1}{2} < 1$. By [Theorem 10.10.2](#), $\mathbf{A}_2^m \rightarrow \mathbf{O}$ (so also bounded and convergent). Explicitly $\mathbf{A}_2^2 = -\frac{1}{4}\mathbf{I}$, so $\mathbf{A}_2^{2l} = (-\frac{1}{4})^l \mathbf{I}$ and $\mathbf{A}_2^{2l+1} = (-\frac{1}{4})^l \mathbf{A}_2$, and every entry has modulus at most $4^{-l} \rightarrow 0$.
- (c) $\mathbf{A}_3 = (1) \oplus \mathbf{J}_2(\frac{1}{3})$ is already a Jordan matrix. Its eigenvalues are 1, with a single block of size 1, and $\frac{1}{3}$, with $|\frac{1}{3}| < 1$. So condition (i) or (ii) holds for each eigenvalue and $\mathbf{A}_3^m$ converges by [Theorem 10.10.4](#). With $\mathbf{P} = \mathbf{I}$, the matrix $\mathbf{E}$ of the proof is $\text{diag}(1, 0, 0)$,

so

$$\lim_m \mathbf{A}_3^m = \text{diag}(1, 0, 0),$$

the projection onto $E_1(\mathbf{A}_3) = \text{span}(\mathbf{e}_1)$ along $G_{1/3}(\mathbf{A}_3) = \text{span}(\mathbf{e}_2, \mathbf{e}_3)$. Explicitly, by Lemma 10.9.1,

$$\mathbf{A}_3^m = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3^{-m} & m 3^{-(m-1)} \\ 0 & 0 & 3^{-m} \end{pmatrix},$$

and $m 3^{-(m-1)} = 3m 3^{-m} \rightarrow 0$ by (B1).

10.10.5 Markov chains, finished

Chapter 9 proved that a Markov chain settles down under three hypotheses: $\mathbf{A}$ diagonalizable over $\mathbb{C}$, $a_{\mathbf{A}}(1) = 1$, and $|\lambda| < 1$ for every other eigenvalue (Theorem 9.11.4). The first hypothesis can now be dropped, and more: for a stochastic matrix, the block condition of Theorem 10.10.3 holds automatically, because the powers of a stochastic matrix have entries in $[0, 1]$.

Corollary 10.10.5. Convergence of a Markov Chain

Let $\mathbf{A} \in M_n(\mathbb{R})$ be stochastic (Definition 9.11.1), $n \geq 1$. Then:

- (a) $(\mathbf{A}^m)$ is bounded; consequently every eigenvalue $\lambda \in \mathbb{C}$ of $\mathbf{A}$ satisfies $|\lambda| \leq 1$, and every eigenvalue with $|\lambda| = 1$ has only 1×1 blocks. In particular $a_{\mathbf{A}}(1) = g_{\mathbf{A}}(1)$.
- (b) $(\mathbf{A}^m)$ converges if and only if 1 is the **only** eigenvalue of $\mathbf{A}$ of modulus 1; and then $\lim_m \mathbf{A}^m$ is the projection onto $E_1(\mathbf{A})$ along $\bigoplus_{\lambda \neq 1} G_{\lambda}(\mathbf{A})$.

Proof.

- (a) By Proposition 9.11.3 (b), every $\mathbf{A}^m$ is stochastic, so its entries are nonnegative and each of its columns sums to 1; hence every entry lies in $[0, 1]$ and $(\mathbf{A}^m)$ is bounded with $c = 1$. Now apply the $(\Rightarrow)$ direction of Theorem 10.10.3. Since $1 \in \text{spec}(\mathbf{A})$ by Proposition 9.11.3 (c), the 1-blocks exist and all have size 1, so $a_{\mathbf{A}}(1) = g_{\mathbf{A}}(1)$ by Corollary 10.3.2 (a), (b).
- (b) By (a), every eigenvalue has $|\lambda| \leq 1$ and every eigenvalue of modulus 1 has only 1×1 blocks. So the criterion of Theorem 10.10.4 — each eigenvalue has $|\lambda| < 1$, or equals 1 with only 1×1 blocks — holds precisely when no eigenvalue $\lambda \neq 1$ has $|\lambda| = 1$. The limit is the one given by Theorem 10.10.4. ■

So the only obstruction to a Markov chain settling down is an eigenvalue of modulus 1 other than 1 itself — periodicity, in the language of probability. Chapter 9's warning example, the swap $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, has the eigenvalue -1 and is exactly this obstruction.

