Chapter 1

Finite Dimensional Vector Spaces

If you can't explain it simply, you don't understand it well enough.

– Albert Einstein

THIS chapter will introduce vector spaces (especially finite dimensions), which are fundamental in many areas of mathematics. It provides a simple but general framework for linear algebra and functional spaces, which becomes the cornerstone for modern analysis and applied mathematics in the 20th century. For a comprehensive study, refer to the first few chapters of the book by Peter Lax Lax (2014).

1.1 Vector Space

A vector space V over a field $\mathbb{F}$ (e.g. $\mathbb{R}$ or $\mathbb{C}$) is a set equipped with "addition" and "scalar multiplication" that satisfy *vector axioms* (see Table 1.1).

Axiom	Statement
Associativity	$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
Commutativity	$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
Additive Identity	There exists $\mathbf{0} \in V$ that $\mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v}$
Additive Inverse	For any $\mathbf{v} \in V$, there exists $-\mathbf{v} \in V$, $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$
Compatibility	$a(b\mathbf{v}) = (ab)\mathbf{v}$
Scalar Identity	There exists $1 \in \mathbb{F}$, $1\mathbf{v} = \mathbf{v}$
Distributivity I	$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
Distributivity II	$(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$

Table 1.1: Vector axioms.

Remark 1.1.1. *These axioms are* not *independent. For instance, if we drop the "commutability", then consider*

$$(1+1)(\mathbf{u}+\mathbf{v}) \xlongequal{\text{Distributivity II}} (\mathbf{u}+\mathbf{v}) + (\mathbf{u}+\mathbf{v}) \xlongequal{\text{Associativity}} \mathbf{u}+\mathbf{v}+\mathbf{u}+\mathbf{v}$$

and

$$(1+1)(\mathbf{u}+\mathbf{v}) = 2\mathbf{u} + 2\mathbf{v} = \mathbf{u}+\mathbf{u}+\mathbf{v}+\mathbf{v}$$

Canceling $\mathbf{u}$ and $\mathbf{v}$ on two sides recovers the commutativity. A complete discussion of the independent vector axioms can be found in Rigby and Wiegold (1973).

Definition 1.1.2. *Once the axioms are satisfied, the set V becomes a linear vector space if V is closed under vector addition and scalar multiplication.*

Example 1.1.3. *We list a few common examples of linear vector spaces.*

1. *The set of all polynomials of degree n is a linear vector space.*
2. *The set of all continuous functions defined on a closed interval $\mathcal{I}$ is a linear vector space.*
3. *The set of all infinite sequences $(a_1, a_2, \cdots)$, where each entry $a_i \in \mathbb{R}$, is a linear vector space.*

4. *The set of all Riemann-integrable functions is a linear vector space.*

Definition 1.1.4. *A subset W of a linear space V is called a linear subspace if W is also closed under vector addition and scalar multiplication.*

Theorem 1.1.5. *The following claims are true.*

1. *The sets $\{\mathbf{0}\}$ and V are linear subspaces of V.*
2. *The sum of two subspaces is a subspace. Here the sum of two sets S and T is defined by*
$$S + T := \{s + t \mid s \in S, t \in T\}.$$
3. *The intersection of two subspaces is a subspace.*

Proof. The proof is straightforward, left as an exercise (☞). □

Remark 1.1.6. *Note that the union of two subspaces S and T may not form a linear subspace unless $S \subset T$ or $T \subset S$. Otherwise, there exist $s \in S \backslash T$ and $t \in T \backslash S$, but $s + t \notin S \cup T$.*

Definition 1.1.7. *The intersection $\cap_\sigma V_\sigma$ of all linear subspaces $V_\sigma \subset V$ containing the set S is called the* ***linear span*** *of S.*

Theorem 1.1.8. *The linear span of S is the smallest linear subspace containing S. The linear span of S consists of all elements x of the form*

$$x = \sum_i a_i x_i, \quad x_i \in S, a_i \in \mathbb{F} \tag{1.1}$$

Proof. The first claim is precisely the definition. For the second claim, notice that the form of (1.1) is the smallest linear subspace containing the set S. □

Example 1.1.9. *Let $S = \{x^2 - 1, x^4 - x^2, x^3 - x^2, x^4 - x^3\}$, then the smallest linear subspace containing S is*

$$\{a_4 x^4 + a_3 x^3 + a_2 x^2 - (a_2 + a_3 + a_4) \mid a_2, a_3, a_4 \in \mathbb{F}\}.$$

Definition 1.1.10. *A subset $S \subset V$ is called **spanning set** of V if $V = \{\sum_i \alpha_i s_i \mid s_i \in S, \alpha_i \in \mathbb{F}\}$, that is, every element in V can be written as a linear combination of elements in S.*

Example 1.1.11. *For instance, $\{1, x, x^2, \cdots, x^n\}$ forms a spanning set of all polynomials of degree n. Since any polynomial $p(x) = a_n x^n + \cdots + a_1 x + a_0$ can be written a linear combination of the monomials.*

Definition 1.1.12. *A set $S \subset V$ is **linearly independent** if*

$$\sum_i a_i s_i = \mathbf{0}$$

only has the solution $a_i = 0$ for all i.

