Vector spaces and linear maps

Robert J. Carroll

Last revised: 6 May 2026

---

1 Motivation

Most of the political-economy models that go beyond a single dimension live in vector spaces. A voter's ideal point in a spatial model is a vector in $\mathbb{R}^k$, with k the number of policy dimensions the analyst takes seriously. A legislator's coalition affiliation, viewed across all N legislators, is a $\{0, 1\}$-valued indicator vector in $\mathbb{R}^N$. A social-network connection structure is a matrix. A regression of vote shares on demographic covariates is, formally, a linear map. A Markov chain governing voter migration between parties has a stochastic matrix as its transition rule. None of these objects works without the vocabulary the next two handouts develop.

This handout treats the bookkeeping and the basic structure: what a vector space is, what a linear map is, how matrices encode linear maps once we fix coordinates, and how inner products supply the geometric notions of length and angle that "centrist," "moderate," and "ideologically distant" all rely on. The next handout takes up eigenvalues, diagonalization, and the spectral theorem, with quadratic forms and Perron–Frobenius as the closing payoffs — the dynamical and definitional results that the optimization, Markov-chain, and network-analysis literatures all lean on.

We work throughout over $\mathbb{R}$. The complex-number generalization is mostly a matter of replacing transposes with conjugate transposes and inner products with sesquilinear ones; the structural content carries over. None of the political-economy applications considered in the next several clusters needs complex scalars, so we stay real.

2 Vector spaces and subspaces

$\mathbb{R}^n$ is the canonical example, but a great many of the formal political-economy applications use objects that are *structurally like* $\mathbb{R}^n$ without literally being it. A vote-share allocation across k candidates lives in a $(k-1)$-dimensional simplex inside $\mathbb{R}^k$, not in $\mathbb{R}^k$ itself. A per-legislator effort level across an N-member chamber is a function $\{1, \ldots, N\} \to \mathbb{R}$, not an N-tuple of distinguished coordinates. The affine hull of a coalition's negotiable policies sits inside the ambient policy space as a strict subspace, smaller than the whole. None of these is $\mathbb{R}^n$ on the nose, but each inherits the same algebra: we can add, we can scale, and the operations have the same structural properties they do in $\mathbb{R}^n$. We want one definition that covers all of these at once, so the structural results we prove apply unchanged across all of them. The definition is the *vector space.*

Definition 1. A *vector space over* $\mathbb{R}$ is a set V together with operations $+ : V \times V \to V$ (*vector addition*) and $\cdot : \mathbb{R} \times V \to V$ (*scalar multiplication*) such that, for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and all $a, b \in \mathbb{R}$:

- $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) \quad$ and $\quad \boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$;
- there is a $\mathbf{0} \in V$ with $\boldsymbol{v} + \mathbf{0} = \boldsymbol{v}$ for every $\boldsymbol{v}$, and for each $\boldsymbol{v}$ there is a $-\boldsymbol{v}$ with $\boldsymbol{v} + (-\boldsymbol{v}) = \mathbf{0}$;

- $a(b\boldsymbol{v}) = (ab)\boldsymbol{v}$, $1 \cdot \boldsymbol{v} = \boldsymbol{v}$;
- $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$, $(a + b)\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v}$.

The elements of V are *vectors* and the elements of $\mathbb{R}$ are *scalars*.

Example 2 ($\mathbb{R}^n$)**.** The n-tuples of real numbers, with componentwise addition and componentwise scalar multiplication, form a vector space. The standard PE setting: n is the number of policy dimensions, and a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ represents an alternative or an ideal point.

Example 3 (Function spaces)**.** For any finite set X, the set $\mathbb{R}^X$ of functions $f : X \to \mathbb{R}$ is a vector space under pointwise operations. Two natural readings: X is a set of candidates and f a vote-share allocation; X is a set of voters and f a per-voter quantity (such as ideological position on one dimension).

Example 4 (Polynomials)**.** The set of polynomials of degree at most n with real coefficients forms a vector space of dimension $n + 1$. Less central to PE applications but useful as a structurally rich finite-dimensional non-$\mathbb{R}^n$ example.

A subspace is a subset that inherits the vector-space structure from its parent.