Example. A Two-State Chain

Let $\mathbf{A} = \begin{pmatrix} 3/4 & 1/2 \\ 1/4 & 1/2 \end{pmatrix}$. Decide whether $\mathbf{A}^m$ converges and find the limit.

Solution. $\mathbf{A}$ is stochastic: the entries are nonnegative, and each column sums to 1. Its characteristic polynomial is

$$p_{\mathbf{A}} = (x - \frac{3}{4})(x - \frac{1}{2}) - \frac{1}{8} = x^2 - \frac{5}{4}x + \frac{1}{4} = (x - 1)(x - \frac{1}{4}),$$

so $\text{spec}(\mathbf{A}) = \{1, \frac{1}{4}\}$ and the only eigenvalue of modulus 1 is 1. By [Corollary 10.10.5](#) (b), $\mathbf{A}^m$ converges.

For the limit, $E_1(\mathbf{A}) = \ker(\mathbf{A} - \mathbf{I})$, and $\mathbf{A} - \mathbf{I} = \begin{pmatrix} -1/4 & 1/2 \\ 1/4 & -1/2 \end{pmatrix}$ gives $\text{span}((2, 1))$; the steady state is $\mathbf{v} = (\frac{2}{3}, \frac{1}{3})$, the multiple of $(2, 1)$ whose entries sum to 1. Also $E_{1/4}(\mathbf{A}) = \ker(\mathbf{A} - \frac{1}{4}\mathbf{I})$, and $\mathbf{A} - \frac{1}{4}\mathbf{I} = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 1/4 \end{pmatrix}$ gives $\text{span}((1, -1))$. Since the two eigenvalues are distinct, $G_{1/4}(\mathbf{A}) = E_{1/4}(\mathbf{A})$ and the limit $\mathbf{\Pi}$ is the projection onto $\text{span}((2, 1))$ along $\text{span}((1, -1))$. Computing it in the basis $\mathcal{C} = ((2, 1), (1, -1))$: with $\mathbf{Q} = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}$, $\det \mathbf{Q} = -3$, and $\mathbf{Q}^{-1} = \frac{1}{3} \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$, so

$$\mathbf{\Pi} = \mathbf{Q} \text{diag}(1, 0) \mathbf{Q}^{-1} = \begin{pmatrix} 2 & 0 \\ 1 & 0 \end{pmatrix} \cdot \frac{1}{3} \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{pmatrix}.$$

Check. $\mathbf{\Pi} = \mathbf{v} \mathbf{1}^\top$, as Theorem 9.11.4 predicts; $\mathbf{\Pi}^2 = \mathbf{\Pi}$ since $\mathbf{1}^\top \mathbf{v} = 1$; the columns of $\mathbf{\Pi}$ are the steady state, so every initial probability vector $\mathbf{x}_0$ gives $\mathbf{x}_m = \mathbf{A}^m \mathbf{x}_0 \rightarrow \mathbf{\Pi} \mathbf{x}_0 = \mathbf{v}(\mathbf{1}^\top \mathbf{x}_0) = \mathbf{v}$. And $\mathbf{A}^2 = \begin{pmatrix} 11/16 & 5/8 \\ 5/16 & 3/8 \end{pmatrix}$ is already close to $\mathbf{\Pi}$, the error decaying like 4^{-m} .

10.10.6 The geometric series of a matrix

The scalar identity $\sum_k a^k = (1 - a)^{-1}$ for $|a| < 1$ has a matrix version, with the same proof and the same hypothesis read off the eigenvalues.

Theorem 10.10.6. Neumann Series

Let $\mathbf{A} \in M_n(\mathbb{C})$, $n \geq 1$. Then the series $\sum_{k \geq 0} \mathbf{A}^k$ converges if and only if every eigenvalue λ of $\mathbf{A}$ satisfies $|\lambda| < 1$. In that case $\mathbf{I} - \mathbf{A}$ is invertible and

$$\sum_{k=0}^{\infty} \mathbf{A}^k = (\mathbf{I} - \mathbf{A})^{-1}.$$

Idea. Telescoping gives $(\mathbf{I} - \mathbf{A})(\mathbf{I} + \mathbf{A} + \cdots + \mathbf{A}^K) = \mathbf{I} - \mathbf{A}^{K+1}$, an exact identity for every K . If $\mathbf{I} - \mathbf{A}$ is invertible, solve it for the partial sum and let $K \rightarrow \infty$, where [Theorem 10.10.2](#) kills $\mathbf{A}^{K+1}$. Conversely, the terms of a convergent series tend to $\mathbf{O}$, which is the hypothesis

of that theorem again.

