Eigenvalue theory and quadratic forms

Robert J. Carroll

Last revised: 6 May 2026

---

1 Motivation

The previous handout developed the language: vector spaces, linear maps, matrices, inner products. This handout takes up the structural payoffs that language unlocks, and the payoffs are what political-economy modelers actually want from linear algebra. Three of them in particular will recur across the next several clusters.

The first is dynamics. A linear iteration $\boldsymbol{x}_t = A\boldsymbol{x}_{t-1}$ runs voter learning models, opinion dynamics on social networks, the time evolution of approval ratings under noisy aggregation, and the transition rules of Markov chains. The long-run behavior of such an iteration — whether it converges, where it converges to, how fast it gets there — is governed entirely by the eigenvalues of the matrix A. The eigenvalue with the largest modulus determines the dominant dynamics, and the corresponding eigenvector is the direction in which the long run lies.

The second is geometry. Quadratic forms — functions of the shape $\boldsymbol{x}^\top A\boldsymbol{x}$ for symmetric A — are the canonical mathematical object behind quadratic-loss utilities in spatial voting, Mahalanobis distances in cleavage-structure measurement, and (in the next cluster) the second-order conditions for optimization. Whether a quadratic form is non-negative for every input is exactly the question of whether A is positive semidefinite, which is exactly the question of whether A's eigenvalues are non-negative. The optimization handout in the next cluster will build the standard concavity and quasiconcavity tests on top of this section's definiteness machinery.

The third is the structural reason Markov chains converge. The transition matrix of a Markov chain has special structure (rows sum to one, all entries non-negative) that the Perron–Frobenius theorem exploits to guarantee a unique stationary distribution under irreducibility. The theorem is a consequence of eigenvalue theory and the structural properties of non-negative matrices, and it is the reason the polling-and-learning models that political scientists run actually have the long-run distributions they presume.

The handout walks through the four pieces: eigenvalues and eigenvectors (§2), diagonalization (§3), the spectral theorem for symmetric matrices (§4), quadratic forms and definiteness (§5), and Perron–Frobenius for stochastic matrices (§6). Each piece is calibrated to a specific downstream payoff named in the corresponding section's opener.

2 Eigenvalues and eigenvectors

What does the long run of a political process look like? An analyst watching vote shares evolve under repeated campaign cycles, opinions move on social media, party affiliations shift across generations, or an authoritarian regime's elite turn over under repeated purges, is watching a vector $\boldsymbol{x}_t \in \mathbb{R}^n$ change over time, and the most basic substantive question she can ask is whether $\boldsymbol{x}_t$ settles down — and if so, what it settles to. When the dynamics is linear, $\boldsymbol{x}_t = A\boldsymbol{x}_{t-1}$ for some fixed $n \times n$ matrix

A, the answer turns out to be sharp and somewhat surprising. Among all the directions in which $\boldsymbol{x}_t$ might evolve, the long run is governed by exactly *one*, with everything else decaying away in comparison. The directions that matter are the *eigenvectors* of A — the special directions on which A acts simply, by scaling. The rates at which they grow or shrink under iteration are the *eigenvalues*. The eigenvector whose eigenvalue is largest in modulus is the one that dominates the long run; the others fade. This section develops the apparatus for finding eigenvalues and eigenvectors of a given matrix, and it sets up the next two sections, which respectively (i) say when the per-eigenvector decomposition gives a clean coordinate-wise dynamics (§3) and (ii) identify a structurally important class of matrices for which the decomposition is always available and especially well-behaved (§4).

Definition 1. For a square matrix $A \in \mathbb{R}^{n\times n}$, a scalar $\lambda \in \mathbb{C}$ is an *eigenvalue* of A if there exists a non-zero vector $\boldsymbol{v} \in \mathbb{C}^n$ with $A\boldsymbol{v} = \lambda\boldsymbol{v}$. Such a $\boldsymbol{v}$ is an *eigenvector* of A corresponding to λ. The set of all eigenvalues of A is the *spectrum* of A, written $\operatorname{spec}(A)$.