Definition 5. A subset $W \subseteq V$ is a *subspace* of V if $\mathbf{0} \in W$ and W is closed under addition and scalar multiplication: $\boldsymbol{u}, \boldsymbol{v} \in W$ and $a \in \mathbb{R}$ imply $\boldsymbol{u} + \boldsymbol{v} \in W$ and $a\boldsymbol{u} \in W$. Equivalently, W is closed under arbitrary finite *linear combinations* $\sum_i a_i \boldsymbol{v}_i$.

Example 6 (Coalition affiliation as a vector)**.** For an N-member legislature with member set $\{1, \ldots, N\}$, a coalition $C \subseteq \{1, \ldots, N\}$ is encoded as the indicator vector $\mathbf{1}_C \in \mathbb{R}^N$ with ith entry equal to 1 if $i \in C$ and 0 otherwise. The coalition lattice from order theory becomes the set $\{0, 1\}^N \subseteq \mathbb{R}^N$ of indicator vectors, with set operations corresponding to coordinatewise Boolean operations on indicators ($\mathbf{1}_{C \cap C'}$ is the coordinatewise minimum, $\mathbf{1}_{C \cup C'}$ the coordinatewise maximum). The whole $\mathbb{R}^N$ is the natural vector space the coalition data lives in, and the indicator-vectors are a structurally distinguished subset rather than a subspace (they are not closed under addition).

3 Linear independence, span, basis, dimension

How many dimensions of conflict are there in U.S. politics? Roll-call voting analysis (Poole and Rosenthal's NOMINATE) typically finds two; the issue-bundle scaling literature finds more; some media-economy and median-voter models work with one. Whichever answer is right empirically, "dimension" is doing real work in the question — it is the number of independent ways in which a position can be specified, the size of a minimal description of the space the agents are choosing from. Getting that notion correct is the work of this section. The substantive payoff is a theorem: for any finite-dimensional vector space, the dimension is a well-defined integer, intrinsic to the space and independent of the basis we use to compute it. Once we commit to a vector space, its dimension is fixed, and the structural conclusions we draw about "dimensional" arguments (existence of Condorcet winners in one dimension, Plott chaos in two or more, etc.) rest on this well-definedness.

Definition 7. The *span* of a finite collection $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in V$ is the set of all linear combinations:

$$\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\} = \left\{ \sum_{i=1}^{m} a_i \boldsymbol{v}_i : a_1, \ldots, a_m \in \mathbb{R} \right\}.$$

The span is always a subspace of V.

Definition 8. A finite collection $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m \in V$ is *linearly independent* if the only linear combination equal to $\mathbf{0}$ is the trivial one: $\sum_i a_i \boldsymbol{v}_i = \mathbf{0}$ implies $a_1 = \cdots = a_m = 0$. Otherwise, the collection is *linearly dependent.*

Definition 9. A finite collection $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in V$ is a *basis* of V if it is linearly independent and spans V. A vector space is *finite-dimensional* if it has a finite basis; otherwise it is *infinite-dimensional.*

The standard basis of $\mathbb{R}^n$ is $\boldsymbol{e}_1 = (1, 0, \ldots, 0), \ldots, \boldsymbol{e}_n = (0, \ldots, 0, 1)$. The standard basis of $\mathbb{R}^X$ (for finite X) is the family of indicator functions $\{\mathbf{1}_{\{x\}} : x \in X\}$.

The substantive theorem is that the size of a basis is intrinsic.

Theorem 10 (Dimension is well-defined)**.** *If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m$ are both bases of the same finite-dimensional vector space V, then $n = m$.*

The proof uses the *Steinitz exchange lemma*: in a finite-dimensional vector space, a linearly independent set can be extended to a basis, and a spanning set can be reduced to a basis, so any linearly independent set has size at most that of any spanning set. Applying this twice gives $n \leq m$ and $m \leq n$. We omit the algorithmic details; Axler (2015, Ch. 2) or Halmos (1958, Ch. 1) give the standard treatment.

Definition 11. The *dimension* of a finite-dimensional vector space V, written $\dim V$, is the cardinality of any basis.