Definition 1.1.13. *If a spanning set $S \subset V$ is linearly independent, then S forms a **basis** for V. When the basis S is finite, the cardinality $|S|$ is the **dimension** of the space.*

Remark 1.1.14. *The concept of dimension does not depend on the choice of the linearly independent spanning set. If two linearly independent sets S_α, S_β have different numbers of elements $n_\alpha > n_\beta$, then each element $s_{\beta,j} \in S_\beta$ can be represented by*

$$\sum_{i=1}^{n_\alpha} a_{ij} s_{\alpha,i} = s_{\beta,j}$$

While the under-determined linear system

$$\sum_{i=1}^{n_\alpha} a_{ij} c_j = 0, \quad j = 1, \cdots, n_\beta$$

permits a nonzero solution $(c_1, \cdots, c_{n_\beta})$. It implies that there exists a set of nonzero coefficients that

$$\sum_{j=1}^{n_\beta} c_j s_{\beta,j} = \mathbf{0}.$$

It contradicts linear independence.

Example 1.1.15. *In $\mathbb{R}^3$, the vectors $\{(1,0,0), (0,1,0), (0,0,1)\}$ are a linearly inde-*

pendent spanning set. But $\{(1,0,0),(0,1,0),(0,0,1),(1,0,1)\}$ *is not a linearly independent spanning set.*

Example 1.1.16. *The set of all polynomials is a linear space. Still, it does not have a finite dimension since any linear combination of a finite number of polynomials has a finite degree.*

1.2 Inner Product

An inner product in a linear space V over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$) is a $\mathbb{F}$ valued function of two vectors $x, y \in V$, denoted by $\langle x, y\rangle$, having the following properties:

1. sesquilinearity: $\langle ax_1 + bx_2, y\rangle = a\langle x_1, y\rangle + b\langle x_2, y\rangle$.
2. skew symmetry (Hermitian): $\langle x, y\rangle = \overline{\langle y, x\rangle}$.
3. positivity: $\langle x, x\rangle > 0$ if $x \neq \mathbf{0}$. $\langle x, x\rangle = 0$ if and only if $x = \mathbf{0}$.

Remark 1.2.1. *The inner product induces a norm function, denoted by* $\|\cdot\|$,

$$\|x\| = \langle x, x\rangle^{1/2}.$$

To prove this indeed defines a norm, we need the following famous inequality to derive the triangle inequality $\|x + y\| \leq \|x\| + \|y\|$.

Lemma 1.2.2 (Cauchy-Schwartz)**.**

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

Proof. Without loss of generality, we assume $y \neq \mathbf{0}$. First, we choose $\alpha \in \mathbb{C}$ that $|\alpha| = 1$ to make $\alpha\langle x, y\rangle \in \mathbb{R}$. For any $t \in \mathbb{R}$,

$$0 \leq \inf_{t\in\mathbb{R}} \|\alpha x + ty\|^2 = \|x\|^2 + \inf_{t\in\mathbb{R}} \left(t^2\|y\|^2 + 2t\alpha\langle x, y\rangle\right) = \|x\|^2 - \frac{|\alpha\langle x, y\rangle|^2}{\|y\|^2}.$$

The infimum is achieved at $t = -\frac{\langle \alpha x, y\rangle}{\|y\|^2} \in \mathbb{R}$, which implies $\|x\|\|y\| \geq |\alpha\langle x, y\rangle| = |\langle x, y\rangle|$. $\square$

The proof of the following sub-additivity is left as an exercise ( ).

Corollary 1.2.3 (sub-additivity). *The triangle inequality holds.*

$$\|x+y\| \leq \|x\| + \|y\|.$$

Definition 1.2.4. *A linear vector space with an inner product is called* ***inner product space.***

Example 1.2.5 (Inner product).

1. *In the usual Euclidean space* $\mathbb{R}^n$ *or* $\mathbb{C}^n$*, the inner product is define as*

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i \overline{y_i}.$$

2. *Let* $\mathcal{C}_\infty([0,1])$ *be the set of continuous functions on* $[0,1]$*, then*

$$\langle f, g\rangle = \int_0^1 f(x)\overline{g(x)}dx$$

defines an inner product on $\mathcal{C}_\infty([0,1])$.

3. *Let* $\mathcal{C}_\infty^m([0,1])$ *be the set of* m*-times continuously differentiable functions on* $[0,1]$*, then we can define the inner product by*

$$\langle f, g\rangle = \int_0^1 \left(\sum_{k=0}^{m} \frac{d^k f(x)}{dx^k} \overline{\frac{d^k g(x)}{dx^k}} \right) dx.$$

Definition 1.2.6. *Two vectors* x *and* y *are* ***orthogonal*** *if* $\langle x, y\rangle = 0$.

Example 1.2.7 (Orthogonality).

1. *In Euclidean space, the definition of orthogonality agrees with the geometric orthogonality.*

2. *With the inner product* $\langle f, g\rangle = \int_{-1}^{1} f(x)\overline{g(x)}dx$*, the functions* $\sin(k\pi x)$ *and* $\sin(m\pi x)$ *are orthogonal as long as the integers* $m \neq k$.