Proof. ($\Leftarrow$) Suppose $|\lambda| < 1$ for every $\lambda \in \text{spec}(\mathbf{A})$. In particular $1 \notin \text{spec}(\mathbf{A})$, so $0 \notin \text{spec}(\mathbf{I} - \mathbf{A})$ — a vector with $(\mathbf{I} - \mathbf{A})\mathbf{v} = \mathbf{0}$ is a vector with $\mathbf{A}\mathbf{v} = \mathbf{v}$ — and $\mathbf{I} - \mathbf{A}$ is invertible (Theorem 9.2.6). Put $\mathbf{S}_K = \sum_{k=0}^K \mathbf{A}^k$. Multiplying out and canceling,

$$(\mathbf{I} - \mathbf{A})\mathbf{S}_K = \sum_{k=0}^K \mathbf{A}^k - \sum_{k=1}^{K+1} \mathbf{A}^k = \mathbf{I} - \mathbf{A}^{K+1},$$

hence $\mathbf{S}_K = (\mathbf{I} - \mathbf{A})^{-1}(\mathbf{I} - \mathbf{A}^{K+1})$. By Theorem 10.10.2, $\mathbf{A}^{K+1} \rightarrow \mathbf{0}$. Each entry of $\mathbf{S}_K$ is a fixed linear combination of the entries of $\mathbf{I} - \mathbf{A}^{K+1}$, with coefficients from $(\mathbf{I} - \mathbf{A})^{-1}$, so by (B3) it converges to the corresponding combination for $\mathbf{I}$, that is, $\mathbf{S}_K \rightarrow (\mathbf{I} - \mathbf{A})^{-1}\mathbf{I} = (\mathbf{I} - \mathbf{A})^{-1}$.

($\Rightarrow$) Suppose $\sum_k \mathbf{A}^k$ converges. Fix a position (r, s) ; the series of complex numbers $\sum_k (\mathbf{A}^k)_{rs}$ converges, so $(\mathbf{A}^k)_{rs} \rightarrow 0$ by (B4). This holds at every position, so $\mathbf{A}^k \rightarrow \mathbf{0}$, and Theorem 10.10.2 gives $|\lambda| < 1$ for every eigenvalue. This proves the theorem. ■

Example. Inverting by Summing a Series

Let $\mathbf{A} = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$. Verify that $\sum_k \mathbf{A}^k$ converges, sum it, and compare with $(\mathbf{I} - \mathbf{A})^{-1}$.

Solution. $p_{\mathbf{A}} = x^2 - \frac{1}{4}$, so $\text{spec}(\mathbf{A}) = \{\frac{1}{2}, -\frac{1}{2}\}$ and both eigenvalues have modulus $\frac{1}{2} < 1$. By Theorem 10.10.6 the series converges.

Summing it. $\mathbf{A}^2 = \frac{1}{4}\mathbf{I}$, so $\mathbf{A}^{2l} = 4^{-l}\mathbf{I}$ and $\mathbf{A}^{2l+1} = 4^{-l}\mathbf{A}$. Grouping the partial sums by parity, which is legitimate because each entry sequence splits into two convergent pieces (B3),

$$\sum_{k \geq 0} \mathbf{A}^k = \left(\sum_{l \geq 0} 4^{-l} \right) (\mathbf{I} + \mathbf{A}) = \frac{1}{1 - \frac{1}{4}} (\mathbf{I} + \mathbf{A}) = \frac{4}{3} \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} = \begin{pmatrix} 4/3 & 2/3 \\ 2/3 & 4/3 \end{pmatrix},$$

the scalar sum being the case $n = 1$ of Theorem 10.10.6 applied to the 1×1 matrix $(\frac{1}{4})$.

The inverse. $\mathbf{I} - \mathbf{A} = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$ has determinant $1 - \frac{1}{4} = \frac{3}{4}$, so by Theorem 0.9.8

$$(\mathbf{I} - \mathbf{A})^{-1} = \frac{4}{3} \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix},$$

in agreement.

Remark. The first thing every criterion in this section looks at is the largest modulus of an eigenvalue,

$$\rho(\mathbf{A}) = \max\{|\lambda| : \lambda \in \text{spec}(\mathbf{A})\},$$

called the **spectral radius** of $\mathbf{A}$. Two of the criteria need nothing else: $\mathbf{A}^m \rightarrow \mathbf{O}$ if and only if $\rho(\mathbf{A}) < 1$, and $\sum_k \mathbf{A}^k$ converges if and only if $\rho(\mathbf{A}) < 1$. The other two need more when $\rho(\mathbf{A}) = 1$: boundedness also asks for the block sizes at modulus 1, and convergence asks in addition that 1 be the only eigenvalue of modulus 1. Chapter 16 measures matrices by norms, proves that the spectral radius never exceeds the norm of $\mathbf{A}$ induced by any vector norm, and recovers $\rho(\mathbf{A})$ as a limit built from the norms of the powers $\mathbf{A}^m$; that is what turns these exact criteria into usable estimates.