So $\dim \mathbb{R}^n = n$, $\dim \mathbb{R}^X = |X|$, and $\dim(\text{degree-}\leq\text{-}n \text{ polynomials}) = n + 1$.[1]

4 Linear maps

Vector spaces themselves are static. The action is in the maps between them. A regression converts a vector of covariates into a predicted outcome. A weighted-voting rule converts a coalition-membership vector into a winning-or-losing decision. A network's adjacency matrix converts a vector of node-level quantities into a vector of neighbor-aggregates. A Markov-chain transition rule converts a probability distribution over states at time t into a probability distribution over states at $t + 1$. Each of these is a function $T : V \to W$ between vector spaces, and each respects the vector-space structure on both sides in a particular way: T sends sums to sums and scalar multiples to scalar multiples. The structural-respecting condition is what makes the function a *linear map*, and it is essentially every interesting kind of map between vector spaces in this project.

[1] The dimension of the policy space is one of those quantities that political-economy models choose almost in passing but on which substantive conclusions can hinge. The single-dimensional median voter theorem (Black 1948, Downs 1957) requires that policy be one-dimensional with single-peaked preferences; the model's predictions break down in two or more dimensions, where Plott's (1967) results show generic non-existence of Condorcet winners. Empirical work using NOMINATE (Poole and Rosenthal, Enderton, 2001, mentioned here only for the dimension-counting question, not for the technique itself) typically finds two dimensions explain most of the variation in U.S. congressional roll-call voting since the early Republic, with the second dimension shifting in interpretation across eras (race in the mid-20th century, region earlier, fiscal-versus-cultural in the late 20th). The empirical question of how many dimensions a polity actually has is therefore live, with substantive theoretical implications for which existence theorems apply and which do not. The handout's pedagogical point is that dimension is a well-defined invariant; the political-science point is that picking the dimension wrong is picking the model wrong.

Definition 12. Let V and W be vector spaces. A function $T : V \to W$ is a *linear map* (or *linear transformation*) if, for all $\boldsymbol{u}, \boldsymbol{v} \in V$ and all $a \in \mathbb{R}$:

$$T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v}), \qquad T(a\boldsymbol{v}) = aT(\boldsymbol{v}).$$

Example 13 (Aggregation rule as a linear map)**.** A *weighted-voting rule* on an N-member legislature with weights $w_1, \ldots, w_N \in \mathbb{R}$ is the linear map $T : \mathbb{R}^N \to \mathbb{R}$ defined by $T(\boldsymbol{x}) = \sum_i w_i x_i$. Restricted to indicator vectors $\mathbf{1}_C$ with C a coalition, this returns the coalition's total weight. Linearity — $T(\mathbf{1}_C + \mathbf{1}_{C'}) = T(\mathbf{1}_C) + T(\mathbf{1}_{C'})$ when the coalitions are disjoint — is the additivity property weighted-voting rules are usually presented with. Non-additive rules (rules that take into account which combinations of members are present) are not linear maps; the question of which voting rules are linear is the question of which are weighted.

Example 14 (Regression)**.** A regression of an outcome $y \in \mathbb{R}$ on covariates $\boldsymbol{x} \in \mathbb{R}^k$ via the model $y = \boldsymbol{\beta}^\top \boldsymbol{x} + \epsilon$ has the deterministic part as a linear map from $\mathbb{R}^k$ to $\mathbb{R}$. The map's coefficient vector $\boldsymbol{\beta}$ is what regression estimation tries to recover; the residual ϵ is the noise. Linearity in the covariates is the structural assumption that the literature on nonlinear regression weakens.

Two associated subspaces are central:

Definition 15. For a linear map $T : V \to W$, the *kernel* of T is

$$\ker T = \{\boldsymbol{v} \in V : T(\boldsymbol{v}) = \mathbf{0}\},$$

a subspace of V. The *image* of T is

$$\operatorname{im} T = \{T(\boldsymbol{v}) : \boldsymbol{v} \in V\},$$

a subspace of W. The *rank* of T is $\dim \operatorname{im} T$, and the *nullity* of T is $\dim \ker T$.

The relationship between kernel and image is the rank-nullity theorem.

Theorem 16 (Rank-nullity)**.** *Let $T : V \to W$ be a linear map with V finite-dimensional. Then*

$$\dim V = \dim \ker T + \dim \operatorname{im} T.$$

Proof. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ be a basis of $\ker T$, and extend to a basis $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ of V, where $k + m = \dim V$. We show that $T(\boldsymbol{v}_1), \ldots, T(\boldsymbol{v}_m)$ is a basis of $\operatorname{im} T$.