The mutually orthogonal vectors provide an effective basis. Suppose that the set

$$S = \{s_1, s_2, \cdots, s_n\} \subset V$$

satisfies

$$\langle s_i, s_j \rangle = 0, \quad i \neq j.$$

Then they are linearly independent (why?). To represent any $f \in V$ as a linear combination of the basis S, that is,

$$f = \sum_{i=1}^{n} \alpha_i s_i.$$

We take the inner product with s_k, and find that

$$\langle f, s_k \rangle = \sum_{i=1}^{n} \alpha_i \langle s_i, s_k \rangle = \alpha_k \|s_k\|^2 \Rightarrow \alpha_k = \frac{\langle f, s_k \rangle}{\|s_k\|^2}.$$

The **Gram-Schmidt** method constructs a mutually orthogonal set from any linearly independent set (not necessarily a basis) of vectors $x_1, x_2, \cdots, x_n \in V$. Take $\psi_1 = x_1$, then construct $\psi_2 = x_2 - a\psi_1$ that satisfies the orthogonality $\langle \psi_1, \psi_2 \rangle = 0$, and we obtain $a = \frac{\langle x_2, \psi_1 \rangle}{\|\psi_1\|^2}$. Iteratively using this idea, we can construct

$$\psi_k = x_k - \sum_{j=1}^{k-1} \frac{\langle x_k, \psi_j \rangle}{\|\psi_j\|^2} \psi_j, \quad k = 2, \cdots, n$$

We can check the mutual orthogonality $\langle \psi_k, \psi_j \rangle = 0$ for $j < k$ (prove it).

Example 1.2.8. *With the inner product* $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)dx$, *the mutually orthogonal polynomials* $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = x^2 - \frac{1}{3}$, *etc. can be constructed by applying Gram-Schmidt method to the linearly independent set* $\{1, x, x^2, \cdots\}$. *These polynomials are called* ***Legendre polynomials.***

1.3 Spectral Theory for Matrices

Let V and U be linear vector spaces over the same field $\mathbb{F}$.

Definition 1.3.1. *A transformation* $T : V \to U$ *is called* ***linear*** *if it satisfies*

$$T(x + y) = Tx + Ty, \quad T(ax) = aTx.$$

Using the above definition, one can prove the following theorem easily (proof omitted).

Theorem 1.3.2. *The image of a linear subspace $W \subset V$ under a linear transformation $T : V \to U$ is a linear subspace of U.*

Let's consider a linear transformation $T : V \to V$ on a finite-dimensional vector space V that maps into itself. Let $(x_1, \cdots, x_n)$ be a linearly independent basis of V. Suppose $y_k = Tx_k \in V$, then there exists coefficients $\{a_{ik}\}_{1\le i,k\le n}$

$$y_k = \sum_{i=1}^{n} a_{ik}x_i.$$

Then any element $x = \sum_{j=1}^{n} c_j x_j$, its image under T is

$$Tx = \sum_{j=1}^{n} c_j y_j = \sum_{j=1}^{n} a_j \sum_{i=1}^{n} a_{ij}u_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij}c_j \right) x_i.$$

The effect of the linear transformation T can be viewed as the transformation of coefficients under matrix multiplication with $A = (a_{ij}) \in \mathbb{F}^{n\times n}$. With a different basis $(z_1, \cdots, z_n)$ of V, there exist the coefficients $\Theta = (\theta_{ij}) \in \mathbb{F}^{n\times n}$ that

$$z_i = \sum_{l=1}^{m} \theta_{li}x_l,$$

then the effect of T under the new basis $(z_1, \cdots, z_n)$ can be viewed as the matrix multiplication with $\Theta^{-1}A\Theta$.

Definition 1.3.3. *The transformation in the form of $A' = \Theta^{-1}A\Theta$ is called **similarity transformation**.*

Out of all possible Θ, we wish to find a particular change of basis that can make the resulting matrix $A' = \Lambda$ diagonal (this is not always guaranteed). That is, $C\Theta = \Lambda\Theta$, if $\mathbf{v}_i$ is the ith column of Θ, then it is equivalent to the following problem

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i,$$

where λ_i is the ith entry on the diagonal of Λ. This is also called the **eigenvalue problem**.

Definition 1.3.4. *An eigenpair of $A \in \mathbb{C}^{n\times n}$ is a pair $(\lambda, \mathbf{v})$, $\lambda \in \mathbb{C}$, $\mathbf{v} \in \mathbf{C}^n$ that*

$$A\mathbf{v} = \lambda \mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}.$$

The vector $\mathbf{v}$ is the eigenvector *and λ is the* eigenvalue.

Theorem 1.3.5. *A complex number $\lambda \in \mathbb{C}$ is an eigenvalue of A iff $\det(A-\lambda I) = 0$.*

The polynomial $\det(A - \lambda I)$ is the characteristic polynomial with exactly n complex roots (possibly with repetition). If an eigenvalue has a multiplicity ≥ 2 (algebraic multiplicity), it does not mean an equal number of eigenvectors (geometric multiplicity). For instance, the following matrix (Jordon block)

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

There is only one eigenvalue equal to 1 of multiplicity 3, but only one eigenvector $(1, 0, 0)$. The geometric multiplicity cannot be larger than the algebraic multiplicity.