10.10.7 Exercises

10.10.7.1 A. Check your understanding

Exercise. A1

- Define entrywise convergence and entrywise boundedness for a sequence of matrices.
- Write down the entries of $\mathbf{J}_k(\lambda)^m$, and say which one shows that a block of size ≥ 2 with $|\lambda| = 1$ has unbounded powers.
- State the three criteria: $\mathbf{A}^m \rightarrow \mathbf{O}$, $(\mathbf{A}^m)$ bounded, $(\mathbf{A}^m)$ convergent.
- True or false: if every entry of $\mathbf{A}$ has modulus less than 1, then $\mathbf{A}^m \rightarrow \mathbf{O}$. Justify your answer.
- True or false: if $(\mathbf{A}^m)$ is bounded then it converges. Justify your answer.
- For which $\mathbf{A}$ does $\sum_k \mathbf{A}^k$ converge, and what is the sum?

Solution.

- $\mathbf{M}_m \rightarrow \mathbf{M}$ if $(\mathbf{M}_m)_{rs} \rightarrow \mathbf{M}_{rs}$ for every position (r, s) ; $(\mathbf{M}_m)$ is bounded if there is one constant c with $|(\mathbf{M}_m)_{rs}| \leq c$ for all m and all (r, s) ([Definition 10.10.1](#)).
- $(\mathbf{J}_k(\lambda)^m)_{r, r+j} = \binom{m}{j} \lambda^{m-j}$ for $0 \leq j \leq k-1$, with zeros below the diagonal (Lemma 10.9.1 with $p = x^m$). The entry with $j = 1$ is $m\lambda^{m-1}$, of modulus m when $|\lambda| = 1$.
- $\mathbf{A}^m \rightarrow \mathbf{O}$ if and only if every eigenvalue has $|\lambda| < 1$ ([Theorem 10.10.2](#)); $(\mathbf{A}^m)$ is bounded if and only if every eigenvalue has $|\lambda| \leq 1$ and those with $|\lambda| = 1$ have only 1×1 blocks ([Theorem 10.10.3](#)); $(\mathbf{A}^m)$ converges if and only if every eigenvalue has $|\lambda| < 1$ or equals 1 with only 1×1 blocks ([Theorem 10.10.4](#)).
- False. $\mathbf{A} = \begin{pmatrix} 0.9 & 0.9 \\ 0.9 & 0.9 \end{pmatrix}$ has the eigenvalue 1.8, and $\mathbf{A}^m \mathbf{1} = 1.8^m \mathbf{1}$ is unbounded; see the warning after [Theorem 10.10.2](#).
- False. $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ satisfies $\mathbf{A}^2 = \mathbf{I}$, so $\mathbf{A}^m$ alternates between $\mathbf{I}$ and $\mathbf{A}$ and is bounded but not convergent. Its eigenvalues are ± 1 , and the eigenvalue -1 is what [Theorem 10.10.4](#) excludes.
- Exactly when every eigenvalue has $|\lambda| < 1$, and then the sum is $(\mathbf{I} - \mathbf{A})^{-1}$ ([Theorem 10.10.6](#)).

10.10.7.2 B. Practice

Exercise. B1: Decide the behavior

For each matrix over $\mathbb{C}$, decide whether $\mathbf{A}^m \rightarrow \mathbf{O}$, whether $(\mathbf{A}^m)$ is bounded, and whether $(\mathbf{A}^m)$ converges. Justify your answer.

- (a) $\mathbf{A}_1 = \begin{pmatrix} 1/2 & 1 \\ 0 & 1/2 \end{pmatrix}$.
- (b) $\mathbf{A}_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.
- (c) $\mathbf{A}_3 = \mathbf{J}_2(1) \oplus \mathbf{J}_1(\frac{1}{2}) \in M_3(\mathbb{C})$.
- (d) $\mathbf{A}_4 = \begin{pmatrix} 0 & 0 & 5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Solution.