Spanning. Any $T(\boldsymbol{w}) \in \operatorname{im} T$ has $\boldsymbol{w} = \sum_i a_i \boldsymbol{u}_i + \sum_j b_j \boldsymbol{v}_j$ for some scalars. Applying T and using $T(\boldsymbol{u}_i) = \mathbf{0}$ gives $T(\boldsymbol{w}) = \sum_j b_j T(\boldsymbol{v}_j)$.

Linear independence. If $\sum_j c_j T(\boldsymbol{v}_j) = \mathbf{0}$, then $T(\sum_j c_j \boldsymbol{v}_j) = \mathbf{0}$, so $\sum_j c_j \boldsymbol{v}_j \in \ker T$. Express $\sum_j c_j \boldsymbol{v}_j = \sum_i d_i \boldsymbol{u}_i$ for some d_i. The basis $\boldsymbol{u}_1, \ldots, \boldsymbol{v}_m$ of V is linearly independent, so all $c_j = 0$ (and all $d_i = 0$).

So $\dim \operatorname{im} T = m = \dim V - k = \dim V - \dim \ker T$, which rearranges to the claim. $\square$

The theorem has two readings. The arithmetic reading: dimensions on the two sides of T are conserved up to the kernel correction. The structural reading: T "loses" information equal in dimension to its kernel and "carries" information equal to its image, with the total bounded by the source's dimension. In the regression setting (Example 14), a rank-deficient covariate matrix has a non-trivial kernel, which is the formal version of multicollinearity — the parameter vector $\boldsymbol{\beta}$ is not identified because there is a non-zero $\boldsymbol{\beta}$ that maps to the zero residual.

5 Matrices and matrix representations

Working political economists almost never compute with abstract linear maps. They compute with *matrices* — the regression coefficient vector, the transition matrix of a Markov chain, the adjacency matrix of a social network, the design matrix of a regression. The relationship between the matrices that show up in practice and the linear maps of the previous section is what this section pins down. Once we fix a basis of the source vector space and a basis of the target, every linear map gets encoded as a rectangular array of scalars, and every rectangular array of scalars encodes a linear map. Matrix multiplication corresponds to function composition. The two perspectives are interchangeable: choosing whether to think in terms of the abstract linear map or the concrete matrix is a matter of whether the immediate question is structural (where the linear-map view is cleaner) or computational (where the matrix view is what the computer eats).

Definition 17. An $m \times n$ *matrix* is a rectangular array $A = (a_{ij})$ of real numbers with m rows and n columns, $1 \le i \le m$, $1 \le j \le n$. Equivalently, A is a function from $\{1, \dots, m\} \times \{1, \dots, n\}$ to $\mathbb{R}$.

Definition 18. An $m \times n$ matrix A defines a linear map $T_A : \mathbb{R}^n \to \mathbb{R}^m$ by $T_A(\boldsymbol{x}) = A\boldsymbol{x}$, where $(A\boldsymbol{x})_i = \sum_j a_{ij} x_j$. The map T_A is the linear map whose value on the jth standard basis vector $\boldsymbol{e}_j$ is the jth column of A.

Proposition 19. *Every linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is T_A for a unique $m \times n$ matrix A, namely the matrix whose jth column is $T(\boldsymbol{e}_j)$.*

The proposition extends from $\mathbb{R}^n$ to a general finite-dimensional V once a basis of V is fixed: a linear map $T : V \to W$ is encoded by a matrix A relative to chosen bases $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ of V and $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_m\}$ of W, with a_{ij} the ith coordinate of $T(\boldsymbol{v}_j)$ in the W-basis. Different choices of bases give different matrices for the same linear map; the change-of-basis machinery (Definition 23 below) tracks the dependence.

Definition 20. For an $m \times n$ matrix A and an $n \times p$ matrix B, the *matrix product* AB is the $m \times p$ matrix with

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Proposition 21. *Matrix multiplication corresponds to composition of linear maps: if A encodes $T : \mathbb{R}^n \to \mathbb{R}^m$ and B encodes $S : \mathbb{R}^p \to \mathbb{R}^n$, then AB encodes $T \circ S : \mathbb{R}^p \to \mathbb{R}^m$.*

Proof. $(T \circ S)(\boldsymbol{x}) = T(S(\boldsymbol{x})) = T(B\boldsymbol{x}) = AB\boldsymbol{x} = (AB)\boldsymbol{x}$, using the definitions of T, S, and matrix multiplication. The encoding-by-matrix is unique by the previous proposition. $\square$

The associativity of matrix multiplication, $A(BC) = (AB)C$, is then a corollary of the associativity of function composition. Non-commutativity ($AB \neq BA$ in general) is the corresponding shadow of the non-commutativity of function composition.