Theorem 1.3.6. *If A has n distinct eigenvalues, then A has n linearly independent eigenvectors.*

Proof. It remains to prove the eigenvectors are linearly independent; otherwise

$$\sum_{i=1}^n a_i \mathbf{v}_i = \mathbf{0}.$$

Repeatedly apply A to both sides, then

$$\sum_{i=1}^n a_i \lambda_i^{k-1} \mathbf{v}_i = \mathbf{0}, \quad k = 1, 2, \cdots, n$$

However, the Vandermonde matrix $M = (\lambda_i^{k-1})_{1\leq i\leq n, 1\leq k\leq n}$ is not singular, contradiction. □

Although the *diagonalization* for any matrix A is not possible, there exists one particular class of matrices satisfying our needs – *self-adjoint matrices*.

Definition 1.3.7. *The **transpose** of a matrix $A = (a_{ij})$ is $A^T = (a_{ji})$. The **adjoint** of a matrix A is defined as $A^* = \overline{A}^T$ (conjugate transpose).*

Definition 1.3.8. *If $A^* = A$, then the matrix A is called **self-adjoint** or **Hermitian**.*

Using the symmetric property, we can easily verify the following theorems for self-adjoint matrices.

Theorem 1.3.9. *If A is self-adjoint, then*

1. *$\langle Ax, x\rangle \in \mathbb{R}$.*
2. *All eigenvalues of A are real.*

Theorem 1.3.10. *If A is self-adjoint, then there exists an orthogonal basis*

$$\{x_1, \cdots, x_n\}$$

such that

1. *Each x_i is an eigenvector, that is, $Ax_i = \lambda_i x_i$.*
2. *$\langle x_i, x_j\rangle = \delta_{ij}$ (Kronecker delta).*
3. *The matrix*
$$Q = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$$
is unitary, that is, $Q^* = Q^{-1}$.
4. *$Q^*AQ = \Lambda$, the matrix Λ is diagonal with real diagonal entries λ_i.*

Definition 1.3.11. *The spectral radius $\rho(A) := \max\{|\lambda_1|, \cdots, |\lambda_n|\}$.*

It is straightforward (taking an eigenvector) to verify the following upper bound for $\rho(A)$.

Theorem 1.3.12. *If $\|\cdot\|$ is any induced matrix norm that*

$$\|A\| = \sup_{\|x\|=1} \|Ax\|,$$

then $\rho(A) \leq \|A\|$. Moreover, $\rho(A) \leq \|A^n\|^{1/n}$ for any $n \in \mathbb{N}$.

Especially, if we take the Euclidean 2-norm $\|\cdot\|_2$, then $\rho(A) = \|A\|_2$ when A is self-adjoint (why?).

Lemma 1.3.13. *If $\|\cdot\|$ is any induced matrix norm, then $\|AB\| \leq \|A\|\|B\|$.*

Theorem 1.3.14 (Gelfand's Formula)**.** *For any induced matrix norm $\|\cdot\|$,*

$$\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k}.$$

Proof. If A is diagonalizable, then the theorem is trivial. Otherwise, let A be similar to its Jordan form Horn and Johnson (2012),

$$A = P^{-1}JP, \quad J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{pmatrix},$$

each Jordan block J_i has size r_i. Then

$$\|A^n\|^{1/n} = \|P^{-1}J^nP\|^{1/n} \leq \|P^{-1}\|^{1/n}\|P\|^{1/n}\|J^n\|^{1/n} \xrightarrow{n\to\infty} \|J^n\|^{1/n},$$

and

$$\|A^n\|^{1/n} = \frac{(\|P\|\|P^{-1}J^nP\|\|P^{-1}\|)^{1/n}}{\|P\|^{1/n}\|P^{-1}\|^{1/n}} \geq \|J^n\|^{1/n}(\|P\|\|P^{-1}\|)^{-1/n} \xrightarrow{n\to\infty} \|J^n\|^{1/n}$$

The diagonally block matrix J^n is

$$J^n = \begin{pmatrix} J_1^n & & \\ & \ddots & \\ & & J_k^n \end{pmatrix},$$

Let $|J|$ be the entrywise absolute value of J, then

$$|J|^n = \begin{pmatrix} |J_1|^n & & \\ & \ddots & \\ & & |J_k|^n \end{pmatrix}.$$

Take a sequence of matrices B_m such that $B_m = |J| + \text{diag}(\varepsilon_{m,1}, \cdots, \varepsilon_{m,n})$ such that $\varepsilon_{m,\cdot} \to 0^+$ as $m \to \infty$ and the diagonal entries of B_m are distinct. Then B_m is diagonalizable for each m and $\lim_{m\to\infty} B_m = |J|$.

$$\limsup_{n\to\infty} \|J^n\|^{1/n} \leq \limsup_{n\to\infty} \||J|^n\|^{1/n} \leq \limsup_{n\to\infty} \|B_m^n\|^{1/n} = \rho(B_m).$$

Passing $m \to \infty$, and use the previous theorem 1.3.12

$$\limsup_{n\to\infty} \|J^n\|^{1/n} \leq \lim_{m\to\infty} \rho(B_m) = \rho(|J|) = \rho(J) = \rho(A) \leq \liminf_{n\to\infty} \|A^n\|^{1/n}.$$

□

1.4 Min-max Principle

The min-max principle characterizes a variational property of the eigenvalues. It is also called the *Courant-Fischer-Weyl* min-max principle.