- (a) The only eigenvalue is $\frac{1}{2}$, of modulus < 1 , so $\mathbf{A}_1^m \rightarrow \mathbf{O}$ by [Theorem 10.10.2](#); hence it is bounded and convergent. (The superdiagonal 1 is irrelevant: by Lemma 10.9.1 the $(1, 2)$ -entry of $\mathbf{A}_1^m$ is $m 2^{-(m-1)} \rightarrow 0$ by (B1).)
- (b) $p_{\mathbf{A}_2} = x^2 + 1$, so $\text{spec}(\mathbf{A}_2) = \{i, -i\}$, both of modulus 1, and the eigenvalues are distinct, so every block has size 1. By [Theorem 10.10.3](#), $(\mathbf{A}_2^m)$ is **bounded**; it does not tend to $\mathbf{O}$ ([Theorem 10.10.2](#)) and does not converge, since the eigenvalue i has modulus 1 and is not 1 ([Theorem 10.10.4](#)). Concretely $\mathbf{A}_2^4 = \mathbf{I}$, so the powers cycle through four matrices.
- (c) The eigenvalue 1 has a block of size 2, so [Theorem 10.10.3](#) (ii) fails: $(\mathbf{A}_3^m)$ is **unbounded**, hence neither convergent nor tending to $\mathbf{O}$. The $(1, 2)$ -entry of $\mathbf{A}_3^m$ is m .
- (d) $\mathbf{A}_4^2 = \mathbf{O}$, so $\mathbf{A}_4^m = \mathbf{O}$ for $m \geq 2$ and $\mathbf{A}_4^m \rightarrow \mathbf{O}$ trivially: all eigenvalues are 0. The large entry 5 does not matter, illustrating the warning after [Theorem 10.10.2](#).

Exercise. B2: Computing limits

- (a) Let $\mathbf{A} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$. Show that $\mathbf{A}^m$ converges and find the limit, both by [Theorem 10.10.4](#) and by a direct computation of $\mathbf{A}^m$.
- (b) Let $\mathbf{B} = \begin{pmatrix} 1 & 1/3 \\ 0 & 1/3 \end{pmatrix}$. Find $\lim_m \mathbf{B}^m$ and identify it as a projection.

Solution.

- (a) $p_{\mathbf{A}} = x^2 - x = x(x - 1)$, so $\text{spec}(\mathbf{A}) = \{0, 1\}$. The eigenvalue 0 has $|0| < 1$ and the eigenvalue 1 has $a_{\mathbf{A}}(1) = 1$, so its single block has size 1; by [Theorem 10.10.4](#) the powers converge, and the limit is the projection onto $E_1(\mathbf{A})$ along $G_0(\mathbf{A}) = E_0(\mathbf{A})$.
Now $\mathbf{A} - \mathbf{I} = \begin{pmatrix} -1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}$ gives $E_1(\mathbf{A}) = \text{span}((1, 1))$ and $E_0(\mathbf{A}) = \ker \mathbf{A} =$

$\text{span}((1, -1))$. With $\mathbf{Q} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and $\mathbf{Q}^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$,

$$\lim_m \mathbf{A}^m = \mathbf{Q} \text{diag}(1, 0) \mathbf{Q}^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \mathbf{A}.$$

Directly. $\mathbf{A}^2 = \mathbf{A}$, so $\mathbf{A}^m = \mathbf{A}$ for every $m \geq 1$ and the limit is $\mathbf{A}$. ($\mathbf{A}$ is itself a projection, and the theorem returns it unchanged.)

- (b) $\mathbf{B}$ is upper triangular, so $\text{spec}(\mathbf{B}) = \{1, \frac{1}{3}\}$ (Theorem 7.4.6). Both criteria hold: $|\frac{1}{3}| < 1$, and $a_{\mathbf{B}}(1) = 1$ so the 1-block has size 1. Solving $(\mathbf{B} - \mathbf{I})\mathbf{x} = \mathbf{0}$ with $\mathbf{B} - \mathbf{I} = \begin{pmatrix} 0 & 1/3 \\ 0 & -2/3 \end{pmatrix}$ gives $E_1(\mathbf{B}) = \text{span}(\mathbf{e}_1)$; solving $(\mathbf{B} - \frac{1}{3}\mathbf{I})\mathbf{x} = \mathbf{0}$ with $\mathbf{B} - \frac{1}{3}\mathbf{I} = \begin{pmatrix} 2/3 & 1/3 \\ 0 & 0 \end{pmatrix}$ gives $E_{1/3}(\mathbf{B}) = \text{span}((1, -2))$. With $\mathbf{Q} = \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix}$ we have $\det \mathbf{Q} = -2$ and $\mathbf{Q}^{-1} = \frac{1}{-2} \begin{pmatrix} -2 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1/2 \\ 0 & -1/2 \end{pmatrix}$. Multiplying in two steps, $\mathbf{Q} \text{diag}(1, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and then

$$\lim_m \mathbf{B}^m = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \mathbf{Q}^{-1} = \begin{pmatrix} 1 & 1/2 \\ 0 & 0 \end{pmatrix}.$$