The complex scalars are needed because real matrices can have complex eigenvalues (the standard rotation matrix in $\mathbb{R}^2$ has eigenvalues $e^{\pm i\theta}$). For most political-economy applications the matrices we work with will turn out to have real eigenvalues — symmetric matrices and stochastic matrices both do — but the general definition is over $\mathbb{C}$.

Proposition 2. *λ is an eigenvalue of A iff $\det(A - \lambda\mathbf{I}) = 0$. The function $p_A(\lambda) = \det(A - \lambda\mathbf{I})$ is a degree-n polynomial in λ, called the* characteristic polynomial *of A.*

Proof. λ is an eigenvalue iff there is a non-zero $\boldsymbol{v}$ with $(A - \lambda\mathbf{I})\boldsymbol{v} = \mathbf{0}$, iff $A - \lambda\mathbf{I}$ has a non-trivial kernel, iff $A - \lambda\mathbf{I}$ is non-invertible, iff $\det(A - \lambda\mathbf{I}) = 0$. $\square$

By the fundamental theorem of algebra, p_A has exactly n roots in $\mathbb{C}$ counted with multiplicity. So an $n \times n$ matrix has n eigenvalues in $\mathbb{C}$ counted with multiplicity, though the count drops when multiple eigenvalues coincide.

Example 3 (Voter learning as a linear iteration)**.** A common social-learning model has N voters update their beliefs each period as a weighted average of their neighbors' previous beliefs: $\boldsymbol{x}_t = A\boldsymbol{x}_{t-1}$, where A is the matrix of weights (each row summing to 1, with a_{ij} the weight voter i puts on voter j's belief). The matrix A has $\mathbf{1}$ (the all-ones vector) as an eigenvector with eigenvalue 1, by the row-sum condition. Other eigenvalues of A are smaller in modulus (under standard assumptions), so iterates of A pull every initial belief vector toward the span of $\mathbf{1}$ — which is to say, toward consensus. The eigenvalue 1 corresponds to the consensus direction; the smaller eigenvalues correspond to disagreement modes that decay over time.

3 Diagonalization

What makes a high-dimensional linear dynamics tractable is when the directions identified by the eigenvalues actually *decouple*. When A has enough independent eigenvectors to form a basis of $\mathbb{R}^n$ — the case this section identifies and works through — the dynamics $\boldsymbol{x}_t = A\boldsymbol{x}_{t-1}$ decomposes into independent one-dimensional sub-dynamics, one per eigenvector. Each eigenvector represents a *mode* of the system that evolves on its own: a consensus mode in a voter-learning model, a dominant cleavage in a measurement application, a principal axis in a PCA decomposition. Each mode decays or grows at the rate set by its own eigenvalue, and crucially the modes do not interact — a high-dimensional vector trajectory is just a sum of independent one-dimensional pieces. The long-run behavior is therefore the long-run behavior of each piece in parallel, and typically only the piece with the largest-modulus eigenvalue survives, with the rest fading away. The political-economy reading is consequential: predictions about a multi-actor, multi-issue political dynamic reduce, when diagonalization is available, to predictions about a one-dimensional summary. Most of what makes spectral methods practically useful in applied political science — eigenvector centrality on networks, principal-component scaling, structural-vector autoregression in time series — depends on this kind of decomposition being available. The formal statement that it is available is the diagonalization theorem.

Definition 4. A square matrix $A \in \mathbb{R}^{n\times n}$ is *diagonalizable* (over $\mathbb{C}$) if there exists an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$. The columns of P are eigenvectors of A; the diagonal entries of D are the corresponding eigenvalues.

Proposition 5. *A is diagonalizable iff A has n linearly independent eigenvectors.*

Proof. ($\Leftarrow$) If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are linearly independent eigenvectors of A with eigenvalues $\lambda_1, \ldots, \lambda_n$, take P to be the matrix with columns $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then $AP = PD$ by direct computation, so $A = PDP^{-1}$. ($\Rightarrow$) If $A = PDP^{-1}$, the columns of P are eigenvectors (by the same direct computation), and P invertible means they are linearly independent. $\square$

The factorization makes iterated dynamics transparent: $A^t = PD^tP^{-1}$, and D^t is the diagonal matrix of tth powers of the eigenvalues. So $\boldsymbol{x}_t = A^t\boldsymbol{x}_0 = PD^tP^{-1}\boldsymbol{x}_0$, which in the eigenvector basis is just per-coordinate scaling.