Definition 22. The $n \times n$ *identity matrix* $\mathbf{I}_n$ has $(\mathbf{I}_n)_{ij} = 1$ if $i = j$ and 0 otherwise; it encodes the identity map on $\mathbb{R}^n$. The *transpose* $A^\top$ of an $m \times n$ matrix A is the $n \times m$ matrix with $(A^\top)_{ij} = a_{ji}$. We use $^\top$ rather than the more common A^T to avoid clashing with T for a generic linear map.

Definition 23. Let $\mathcal{B} = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ and $\mathcal{B}' = (\boldsymbol{v}'_1, \dots, \boldsymbol{v}'_n)$ be two bases of V. The *change-of-basis matrix* P from $\mathcal{B}$ to $\mathcal{B}'$ has jth column equal to the coordinates of $\boldsymbol{v}_j$ in the basis $\mathcal{B}'$. If A is the matrix of $T : V \to V$ in basis $\mathcal{B}$, then the matrix in $\mathcal{B}'$ is PAP^{-1}.

The change-of-basis transformation $A \mapsto PAP^{-1}$ is called *conjugation* of A by P, and it is the algebraic content of saying "the linear map is the same, but its coordinate representation changed."

6 Inner products and orthogonality

What does it mean for two voters to be *ideologically close*? For a candidate to be *moderate*? For one policy proposal to be a *hedge* against another? Each is a geometric question — about distances, angles, or perpendicularity in the policy space — and the algebraic vector-space machinery developed so far does not by itself answer any of them. We can add and scale ideal points in $\mathbb{R}^k$, and we can map ideal points to outcomes by linear regressions, but we cannot yet say two ideal points are "close" in any sense more precise than the $\mathbb{R}^n$ coordinate distance. The vocabulary that supplies the missing geometry is the *inner product*: a single bilinear function on $V \times V$ that gives a length for every vector, an angle between any two, and the orthogonality and projection structure that nearly every spatial-modeling argument in political economy — and the spectral theorem in the next handout — rests on.

Definition 24. A (real) *inner product* on a vector space V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ that is:

- *bilinear*: linear in each argument separately;
- *symmetric*: $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$;
- *positive-definite*: $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$ with equality iff $\boldsymbol{v} = \boldsymbol{0}$.

The associated *norm* is $\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$.

The standard inner product on $\mathbb{R}^n$ is $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y} = \sum_i x_i y_i$, with the corresponding standard norm $\|\boldsymbol{x}\| = \sqrt{\sum_i x_i^2}$ (the Euclidean norm). Most of the political-economy applications use this standard inner product, but the abstract definition lets us recognize the structure in non-Euclidean settings (weighted inner products, function-space inner products) without rederiving the basic results.

Theorem 25 (Cauchy–Schwarz)**.** *For all $\boldsymbol{u}, \boldsymbol{v} \in V$ in any inner-product space,*

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \|\boldsymbol{v}\|.$$

Equality holds iff $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent.

The proof is the standard one: expand $\|\boldsymbol{u} - t\boldsymbol{v}\|^2 \geq 0$ at the optimal $t = \langle \boldsymbol{u}, \boldsymbol{v} \rangle / \|\boldsymbol{v}\|^2$. We omit the algebra; see Axler (2015, §6.A).

Two vectors are *orthogonal*, written $\boldsymbol{u} \perp \boldsymbol{v}$, if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$. A finite collection $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ is *orthogonal* if every pair is, and *orthonormal* if additionally each $\|\boldsymbol{u}_i\| = 1$. An orthonormal basis is a basis that is also orthonormal.

Theorem 26 (Gram–Schmidt)**.** *Every linearly independent finite set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ in an inner-product space can be turned into an orthonormal set $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ with the same span by an explicit iterative algorithm.*

The algorithm: $\boldsymbol{u}_1 = \boldsymbol{v}_1/\|\boldsymbol{v}_1\|$; for $k > 1$, project out the components along $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{k-1}$ from $\boldsymbol{v}_k$ and normalize. The proof is just verifying that the construction has the claimed properties; see Axler (2015, §6.B).