This is the projection onto $\text{span}(\mathbf{e}_1)$ along $\text{span}((1, -2))$: it fixes $\mathbf{e}_1$ and sends $(1, -2)$ to $(1 - 1, 0) = \mathbf{0}$. *Check.* $\mathbf{B}^m = \begin{pmatrix} 1 & c_m \\ 0 & 3^{-m} \end{pmatrix}$ with $c_1 = \frac{1}{3}$ and $c_{m+1} = \frac{1}{3}(1 + c_m)$, read off $\mathbf{B}^{m+1} = \mathbf{B}^m \mathbf{B}$; the first values are $\frac{1}{3}, \frac{4}{9}, \frac{13}{27}$, climbing towards the fixed point $c = \frac{1}{3}(1 + c)$, that is $c = \frac{1}{2}$.

Exercise. B3: A Neumann series

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

- Explain why $\sum_k \mathbf{A}^k$ converges.
- Compute $(\mathbf{I} - \mathbf{A})^{-1}$ directly.
- Compute the partial sum $\mathbf{I} + \mathbf{A} + \mathbf{A}^2$ and compare it with (b).

Solution.

- $\mathbf{A}$ is upper triangular with both diagonal entries $\frac{1}{3}$, so $\text{spec}(\mathbf{A}) = \{\frac{1}{3}\}$ (Theorem 7.4.6) and $|\frac{1}{3}| < 1$. By Theorem 10.10.6 the series converges, with sum $(\mathbf{I} - \mathbf{A})^{-1}$.
- $\mathbf{I} - \mathbf{A} = \begin{pmatrix} 2/3 & -1/3 \\ 0 & 2/3 \end{pmatrix}$, of determinant $\frac{4}{9}$. By Theorem 0.9.8,

$$(\mathbf{I} - \mathbf{A})^{-1} = \frac{9}{4} \begin{pmatrix} 2/3 & 1/3 \\ 0 & 2/3 \end{pmatrix} = \begin{pmatrix} 3/2 & 3/4 \\ 0 & 3/2 \end{pmatrix}.$$

- (c) Write $\mathbf{A} = \frac{1}{3}(\mathbf{I} + \mathbf{M})$ with $\mathbf{M} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\mathbf{M}^2 = \mathbf{O}$; then $\mathbf{A}^k = 3^{-k}(\mathbf{I} + k\mathbf{M})$ by the binomial theorem, the two summands commuting. So

$$\mathbf{I} + \mathbf{A} + \mathbf{A}^2 = (1 + \frac{1}{3} + \frac{1}{9})\mathbf{I} + (\frac{1}{3} + \frac{2}{9})\mathbf{M} = \frac{13}{9}\mathbf{I} + \frac{5}{9}\mathbf{M} = \begin{pmatrix} 13/9 & 5/9 \\ 0 & 13/9 \end{pmatrix}.$$

Summing the whole series the same way, $(\mathbf{I} - \mathbf{A})^{-1} = (\sum_{k \geq 0} 3^{-k})\mathbf{I} + (\sum_{k \geq 0} k3^{-k})\mathbf{M}$, so comparing with (b) reads off $\sum_k 3^{-k} = \frac{3}{2}$ and $\sum_k k3^{-k} = \frac{3}{4}$. The two partial sums $\frac{13}{9}$ and $\frac{5}{9}$, about 1.44 and 0.56, are on their way to $\frac{3}{2}$ and $\frac{3}{4}$; the convergence is geometric, as the factor 3^{-k} shows.

10.10.7.3 C. Going deeper

Exercise. C1: Bounded but not convergent

- (a) Give an example of $\mathbf{A} \in M_2(\mathbb{R})$ with $(\mathbf{A}^m)$ bounded but not convergent, and one with $(\mathbf{A}^m)$ convergent but $\mathbf{A}^m \not\rightarrow \mathbf{O}$.
- (b) Prove that if $\mathbf{A}^m \rightarrow \mathbb{L}$, then $\mathbb{L}^2 = \mathbb{L}$ and $\mathbf{A}\mathbb{L} = \mathbb{L}\mathbf{A} = \mathbb{L}$.
- (c) Deduce that if $\mathbf{A}^m \rightarrow \mathbb{L}$ with $\mathbb{L}$ invertible, then $\mathbb{L} = \mathbf{I}$ and $\mathbf{A} = \mathbf{I}$.
- (d) Give an invertible $\mathbf{A}$ with $\mathbf{A}^m \rightarrow \mathbb{L}$ and $\mathbb{L} \notin \{\mathbf{O}, \mathbf{I}\}$, so that the invertibility of $\mathbb{L}$, not of $\mathbf{A}$, is what part (c) needs.