Two facts about eigenvalues are worth noting before we move to symmetric matrices.

Proposition 6. *If $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of A with corresponding eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$, then the eigenvectors are linearly independent. In particular, an $n \times n$ matrix with n distinct eigenvalues is diagonalizable.*

The proof is a clean induction on k; see Axler (2015, §5.A).

When eigenvalues coincide, diagonalization can fail. The standard non-diagonalizable example is the matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, which has 0 as a double eigenvalue but only a one-dimensional space of eigenvectors. Such matrices are called *defective*; the structural patch is the Jordan canonical form, which we will not pursue here. Most matrices arising in PE applications are diagonalizable, and the symmetric matrices of the next section are always diagonalizable.

4 Symmetric matrices and the spectral theorem

A great many of the matrices a political-economy modeler actually computes with are symmetric, and not by accident. The covariance matrix of any vector of random variables is symmetric, by the symmetry of covariance: $\operatorname{Cov}(X_i, X_j) = \operatorname{Cov}(X_j, X_i)$. The Hessian of a twice-differentiable utility function is symmetric, by the equality of mixed partial derivatives (Clairaut's theorem). The adjacency matrix of any undirected network — where "connection" is a symmetric relation — is symmetric. Distance matrices, similarity matrices, and correlation matrices in measurement work are all symmetric. Symmetric matrices, in short, are the ones the working political scientist actually has in hand when she sits down to compute. And they have a structural property that makes them especially well-behaved: a real symmetric matrix is *always* diagonalizable, can always be diagonalized with an orthonormal eigenbasis, and has all real eigenvalues. This is the spectral theorem, and it is the single most-used result in applied linear algebra. Principal-component analysis is the spectral decomposition of a covariance matrix. Eigenvector centrality on an undirected network is the principal eigenvector of the adjacency matrix. The second-order conditions for optimization, taken up in the next cluster, work because Hessians are symmetric and so spectrally well-behaved. The spectral theorem is the backbone these applications rest on, and the rest of the section develops it.

Theorem 7 (Spectral theorem for real symmetric matrices)**.** *Let $A \in \mathbb{R}^{n\times n}$ be symmetric ($A = A^\top$). Then:*

- *All eigenvalues of A are real.*
- *Eigenvectors corresponding to distinct eigenvalues are orthogonal.*
- *There exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A. Equivalently, $A = QDQ^\top$ for some orthogonal Q (so $Q^\top Q = \mathbf{I}$) and diagonal D.*

Proof sketch. **Real eigenvalues.** If $A\boldsymbol{v} = \lambda \boldsymbol{v}$ with $\boldsymbol{v} \in \mathbb{C}^n$ non-zero, then $\overline{\boldsymbol{v}}^\top A\boldsymbol{v} = \lambda \overline{\boldsymbol{v}}^\top \boldsymbol{v}$. The left side equals $\overline{\boldsymbol{v}}^\top A^\top \boldsymbol{v} = (A\overline{\boldsymbol{v}})^\top \boldsymbol{v} = (\overline{\lambda}\overline{\boldsymbol{v}})^\top \boldsymbol{v} = \overline{\lambda}\overline{\boldsymbol{v}}^\top \boldsymbol{v}$ (using A real and symmetric to commute conjugation past it). Since $\overline{\boldsymbol{v}}^\top \boldsymbol{v} = \|\boldsymbol{v}\|^2 \neq 0$, we get $\lambda = \overline{\lambda}$, so $\lambda \in \mathbb{R}$.

Orthogonality of distinct-eigenvalue eigenvectors. If $A\boldsymbol{v} = \lambda\boldsymbol{v}$ and $A\boldsymbol{w} = \mu\boldsymbol{w}$ with $\lambda \neq \mu$, then $\lambda\langle \boldsymbol{v}, \boldsymbol{w}\rangle = \langle A\boldsymbol{v}, \boldsymbol{w}\rangle = \langle \boldsymbol{v}, A^\top\boldsymbol{w}\rangle = \langle \boldsymbol{v}, A\boldsymbol{w}\rangle = \mu\langle \boldsymbol{v}, \boldsymbol{w}\rangle$, so $(\lambda - \mu)\langle \boldsymbol{v}, \boldsymbol{w}\rangle = 0$, hence $\langle \boldsymbol{v}, \boldsymbol{w}\rangle = 0$.