Definition 1.4.1. *Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint. The Rayleigh quotient $R_A : \mathbb{C}^n - \{\mathbf{0}\} \to \mathbb{R}$ is*

$$R_A(x) := \frac{\langle Ax, x\rangle}{\langle x, x\rangle}.$$

Since $R_A(x)$ is a continuous function on the compact set $\|x\|_2 = 1$, it attains a maximum at a vector x_1. The following maximum principle shows that the vector x_1 is an eigenvector for the largest eigenvalue.

Theorem 1.4.2 (Maximum Principle). *Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint. The following statements hold.*

1. $\lambda_1 = \max_{\|x\|=1}\langle Ax, x\rangle = \langle Ax_1, x_1\rangle$ *is the largest eigenvalue of A, and x_1 is the corresponding normalized eigenvector to λ_1.*

2. *Let $\lambda_k = \max\langle Ax, x\rangle$ subject to the constraints*
   - $\langle x, x_j\rangle = 0$, $j = 1, 2, \cdots, k-1$.
   - $\|x\| = 1$.

   Then λ_k is the k-th eigenvalue of A (sorted in descending order) and x_k is the corresponding normalized eigenvector to λ_k.

Proof. We prove the first statement. Let $\{e_1, \cdots, e_n\}$ be the orthogonal basis formed by the eigenvectors. We may assume the corresponding eigenvalues are sorted in descending order $\lambda_1 \geq \cdots \geq \lambda_n$. Then let $x = \sum_{i=1}^n c_i e_i$ such that $\sum_{i=1}^n |c_i|^2 = 1$, then

$$\langle Ax, x\rangle = \sum_{i=1}^n \lambda_i |c_i|^2 \leq \lambda_1 \sum_{i=1}^n |c_i|^2.$$

The equal sign holds when x is an eigenvector of eigenvalue λ_1. For the second statement, we can derive similarly. □

The above maximum principle implies the following min-max principle of Courant.

Corollary 1.4.3 (Courant min-max principle). *Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint. Then*

$$\lambda_k = \min_{C\in\mathbb{C}^{(k-1)\times n}} \max_{\|x\|=1, Cx=\mathbf{0}} \langle Ax, x\rangle,$$

Proof. It is easy to show

$$\lambda_k \geq \min_{C\in\mathbb{C}^{(k-1)\times n}} \max_{\|x\|=1, Cx=\mathbf{0}} \langle Ax, x\rangle.$$

To show the converse, we consider the space $S = \text{span}\{e_1, e_2, \cdots, e_k\}$, then $\dim(S) = k$. Since $Cx = \mathbf{0}$ defines a linear subspace T of dimension $(n-k+1)$, we have $S \cap T$ has at least dimension one.

$$\max_{\|x\|=1, x\in T} \langle Ax, x\rangle \geq \max_{\|x\|=1, x\in S\cap T} \langle Ax, x\rangle \geq \min_{\|x\|=1, x\in T} \langle Ax, x\rangle = \lambda_k.$$

□

Theorem 1.4.4 (Courant-Fischer). *Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint. Then*

$$\lambda_k = \max_{S, \dim(S)=k} \min_{\|x\|=1, x\in S} \langle Ax, x\rangle$$

and

$$\lambda_k = \min_{T, \dim(T)=n-k+1} \max_{\|x\|=1, x\in T} \langle Ax, x\rangle,$$

where S and T are linear subspaces of $\mathbb{C}^n$.

Proof. We only need to prove the first identity. The idea is similar to Corollary 1.4.3. Take the subspace $U = \text{span}\{e_k, \cdots, e_n\}$, then $\dim(U) = n-k+1$ and $\dim(S \cap U) \geq 1$. Therefore,

$$\min_{\|x\|=1, x\in S} \langle Ax, x\rangle \leq \min_{\|x\|=1, x\in S\cap U} \langle Ax, x\rangle \leq \max_{\|x\|=1, x\in U} \langle Ax, x\rangle = \lambda_k.$$

If we take $S' = \text{span}\{e_1, \cdots, e_k\}$, then we have

$$\max_{S, \dim(S)=k} \min_{\|x\|=1, x\in S} \langle Ax, x\rangle \geq \min_{\|x\|=1, x\in S'} \langle Ax, x\rangle = \lambda_k.$$

□

1.5 Common Matrices

In this section, we introduce a few types of matrices and their properties.