Solution.

- (a) For the first, $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$: $\mathbf{A}^2 = \mathbf{I}$, so $\mathbf{A}^m$ alternates between $\mathbf{I}$ and $\mathbf{A}$ and is bounded but not convergent; its eigenvalues are ± 1 . For the second, $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$: $\mathbf{A}^m = \text{diag}(1, 2^{-m}) \rightarrow \text{diag}(1, 0) \neq \mathbf{O}$.
- (b) Each entry of $\mathbf{A}^m \mathbf{A}^m = \mathbf{A}^{2m}$ is a sum of products of entries of $\mathbf{A}^m$, so by (B3) it converges to the corresponding expression in the entries of $\mathbb{L}$; hence $\mathbf{A}^{2m} \rightarrow \mathbb{L}^2$. But $(\mathbf{A}^{2m})_m$ is a subsequence of $(\mathbf{A}^m)_m$ — more precisely, for each position the sequence $(\mathbf{A}^{2m})_{rs}$ is a subsequence of the convergent sequence $(\mathbf{A}^m)_{rs}$, so it has the same limit $\mathbb{L}_{rs}$. Hence $\mathbb{L}^2 = \mathbb{L}$. Similarly $\mathbf{A}^{m+1} = \mathbf{A}\mathbf{A}^m = \mathbf{A}^m \mathbf{A}$, and the left side is a tail of the convergent sequence, so it tends to $\mathbb{L}$, while the two right sides tend to $\mathbf{A}\mathbb{L}$ and $\mathbb{L}\mathbf{A}$ by (B3). Hence $\mathbf{A}\mathbb{L} = \mathbb{L}\mathbf{A} = \mathbb{L}$.
- (c) Suppose $\mathbf{A}^m \rightarrow \mathbb{L}$ with $\mathbb{L}$ invertible. By (b), $\mathbb{L}^2 = \mathbb{L}$; multiplying by $\mathbb{L}^{-1}$ gives $\mathbb{L} = \mathbf{I}$. Then $\mathbf{A}\mathbb{L} = \mathbb{L}$ reads $\mathbf{A} = \mathbf{I}$. (Conversely $\mathbf{A} = \mathbf{I}$ gives $\mathbf{A}^m = \mathbf{I} \rightarrow \mathbf{I}$, so this case really occurs.)
- (d) $\mathbf{A} = \text{diag}(1, \frac{1}{2})$ is invertible, and $\mathbf{A}^m = \text{diag}(1, 2^{-m}) \rightarrow \text{diag}(1, 0) = \mathbb{L}$, which is neither $\mathbf{O}$ nor $\mathbf{I}$ — and not invertible. So no hypothesis on $\mathbf{A}$ alone forces $\mathbb{L} \in \{\mathbf{O}, \mathbf{I}\}$: by [Theorem 10.10.4](#) the limit is the projection onto $E_1(\mathbf{A})$ along the other generalized eigenspaces, and that projection is $\mathbf{O}$ or $\mathbf{I}$ only in the extreme cases $1 \notin \text{spec}(\mathbf{A})$ and $E_1(\mathbf{A}) = \mathbb{C}^n$.

Exercise. C2: Chapter 9's Markov theorem without diagonalizability

Let $\mathbf{A} \in M_n(\mathbb{R})$ be stochastic with $a_{\mathbf{A}}(1) = 1$ and $|\lambda| < 1$ for every eigenvalue $\lambda \neq 1$.

- (a) Prove that $\mathbf{1}^\top \mathbf{w} = 0$ for every $\mathbf{w} \in G_\lambda(\mathbf{A})$ with $\lambda \neq 1$. *Hint:* $\mathbf{1}^\top (\mathbf{A} - \lambda \mathbf{I}) = (1 - \lambda) \mathbf{1}^\top$.
- (b) Prove that $(\mathbf{A}^m)$ converges to a **stochastic** matrix $\mathbf{\Pi}$, and deduce that $\mathbf{A}$ has a steady state $\mathbf{v}$ with $E_1(\mathbf{A}) = \text{span}(\mathbf{v})$.
- (c) Deduce that $\mathbf{\Pi} = \mathbf{v} \mathbf{1}^\top$ and that $\mathbf{A}^m \mathbf{x}_0 \rightarrow \mathbf{v}$ for every probability vector $\mathbf{x}_0$, **without** assuming that $\mathbf{A}$ is diagonalizable.

Solution.