Orthonormal eigenbasis. Induction on n. Pick a unit eigenvector $\boldsymbol{v}_1$ corresponding to some eigenvalue λ_1 (which exists by the fundamental theorem of algebra applied to the characteristic polynomial). The orthogonal complement $\boldsymbol{v}_1^\perp$ is A-invariant (because A is symmetric and $\boldsymbol{v}_1$ is an eigenvector); the restriction of A to $\boldsymbol{v}_1^\perp$ is again symmetric, so by induction has an orthonormal eigenbasis. Together with $\boldsymbol{v}_1$, this gives an orthonormal eigenbasis of $\mathbb{R}^n$. □

The matrix-form statement $A = QDQ^\top$ is what gets used in computation. The orthogonality $Q^\top Q = \mathbf{I}$ replaces the more general P^{-1} from non-symmetric diagonalization; in particular, $Q^{-1} = Q^\top$, so the inverse of the change-of-basis matrix is computed simply by transposing. This is what makes the spectral decomposition a stable numerical operation, and it is one of the structural reasons symmetric problems are well-behaved in practice.[1]

---

[1] The spectral theorem extends to a wider class of matrices — the *normal* matrices, those satisfying $AA^\top = A^\top A$, of which symmetric matrices are a special case. Over $\mathbb{C}$, the right setting is the *Hermitian* matrices ($A = A^*$, where $*$ is conjugate transpose), and the spectral theorem says: every Hermitian matrix is unitarily diagonalizable with real eigenvalues. The real symmetric case in this handout is the part we need; the Hermitian case is the structural successor for complex-valued problems (signal processing, quantum information), which we do not pursue. The non-normal case can be defective (no eigenbasis exists) and is the home of the Jordan canonical form. The structural picture: symmetric (or Hermitian) is the right structural condition for an orthogonal (or unitary) eigenbasis to exist, with all eigenvalues real; everything beyond that is technical patching of cases that do not enjoy the same nice structure. The standard graduate references are Horn and Johnson (2013, Ch. 2), Lax (2007, Ch. 6), and Axler (2015, §7); for the specifically applied perspective, Strang (2019, Ch. 5) treats the spectral theorem as the structural content of PCA and the singular-value decomposition.

5 Quadratic forms and definiteness

Several political-economy primitives that look quite different at first glance are the same kind of mathematical object underneath. A spatial voter's quadratic-loss utility from a policy at displacement $\boldsymbol{p}-\boldsymbol{x}_v$ from her ideal point is $-\|\boldsymbol{p}-\boldsymbol{x}_v\|^2 = -(\boldsymbol{p}-\boldsymbol{x}_v)^\top(\boldsymbol{p}-\boldsymbol{x}_v)$ — a function of the squared distance. A measurement scholar's Mahalanobis distance between two cleavage-position vectors weights the squared coordinate differences unequally, by the inverse covariance: $d_\Sigma(\boldsymbol{x},\boldsymbol{y})^2 = (\boldsymbol{x}-\boldsymbol{y})^\top\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{y})$. The variance of a linear combination $\boldsymbol{a}^\top\boldsymbol{X}$ of correlated random variables is $\mathrm{Var}(\boldsymbol{a}^\top\boldsymbol{X}) = \boldsymbol{a}^\top\Sigma\boldsymbol{a}$ for the covariance matrix Σ. The second-order Taylor approximation of any twice-differentiable function f around a point $\boldsymbol{x}$, in leading order, has the shape $\frac{1}{2}\boldsymbol{h}^\top H_f(\boldsymbol{x})\boldsymbol{h}$ in the displacement $\boldsymbol{h}$, where H_f is the Hessian. Each of these is a function of the form $\boldsymbol{x}^\top A\boldsymbol{x}$ for some symmetric matrix A — a *quadratic form.* The natural substantive question to ask of a quadratic form is whether it is always non-negative (the voter always loses, the variance is non-negative, the function is locally bowl-shaped), always non-positive, or has mixed signs (the function has a saddle point, the policy weighting is internally inconsistent). That question is the question of *definiteness*, and the spectral theorem of the previous section answers it cleanly in eigenvalue terms. The optimization handout in the next cluster will identify concavity of f with negative semidefiniteness of its Hessian and quasiconcavity with a related condition on the bordered Hessian; the Sylvester-criterion machinery of this section is what those optimization tests actually do at runtime.