1.5.1 Positive (Semi)definite Matrices

Definition 1.5.1. *Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint. If $\langle Ax, x\rangle$ is positive for all nonzero $x \in \mathbb{C}^n$. Then A is positive definite.*

Definition 1.5.2. *Let $A \in \mathbb{C}^{n\times n}$ be self-adjoint. If $\langle Ax, x\rangle$ is nonnegative for all nonzero $x \in \mathbb{C}^n$. Then A is positive semidefinite.*

It is straightforward to verify that the eigenvalues of a positive (semi)definite matrix are positive (nonnegative).

Corollary 1.5.3. *If A and B are self-adjoint and B is positive semi-definite, then*

$$\lambda_i(A+B) \geq \lambda_i(A), \quad i = 1, \cdots, n.$$

Proof. Using the min-max principle,

$$\begin{aligned}
\lambda_i(A+B) &= \max_{S, \dim(S)=i} \min_{\|x\|=1, x\in S} \langle (A+B)x, x\rangle \\
&= \max_{S, \dim(S)=i} \min_{\|x\|=1, x\in S} \langle Ax, x\rangle + \langle Bx, x\rangle \\
&\geq \max_{S, \dim(S)=i} \min_{\|x\|=1, x\in S} \langle Ax, x\rangle = \lambda_i(A).
\end{aligned}$$

□

Many related eigenvalue inequalities (e.g., Weyl inequality) can be found in Horn and Johnson (2012).

Definition 1.5.4. *The Hadmard product of two matrices $A, B \in \mathbb{C}^{n\times n}$ is defined by (entrywise)*

$$(A \odot B)_{ij} = A_{ij}B_{ij}, \quad 1 \leq i, j \leq n.$$

The following theorem states that the Hadamard product of two positive definite matrices is still positive definite.

Theorem 1.5.5 (Schur's Product Theorem). *If $A, B \in \mathbb{C}^{n\times n}$ are positive definite, then $A \odot B$ is also positive definite.*

Proof. Let $A = \sum_{i=1}^n \lambda_i v_i v_i^T$ and $B = \sum_{j=1}^n \mu_j u_j u_j^T$, where (λ_i, v_i), (μ_j, u_j) are the eigenpairs of A and B, respectively. Then

$$A \odot B = \sum_{i=1}^n \sum_{j=1}^n \lambda_i v_i v_i^T \odot \mu_j u_j u_j^T = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j v_i v_i^T \odot u_j u_j^T.$$

The matrix $v_i v_i^T \odot u_j u_j^T = (v_i \odot u_j)(v_i \odot u_j)^T$ can be verified easily, which is positive semidefinite. Therefore, $A \odot B$ is positive semidefinite. To show positive definiteness, we assume $a \neq \mathbf{0}$ such that

$$a^T(v_i \odot u_j)(v_i \odot u_j)^T a = 0$$

for all $1 \leq i, j \leq n$. This implies $(v_i \odot u_j)^T a = v_i^T(u_j \odot a) = 0$ for all $1 \leq i, j \leq n$. Since v_i form an orthonormal basis, we must have $u_j \odot a = \mathbf{0}$ for all $1 \leq j \leq n$, this implies $a = \mathbf{0}$, contradiction. $\square$

A closely related concept is the positive definite kernel.

Definition 1.5.6. *Let $\mathcal{X}$ be a non-empty set. Suppose $\mathcal{K}(x, y)$ is Hermitian, that is, $\mathcal{K}(x, y) = \overline{\mathcal{K}(y, x)}$ on $\mathcal{X} \times \mathcal{X} \to \mathbb{C}$. If*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_i \overline{c_j} \mathcal{K}(x_i, x_j) \geq 0$$

for any set of points $x_1, \cdots, x_n \in \mathcal{X}$ and $c_1, \cdots, c_n \in \mathbb{C}$, then the kernel $\mathcal{K}$ is positive definite.

In short, the kernel $\mathcal{K}$ is positive definite if the *kernel matrix* $K_{ij} := \mathcal{K}(x_i, x_j)$ is positive (semi)definite for arbitrary samples. Using Schur's Product Theorem, the following result is straightforward.

Corollary 1.5.7. *Suppose $\mathcal{K}_1$ and $\mathcal{K}_2$ are positive definite kernels on $\mathcal{X} \times \mathcal{X}$, then the product $\mathcal{K}_1\mathcal{K}_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is also positive definite.*

Example 1.5.8. *The following kernels are positive definite (why?)*

1. $\mathcal{K}(x, y) = x^T y$
2. $\mathcal{K}(x, y) = x^T y + r$, $r \geq 0$
3. $\mathcal{K}(x, y) = (x^T y + r)^n$, $r \geq 0$, $n \in \mathbb{N}$

For convolution kernels, their Fourier transforms (if they exist) determine whether they are positive definite.