- (a) Since $\mathbf{A}$ is stochastic, $\mathbf{1}^\top \mathbf{A} = \mathbf{1}^\top$ (Proposition 9.11.3 (c)), so $\mathbf{1}^\top (\mathbf{A} - \lambda \mathbf{I}) = \mathbf{1}^\top - \lambda \mathbf{1}^\top = (1 - \lambda) \mathbf{1}^\top$. Iterating n times, $\mathbf{1}^\top (\mathbf{A} - \lambda \mathbf{I})^n = (1 - \lambda)^n \mathbf{1}^\top$. Let $\mathbf{w} \in G_\lambda(\mathbf{A})$, so $(\mathbf{A} - \lambda \mathbf{I})^n \mathbf{w} = \mathbf{0}$ (Corollary 10.1.2 (a)). Then

$$0 = \mathbf{1}^\top (\mathbf{A} - \lambda \mathbf{I})^n \mathbf{w} = (1 - \lambda)^n \mathbf{1}^\top \mathbf{w},$$

and $\lambda \neq 1$ gives $(1 - \lambda)^n \neq 0$, so $\mathbf{1}^\top \mathbf{w} = 0$.

- (b) By hypothesis 1 is the only eigenvalue of modulus 1, so $(\mathbf{A}^m)$ converges by [Corollary 10.10.5](#) (b); call the limit $\mathbf{\Pi}$. Every $\mathbf{A}^m$ is stochastic (Proposition 9.11.3 (b)), so all its entries are ≥ 0 and each of its columns sums to 1. Both properties pass to the limit: a limit of real numbers ≥ 0 is ≥ 0 , and a column sum is a finite sum of entries, so it converges to the sum of the limits (Chapter 9 §11). Hence $\mathbf{\Pi}$ is stochastic, and in particular $\mathbf{\Pi} \neq \mathbf{0}$.

By [Theorem 10.10.4](#), $\mathbf{\Pi}$ is the projection onto $E_1(\mathbf{A})$ along $\bigoplus_{\lambda \neq 1} G_\lambda(\mathbf{A})$, so $\text{im } T_{\mathbf{\Pi}} = E_1(\mathbf{A})$. Since $a_{\mathbf{A}}(1) = 1$ and $1 \leq g_{\mathbf{A}}(1) \leq a_{\mathbf{A}}(1)$ (Theorem 9.3.1), $E_1(\mathbf{A})$ is a line. Each column of $\mathbf{\Pi}$ lies in $\text{im } T_{\mathbf{\Pi}} = E_1(\mathbf{A})$ (Definition 0.9.9: column j is $\mathbf{\Pi} \mathbf{e}_j$) and is a probability vector, hence is fixed by $\mathbf{A}$ and is a steady state. Any two probability vectors on the same line are equal: if $\mathbf{w} = c\mathbf{v}$ with both entry sums equal to 1, then $1 = c$. So all the columns of $\mathbf{\Pi}$ are one and the same steady state $\mathbf{v}$, and $E_1(\mathbf{A}) = \text{span}(\mathbf{v})$.

- (c) All columns of $\mathbf{\Pi}$ equal $\mathbf{v}$, which is exactly the statement $\mathbf{\Pi} = \mathbf{v} \mathbf{1}^\top$. (Independently of (b), part (a) confirms it: $\mathbf{v} \mathbf{1}^\top$ fixes $\mathbf{v}$ because $\mathbf{1}^\top \mathbf{v} = 1$, and kills every $\mathbf{w} \in G_\lambda(\mathbf{A})$ with $\lambda \neq 1$ because $\mathbf{1}^\top \mathbf{w} = 0$; by Theorem 10.1.3 (a) those two conditions determine the projection onto $E_1(\mathbf{A}) = G_1(\mathbf{A})$ along the rest, using $\dim G_1(\mathbf{A}) = a_{\mathbf{A}}(1) = 1 = \dim E_1(\mathbf{A})$ and Theorem 1.6.8.) For a probability vector $\mathbf{x}_0$, each entry of $\mathbf{A}^m \mathbf{x}_0$ is a fixed linear combination of entries of $\mathbf{A}^m$, so by (B3)

$$\mathbf{A}^m \mathbf{x}_0 \longrightarrow \mathbf{\Pi} \mathbf{x}_0 = \mathbf{v} (\mathbf{1}^\top \mathbf{x}_0) = \mathbf{v} (x_{01} + \cdots + x_{0n}) = \mathbf{v}.$$

No diagonalizability was used: the blocks for the eigenvalues $\lambda \neq 1$ may have any sizes, since [Theorem 10.10.2](#) does not care. This removes the diagonalizability hypothesis from Theorem 9.11.4.