Definition 8. A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is:

- *positive definite* if $\boldsymbol{x}^\top A\boldsymbol{x} > 0$ for every non-zero $\boldsymbol{x} \in \mathbb{R}^n$;
- *positive semidefinite* if $\boldsymbol{x}^\top A\boldsymbol{x} \geq 0$ for every $\boldsymbol{x} \in \mathbb{R}^n$;
- *negative definite* if $\boldsymbol{x}^\top A\boldsymbol{x} < 0$ for every non-zero $\boldsymbol{x}$;
- *negative semidefinite* if $\boldsymbol{x}^\top A\boldsymbol{x} \leq 0$ for every $\boldsymbol{x}$;
- *indefinite* if it takes both strictly positive and strictly negative values.

The spectral theorem provides the cleanest characterization of definiteness in eigenvalue terms.

Theorem 9 (Definiteness via eigenvalues)**.** *Let $A \in \mathbb{R}^{n\times n}$ be symmetric, with eigenvalues $\lambda_1,\ldots,\lambda_n \in \mathbb{R}$. Then:*

- *A is positive definite iff every $\lambda_i > 0$;*
- *A is positive semidefinite iff every $\lambda_i \geq 0$;*
- *A is negative definite iff every $\lambda_i < 0$;*
- *A is negative semidefinite iff every $\lambda_i \leq 0$;*
- *A is indefinite iff A has both positive and negative eigenvalues.*

Proof. By the spectral theorem, $A = QDQ^\top$ with Q orthogonal and $D = \mathrm{diag}(\lambda_1,\ldots,\lambda_n)$. The change of variables $\boldsymbol{y} = Q^\top\boldsymbol{x}$ is a bijection of $\mathbb{R}^n$ to itself with $\boldsymbol{y}^\top\boldsymbol{y} = \boldsymbol{x}^\top\boldsymbol{x}$ and $\boldsymbol{x}^\top A\boldsymbol{x} = \boldsymbol{y}^\top D\boldsymbol{y} = \sum_i \lambda_i y_i^2$. The signs of $\sum_i \lambda_i y_i^2$ over $\boldsymbol{y} \neq \boldsymbol{0}$ are determined by the signs of λ_i, giving each of the equivalences. $\square$

For matrices small enough to factor by hand (typically 2 × 2 or 3 × 3), Theorem 9 is the practical test: compute eigenvalues, read off signs. For larger matrices a determinant-based test is faster.

Theorem 10 (Sylvester's criterion)**.** *A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive definite iff every leading principal minor (the determinant of the top-left $k \times k$ submatrix, for $k = 1, \ldots, n$) is strictly positive. A is negative definite iff the leading principal minors alternate in sign starting with the* 1×1 *minor strictly negative.*

For positive semidefiniteness, the analogous criterion uses *all* principal minors (not just leading); see Horn and Johnson (2013, §7.2). The proof of Sylvester's criterion is a careful induction on dimension; we omit it.[2]