Theorem 1.5.9 (Bochner). *Suppose $\mathcal{K}(x, y) = f(x - y)$ is Hermitian. If the Fourier transform $\widehat{f}$ is non-negative, then $\mathcal{K}(x, y)$ is positive definite on $\mathbb{R}^n \times \mathbb{R}^n$.*

Proof. The inverse Fourier transform of $\widehat{f}$ is

$$f(x) = \int_{\mathbb{R}} \widehat{f}(\zeta)e^{2\pi i\zeta x}dx$$

Therefore,

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n} f(x_i - x_j)c_i\overline{c_j} &= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\overline{c_j}\int_{\mathbb{R}^n} \widehat{f}(\zeta)e^{2\pi i\zeta(x_i-x_j)}d\zeta \\ &= \int_{\mathbb{R}^n} \widehat{f}(\zeta)\left|\sum_{i=1}^{n} c_i e^{2\pi i\zeta x_i}\right|^2 d\zeta \geq 0.\end{aligned}$$

□

Example 1.5.10. *The following kernels are positive definite (why?).*

1. $\mathcal{K}(x,y) = e^{-\alpha|x-y|}$, $\alpha > 0$
2. $\mathcal{K}(x,y) = e^{-\alpha|x-y|^2}$, $\alpha > 0$
3. $\mathcal{K}(x,y) = \mathrm{sinc}(\alpha|x-y|)$, $\alpha > 0$

1.5.2 Diagonally Dominant Matrices

The diagonally dominant matrices are useful in numerical analysis.

Definition 1.5.11. *Let $A \in \mathbb{C}^{n\times n}$. It is **diagonally dominant** if*

$$|A_{ii}| \geq \sum_{j\neq i}|A_{ij}|, \quad i = 1,\cdots,n.$$

*If the inequality is **strict**, then A is **strictly** diagonally dominant.*

Theorem 1.5.12 (Gershgorin Circle Theorem)**.** *Let $A \in \mathbb{C}^{n\times n}$. The eigenvalues of A are contained in the following set:*

$$\bigcup_{i=1}^{n}\{z \in \mathbb{C} \colon |z - A_{ii}| \leq \sum_{j\neq i}|A_{ij}|\}.$$

Proof. If (λ, x) is an eigenpair of A, that is, $Ax = \lambda x$. Let i be the index of the entry with the largest absolute value, then

$$\sum_{j=1}^{n} A_{ij} x_j = \lambda x_i,$$

which means

$$\sum_{j \neq i} A_{ij} \frac{x_j}{x_i} = \lambda - A_{ii}.$$

Take the absolute value on both sides,

$$|\lambda - A_{ii}| \leq \sum_{j \neq i} |A_{ij}|.$$

□

As a corollary of the Gershgorin Circle Theorem, the strictly diagonally dominant matrices are non-singular. The proof is straightforward (prove it).

Corollary 1.5.13. *Strictly diagonally dominant matrices are non-singular.*

In addition, if a diagonally dominant matrix is also Hermitian, we can conclude positive-(semi)definiteness from its spectrum.

Corollary 1.5.14. *If A is self-adjoint, diagonally dominant, and the diagonal entries of A are non-negative (positive), then A is positive semidefinite (definite).*

1.5.3 Non-negative/Positive Matrices

The positive matrices are useful in probability and graph theory. Note that the positive matrices are **not** the same as positive definite matrices.

Definition 1.5.15. *Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ be a matrix with entries $a_{ij} \geq 0$, then A is a nonnegative matrix.*

Definition 1.5.16. *Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ be a matrix with entries $a_{ij} > 0$, then A is a positive matrix.*

The property of being a positive (non-negative) matrix is closed under matrix multiplication (why?). When $x \in \mathbb{R}^n$ is entrywise positive, we have $Ax > 0$ when A is positive. Repeating this argument, we find $A^n x > 0$. Therefore, if the eigen-decomposition $x = \sum_{i=1}^n c_i e_i$, we will find

$$\frac{1}{\rho(A)^n} A^n x = \sum_{i=1}^n \left(\frac{\lambda_i}{\rho(A)} \right)^n c_i e_i \xrightarrow{n \to \infty} \sum_{|\lambda_i| = \rho(A)} e^{in \arg \lambda_i} c_i e_i, \tag{1.2}$$

which is still entrywise positive for different large n. Observe that

$$\frac{1}{M-N} \left| \sum_{n=N}^{M-1} e^{in\theta} \right| = \frac{1}{M-N} \left| \frac{1 - e^{i(M-N)\theta}}{1 - e^{i\theta}} \right|$$

can be made arbitrarily small if $\theta \neq 0$. Taking the average of the right-hand side of (1.2) with large $n \in \mathbb{N}$, the terms with nonzero arg λ_i will disappear. Then, repeating the same process, we find that $\sum_{\lambda_i = \rho(A)} c_i e_i$ should be entrywise positive. That means, the set $\{\lambda_i = \rho(A)\}$ is non-empty and a corresponding eigenvector is strictly positive.

A related famous result is the following **Perron-Frobenius theorem**, which shows the spectral radius is an eigenvalue of a positive matrix. There are many ways to prove it, see (Horn and Johnson, 2012, Chapter 8).