Definition 27. For a subspace $W \subseteq V$, the *orthogonal projection* of $\boldsymbol{v} \in V$ onto W, written $\mathrm{proj}_W(\boldsymbol{v})$, is the unique $\boldsymbol{w} \in W$ minimizing $\|\boldsymbol{v} - \boldsymbol{w}\|$. Equivalently, $\boldsymbol{v} - \mathrm{proj}_W(\boldsymbol{v})$ is orthogonal to every element of W.

The connection to political-economy modeling runs through the geometric reading of distance.[2]

7 What's next

The next handout takes up eigenvalue theory. The headline objects are:

- *Eigenvalues and eigenvectors* for square matrices, with the diagonalization theorem identifying when a matrix has an eigenbasis. The political-economy applications are dynamical: voter learning as a linear iteration converges at a rate determined by the eigenvalues; eigenvector centrality on a social network identifies influential nodes via the principal eigenvector of the adjacency matrix.
- *The spectral theorem for symmetric matrices.* Symmetric matrices have orthonormal eigenbases and real eigenvalues; this is the structural result behind quadratic forms, second-order conditions in optimization, and most of the dimensionality-reduction techniques (PCA) used in measurement.
- *Quadratic forms and definiteness.* Positive and negative (semi)definiteness of symmetric matrices is exactly the notion needed for concavity and convexity of twice-differentiable functions. The optimization handout will identify a function as concave iff its Hessian is negative semidefinite, and quasiconcave iff a bordered-Hessian sign condition holds; this handout's section on definiteness will be the diagnostic.
- *Stochastic matrices and Perron–Frobenius.* A stochastic matrix has 1 as an eigenvalue with a non-negative eigenvector, and under irreducibility this eigenvalue is simple. The result is the structural reason Markov chains converge to a unique stationary distribution, and it sets up the eventual Markov-chains handout.

For graduate-level treatments at this handout's level of abstraction, Axler (2015) is the standard accessible reference (the second printing's celebrated approach defers determinants to the end).

[2] The inner product is what makes "ideologically close" a precise notion. In a k-dimensional spatial-voting model with voter ideal points $\boldsymbol{x}_v$ and policy alternatives $\boldsymbol{p}$, the standard quadratic-loss utility is $u_v(\boldsymbol{p}) = -\|\boldsymbol{p} - \boldsymbol{x}_v\|^2$, which is a function entirely of inner products: expanding gives $u_v(\boldsymbol{p}) = -\langle \boldsymbol{p}, \boldsymbol{p}\rangle + 2\langle \boldsymbol{p}, \boldsymbol{x}_v\rangle - \langle \boldsymbol{x}_v, \boldsymbol{x}_v\rangle$. The voter's optimal policy is her own ideal point; her preference over alternatives is governed by the inner-product structure of the policy space. "Centrist" becomes an inner-product statement: voter v is centrist if her ideal point $\boldsymbol{x}_v$ projects close to the centroid of the electorate. "Two voters are ideologically close" becomes $\langle \boldsymbol{x}_v - \boldsymbol{x}_{v'}, \boldsymbol{x}_v - \boldsymbol{x}_{v'}\rangle$ small. The empirical-modeling literature on cleavages and party systems is, in formal terms, a literature on the inner-product structure of the policy space; Hinich and Munger (1997) treats the spatial-modeling tradition under exactly this rubric. The choice of inner product is itself substantive: an unweighted Euclidean inner product implicitly weights all policy dimensions equally, which is rarely the empirically right move, and weighted inner products with different weights on different dimensions encode the analytical claim that some dimensions matter more than others.

Halmos (1958) is the classical treatment; Strang (2019) is the more applied / numerical take. For the political-economy-flavored use of multidimensional spatial models, Hinich and Munger (1997) is the canonical reference.

8 Exercises

Exercise 28. Show that the set of indicator vectors $\{\mathbf{1}_C : C \subseteq \{1, \ldots, N\}\} \subseteq \mathbb{R}^N$ is *not* a subspace. (The simplest argument: $\mathbf{1}_{\{1\}} + \mathbf{1}_{\{1\}} = (2, 0, \ldots, 0) \notin \{0,1\}^N$.) What is the smallest subspace of $\mathbb{R}^N$ containing every indicator vector?