6 Stochastic matrices and Perron–Frobenius

Voters switching parties between elections. Opinions diffusing across a fixed social network. The composition of a legislative body changing under term limits and replacement. The status profile of an authoritarian regime's elite under repeated purges. Each of these is a stochastic dynamic in which the relevant state at time $t + 1$ depends only on the state at time t (the Markovian assumption), and each is governed by a single matrix of transition probabilities. The questions a working political scientist actually asks of such a dynamic are: does the long-run composition of the system stabilize? Is it determined by the dynamics alone, regardless of where the system started? And how quickly does it get there? The questions are eigenvalue questions in disguise, and the structural answer to all three of them is the *Perron–Frobenius theorem*. The theorem says: every stochastic matrix has 1 as a real eigenvalue, with a non-negative eigenvector. Under irreducibility (informally: every state is reachable from every other), 1 is the largest eigenvalue in modulus and is simple, and the corresponding non-negative left eigenvector, normalized to sum to 1, is the chain's unique stationary distribution. Under the additional condition of aperiodicity, the chain converges to that stationary distribution from every starting point, with rate of convergence governed by the second-largest eigenvalue's modulus. The result is the structural backbone for the analysis of every Markov-chain-based political-economy model the next several literatures use, and it is what makes the polling-and-learning-and-replacement framework a tractable mathematical setting in the first place.

Definition 11. A square matrix $P \in \mathbb{R}^{n\times n}$ is *(row-)stochastic* if every entry is non-negative and every row sums to 1. A matrix is *column-stochastic* if its transpose is row-stochastic.

The row-stochastic case is the convention here: P_{ij} is the probability of moving from state i to state j. A probability distribution $\boldsymbol{\pi}$ over states is treated as a row vector, evolving under $\boldsymbol{\pi}_{t+1}^\top = \boldsymbol{\pi}_t^\top P$.

---

[2] The next cluster's optimization handout will identify a twice-differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ as *concave* on a convex domain iff its Hessian H_f is negative semidefinite at every point of the domain, and *strictly concave* iff H_f is negative definite. So checking concavity reduces to checking negative semidefiniteness of a symmetric matrix at each point — a Sylvester-criterion check on the leading principal minors. Quasiconcavity is more subtle: f is quasiconcave iff a related matrix called the *bordered Hessian* (the Hessian augmented by a row and column of gradient entries) satisfies a sign condition on its leading principal minors. The bordered-Hessian test is what the constrained-optimization literature uses for the second-order conditions on Lagrangian solutions: an interior maximum of f subject to a constraint $g(\boldsymbol{x}) = 0$ requires the bordered Hessian to be negative semidefinite on the tangent space of the constraint surface, which translates into a Sylvester-style test on the bordered matrix's leading principal minors. The full statement and proof are in the optimization handout to come; this footnote flags that the work this section's definiteness vocabulary will be doing in the next cluster is exactly the matrix-side of every concavity check the formal political-economy modeler ever runs.

The convention varies in the literature; column-stochastic matrices and column-vector probability distributions $\boldsymbol{\pi}_{t+1} = P\boldsymbol{\pi}_t$ are equally common.

Proposition 12. *A row-stochastic matrix P has 1 as an eigenvalue, with the all-ones vector $\mathbf{1}$ as a right eigenvector: $P\mathbf{1} = \mathbf{1}$.*

Proof. The ith entry of $P\mathbf{1}$ is $\sum_j P_{ij} = 1$ by the row-sum condition. $\square$

The right eigenvector with eigenvalue 1 is $\mathbf{1}$. The *left* eigenvector with eigenvalue 1 is the stationary distribution: a row vector $\boldsymbol{\pi}^\top$ with $\boldsymbol{\pi}^\top P = \boldsymbol{\pi}^\top$, $\sum_i \pi_i = 1$, $\pi_i \geq 0$. Existence is guaranteed by the previous proposition (transposing perspectives), but uniqueness and convergence to the stationary distribution are stronger statements that need additional structure on P.

Definition 13. A non-negative matrix P is *irreducible* if for every pair (i, j) there exists $t \geq 1$ such that $(P^t)_{ij} > 0$. (For a Markov chain: every state can be reached from every other state in some finite number of steps.)

Theorem 14 (Perron–Frobenius for stochastic matrices)**.** *Let $P \in \mathbb{R}^{n\times n}$ be a row-stochastic matrix.*

- *Every eigenvalue λ of P satisfies $|\lambda| \leq 1$.*
- 1 *is an eigenvalue, with at least one non-negative left eigenvector.*
- *If P is irreducible, then 1 is a* simple *eigenvalue (one-dimensional eigenspace), the corresponding left eigenvector $\boldsymbol{\pi}$ can be chosen with all entries strictly positive, and (after normalization to $\sum_i \pi_i = 1$) it is the unique stationary distribution.*
- *If P is irreducible and* aperiodic *(the gcd of cycle lengths is 1), then for every initial distribution $\boldsymbol{\pi}_0$, the iterates $\boldsymbol{\pi}_t$ converge to $\boldsymbol{\pi}$ as $t \to \infty$, with rate of convergence determined by the second-largest eigenvalue's modulus.*

The proof has several pieces; Horn and Johnson (2013, Ch. 8) or Meyer (2000, Ch. 8) gives the full development. The key piece is that the iterated matrix powers P^t approach a limit matrix whose every row equals the stationary distribution $\boldsymbol{\pi}^\top$, which is the formal version of "the chain forgets its starting point."

Example 15 (Voter migration)**.** A two-party system with party affiliations $\{D, R\}$ has voters who switch parties between elections according to fixed probabilities. The transition matrix is

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix},$$

where α is the probability a D voter becomes an R, and β the reverse. With $\alpha, \beta \in (0, 1)$, P is irreducible and aperiodic, so by Perron–Frobenius the chain converges to a unique stationary distribution (π_D, π_R). Computing: $\boldsymbol{\pi}^\top P = \boldsymbol{\pi}^\top$ gives $\pi_D(1-\alpha) + \pi_R\beta = \pi_D$ and the sum-to-one constraint gives $\pi_D = \beta/(\alpha+\beta)$, $\pi_R = \alpha/(\alpha+\beta)$. Long-run partisan composition is governed by the ratio of switch rates, regardless of where the chain starts. The second-largest eigenvalue is $1 - \alpha - \beta$, so the convergence rate is governed by how big $\alpha + \beta$ is — high mobility means fast convergence to the stationary distribution.

7 What's next

This handout closes the linear-algebra cluster. Three strands extend it.

Markov chains as a topic in their own right: the next handout in this region will pick up where Perron–Frobenius leaves off, treating discrete-time Markov chains, recurrence and transience, hitting times, mixing rates, and the structural place of Markov chains in formal political-economy modeling (voter migration, opinion dynamics, learning).

Optimization: the second cluster after this one. The optimization handout will identify a function as concave iff its Hessian is negative semidefinite at every interior point of the domain, and quasiconcave iff a bordered-Hessian sign condition holds. Sylvester's criterion from §5 is the diagnostic; the section's quadratic-forms machinery is exactly the input the optimization handout's second-order conditions take as given.

Spectral applications in measurement and inference: principal components analysis is the spectral decomposition of a covariance matrix; eigenvector centrality on a network is the principal eigenvector of the adjacency matrix; spectral clustering uses the eigenvectors of a graph Laplacian. Each is a direct application of the spectral theorem to a substantive measurement problem in political science. We do not develop these here, but the structural backing is now in place.

For graduate-level treatments at this handout's level of abstraction, Axler (2015, §5–§7) is the standard accessible reference. Horn and Johnson (2013) is the canonical encyclopedic treatment of eigenvalue theory and the Perron–Frobenius theorem. Lax (2007) gives a tighter, more applied perspective. For Perron–Frobenius and Markov chains specifically, Meyer (2000, Ch. 8) is the cleanest standalone treatment.

8 Exercises

Exercise 16. Compute the eigenvalues and eigenvectors of $A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. Verify that the eigenvectors are orthogonal (consistent with the spectral theorem, since A is symmetric).

Exercise 17. The rotation matrix $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ has complex eigenvalues $e^{\pm i\theta}$. (a) Verify this directly. (b) Argue, without computing eigenvectors, that R_θ has no real eigenvectors when $\theta \neq 0, \pi$. (Geometric intuition: rotation has no fixed direction.)

Exercise 18. *Voter learning, diagonalized.* Let A be the social-learning matrix of Example 3 with $N = 3$ voters, with weights a_{ij} such that $A = \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \frac{2}{3}\mathbf{I}$ (each voter puts equal weight on all three voters' previous beliefs and on her own). (a) Find the eigenvalues and eigenvectors of A. (b) Diagonalize A. (c) Use the diagonalization to compute A^{10}, and interpret the result.