Exercise 29. In a 2-dimensional spatial-voting model with policy space $\mathbb{R}^2$, three legislators have ideal points $\boldsymbol{x}_1 = (1, 0)$, $\boldsymbol{x}_2 = (0, 1)$, $\boldsymbol{x}_3 = (1, 1)$. (a) Are $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ linearly independent? (b) What is the dimension of $\operatorname{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$? (c) Describe the affine span (linear span shifted to pass through $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$) and interpret it as a coalitional bargaining region.

Exercise 30. Verify that the standard basis vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ form a basis of $\mathbb{R}^n$. Then construct a different basis of $\mathbb{R}^3$ explicitly and compute the change-of-basis matrix P from the standard basis to your basis.

Exercise 31. Let $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be rotation by angle θ counterclockwise about the origin. Show that T_θ is linear. Compute its matrix in the standard basis. (This is the canonical example where geometry and algebra coincide.)

Exercise 32. *Weighted-voting kernel.* Continuing Example 13: let $T : \mathbb{R}^N \to \mathbb{R}$ be a weighted-voting rule with weights $w_1, \ldots, w_N$ summing to $W > 0$. (a) Compute $\dim \ker T$ and $\dim \operatorname{im} T$, and verify rank-nullity. (b) Identify $\ker T$ explicitly as a subspace — which coalitions of weights add to zero? (c) The minimal-winning-coalition condition (the voting threshold $W/2$) cuts the indicator-vector subset $\{0,1\}^N$ into the winning and losing coalitions; this is *not* a kernel-image cut on the linear map. Discuss in one sentence why the linear-algebra apparatus does not by itself characterize the winning condition.

Exercise 33. *Multicollinearity in regression.* A linear regression of $y \in \mathbb{R}^n$ on covariates given by an $n \times k$ design matrix X has the linear map $T : \mathbb{R}^k \to \mathbb{R}^n$, $\boldsymbol{\beta} \mapsto X\boldsymbol{\beta}$. Show that the regression coefficient vector $\boldsymbol{\beta}$ is identified from $X\boldsymbol{\beta}$ iff T is injective, iff $\ker T = \{\mathbf{0}\}$, iff the columns of X are linearly independent. The phenomenon known as multicollinearity is exactly the failure of this last condition.

Exercise 34. Verify on a small example that matrix multiplication corresponds to composition of linear maps. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(x_1, x_2) = (x_1 + 2x_2, 3x_1 - x_2)$ and $S : \mathbb{R}^2 \to \mathbb{R}^2$ by $S(x_1, x_2) = (-x_1 + x_2, 2x_2)$. (a) Compute the matrices A of T and B of S in the standard basis. (b) Compute AB and verify directly that AB encodes $T \circ S$.

Exercise 35. Use Cauchy–Schwarz to prove the triangle inequality for the Euclidean norm: $\|\boldsymbol{u} + \boldsymbol{v}\| \le \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$ for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. (Hint: square both sides.)

Exercise 36. *Projection onto a coalition.* Continuing Example 6: in $\mathbb{R}^N$ with the standard inner product, the subspace $W_C = \operatorname{span}\{\boldsymbol{e}_i : i \in C\}$ is the "C-coalition coordinate plane." (a) Compute $\operatorname{proj}_{W_C}(\boldsymbol{x})$ explicitly for an arbitrary $\boldsymbol{x} \in \mathbb{R}^N$. (b) Show that the squared distance $\|\boldsymbol{x} - \operatorname{proj}_{W_C}(\boldsymbol{x})\|^2$ is exactly the sum of squared values of $\boldsymbol{x}$ outside C. (c) Interpret: if $\boldsymbol{x}$ encodes a per-legislator quantity (effort, support level), the squared distance is the total amount of that quantity attributable to non-C legislators.

Exercise 37. Apply the Gram–Schmidt algorithm to the vectors $(1, 1, 0), (1, 0, 1), (0, 1, 1)$ in $\mathbb{R}^3$ with the standard inner product. Verify that the result is orthonormal and spans the same subspace.

References

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

Enderton, Herbert B. (2001). *A Mathematical Introduction to Logic*. 2nd ed. San Diego: Academic Press.

Halmos, Paul R. (1958). *Finite-Dimensional Vector Spaces*. 2nd ed. Princeton: Van Nostrand.

Hinich, Melvin J. and Michael C. Munger (1997). *Analytical Politics*. Cambridge: Cambridge University Press.

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