Exercise 19. *Hessian of a quadratic-loss utility.* A spatial voter has quadratic-loss utility $u_v(\boldsymbol{p}) = -(\boldsymbol{p} - \boldsymbol{x}_v)^\top A(\boldsymbol{p} - \boldsymbol{x}_v)$ for a symmetric matrix A and ideal point $\boldsymbol{x}_v \in \mathbb{R}^k$. (a) Compute the Hessian of u_v as a function of $\boldsymbol{p}$. (b) Use Theorem 9 to show that u_v is strictly concave iff A is positive definite. (c) Interpret: positive-definite A means the voter has a unique optimum (her ideal point); indefinite A means the voter has saddle points and no unique best policy.

Exercise 20. *Sylvester's criterion in dimension 2.* A 2 × 2 symmetric matrix $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is positive definite iff $a > 0$ and $\det A = ac - b^2 > 0$, by Sylvester's criterion. Verify this directly by computing eigenvalues and showing both are positive iff $a > 0$ and $ac > b^2$.

Exercise 21. *Eigenvector centrality.* A small social network on $N = 4$ legislators has adjacency matrix

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

(a) Verify A is symmetric. (b) Find the largest eigenvalue and its corresponding eigenvector. (c) The principal eigenvector, normalized so its entries sum to 1, is the eigenvector centrality of the network. Which legislator is most central? Discuss in one sentence whether the answer matches what one would expect from the network's adjacency structure.

Exercise 22. *Voter migration via Perron–Frobenius.* Continuing Example 15: take $\alpha = 0.1$ and $\beta = 0.05$. (a) Compute the stationary distribution. (b) Compute the second-largest eigenvalue and the corresponding decay rate. (c) Suppose the initial partisan composition is $\boldsymbol{\pi}_0 = (0.5, 0.5)$ and the chain has been running for many elections. Approximately what is $\boldsymbol{\pi}_{20}$? (d) Discuss in one sentence: under what additional assumption (specifically, beyond what Perron–Frobenius requires) would the convergence to the stationary distribution be rapid versus slow?

Exercise 23. A non-symmetric 2 × 2 matrix A has complex eigenvalues $\lambda = a \pm bi$ with $a, b \in \mathbb{R}$, $b \neq 0$. (a) Show that $|\lambda|^2 = a^2 + b^2$. (b) Argue that the iterated dynamics $\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t$ converges to $\mathbf{0}$ iff $a^2 + b^2 < 1$. (c) Discuss in one sentence: what is the political-economy reading of complex eigenvalues in a dynamical model? (Hint: the imaginary part corresponds to oscillation; the real part corresponds to growth or decay.)

Exercise 24. *Mahalanobis distance as a quadratic form.* The Mahalanobis distance between two points $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^k$ with respect to a symmetric positive-definite matrix Σ is $d_\Sigma(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{y})}$. (a) Show that d_Σ is a metric on $\mathbb{R}^k$ (i.e., satisfies the metric axioms). (b) The standard interpretation in measurement uses Σ as a covariance matrix, and Σ^{-1} then weights dimensions inversely to their variance, so high-variance dimensions count for less. Discuss in one sentence why this is the right scaling for a measure of cleavage-based political distance.

Exercise 25. Give an example of a 2 × 2 matrix that is diagonalizable but not symmetric. Then give an example that is not diagonalizable. (Hint: distinct eigenvalues guarantee diagonalizability; coincident eigenvalues are where defectiveness can occur.)

References

Axler, Sheldon (2015). *Linear Algebra Done Right.* 3rd ed. Cham: Springer.

Horn, Roger A. and Charles R. Johnson (2013). *Matrix Analysis.* 2nd ed. Cambridge: Cambridge University Press.

Lax, Peter D. (2007). *Linear Algebra and Its Applications.* 2nd ed. Hoboken, NJ: Wiley.

Meyer, Carl D. (2000). *Matrix Analysis and Applied Linear Algebra.* Philadelphia: SIAM.

Strang, Gilbert (2019). *Linear Algebra and Learning from Data.* Wellesley, MA: Wellesley-Cambridge Press.