Theorem 1.5.17 (Perron-Frobenius). *Let $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ be a positive matrix. Then*

1. *The spectral radius $\rho(A)$ is the leading eigenvalue of A, and the eigenvalue is simple.*
2. *The associated eigenvector can be normalized to have strictly positive entries.*
3. *There are no other eigenvectors with strictly positive entries.*

Proof. ① We first show $\rho(A)$ is an eigenvalue. Let $(\lambda, \mathbf{v})$ be an eigenpair with $|\lambda| = \rho(A)$, we prove $\lambda \in \mathbb{R}^+$. Otherwise $\lambda = \rho(A)e^{i\theta}$ for certain $\theta \in (0, 2\pi)$, then there exists $m \in \mathbb{N}$ that $\Re e^{im\theta} < 0$. While A^m is still a positive matrix, $(\lambda^m, \mathbf{v})$ is an eigenpair of A^m, then if $\epsilon > 0$,

$$(A^m - \epsilon I)\mathbf{v} = (\lambda^m - \varepsilon)\mathbf{v}$$

implying that $\rho(A^m - \epsilon I) \geq |\lambda^m - \varepsilon| > \rho(A)^m$. On the other hand, if we take the matrix norm $\|A\| = \max_{1 \leq i,j \leq n} |a_{ij}|$ and choose ϵ small enough that $A^m - \epsilon I$ is still a positive matrix, then use Gelfand's formula,

$$\begin{aligned}\rho(A)^m < \rho(A^m - \epsilon I) &= \lim_{k\to\infty} \|(A^m - \epsilon I)^k\|^{1/k} \\ &\le \lim_{k\to\infty} \|(A^m)^k\|^{1/k} = \rho(A^m) \le \rho(A)^m.\end{aligned}$$

Contradiction.

② We show that the geometric multiplicity of $\lambda = \rho(A)$ must be one. Otherwise, there are two independent eigenvectors $\mathbf{v} > 0$ and $\mathbf{u}$ (we can assume it is real), then there exists a constant $t \in \mathbb{R}$ that $\mathbf{v}' = \mathbf{v} + t\mathbf{u} \geq 0$ with at least one entry equal to zero, then $\mathbf{v}'$ is an eigenvector of λ but $A\mathbf{v}'$ should be entrywise positive, contradiction.

③ We then show that the algebraic multiplicity of $\lambda = \rho(A)$ is also one. Take $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ be the eigenvectors of A and A^T corresponding to $\rho(A)$, then $A\mathbf{v} = \rho(A)\mathbf{v}$ and $\mathbf{w}^T A = \rho(A)\mathbf{w}^T$. Let $(\mathbf{w}, \mathbf{u}_1, \cdots, \mathbf{u}_{n-1})$ be an orthonormal basis for $\mathbb{R}^n$. If $\mathbf{u} \in \operatorname{span}\{\mathbf{u}_1, \cdots, \mathbf{u}_{n-1}\}$,

$$\mathbf{w}^T A\mathbf{u} = \rho\mathbf{w}^T\mathbf{u} = 0.$$

Therefore, $A\mathbf{u} \in \operatorname{span}\{\mathbf{u}_1, \cdots, \mathbf{u}_{n-1}\}$. Denote

$$P = \begin{pmatrix}\mathbf{v} & U\end{pmatrix}, \quad U = \begin{pmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_{n-1}\end{pmatrix}.$$

We can verify that

$$AP = P\left(\begin{array}{c|ccc}\rho(A) & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right)$$

for some matrix $B = U^{-1}AU \in \mathbb{R}^{(n-1)\times(n-1)}$. It means the Jordan block for $\rho(A)$ is of size 1×1.

④ If there are two eigenvectors with strictly positive entries. Suppose $(\mu, \mathbf{u})$ is another eigenpair that $\mathbf{u} > 0$ and $\mu < \rho(A)$, then we can take $t < 0$ that $\mathbf{u}' = \mathbf{u} + t\mathbf{v} \geq 0$ with at least one entry equal to zero. Then $A\mathbf{u}' = \mu\mathbf{u} + t\rho(A)\mathbf{v} < \mu\mathbf{u} + t\mu\mathbf{v} = \mu(\mathbf{u} + t\mathbf{v})$ has a negative entry, contradiction. □

For non-negative matrices, we can view them as the limiting process of the positive matrices.

Corollary 1.5.18 (Non-negative Matrices)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a non-negative matrix, then the spectral radius $\rho(A)$ is an eigenvalue of A and there is a non-negative*

Continue on next page ⇓

Continue from last page ⇑

eigenvector for it.

Example 1.5.19 (Markov Chain). *Let $P \in \mathbb{R}^{n\times n}$ be a stochastic matrix, that is, P is a non-negative matrix with*

$$\sum_{j=1}^{n} P_{ij} = 1, \quad \forall 1 \le i \le n.$$

The vector $\mathbf{1}$ is an eigenvector with eigenvalue 1. Thus, 1 is the spectral radius. Furthermore, the linear equation $(I-P)x = y$ has a unique solution x that $\langle x, \mathbf{1}\rangle = 0$.

Example 1.5.20 (PageRank). *The "pagerank" refers to an algorithm to rank webpages according to the number of links between the webpages. The "importannce" or "ranking" of P_i, denoted by $r(P_i)$, satisfies*

$$r(P_i) = \sum_{P_j \in d(B_i)} \frac{r(P_j)}{|d(B_j)|},$$

where $d(B_i)$ is the set of pages pointing to P_i. The right-hand side denotes the average contribution of "importance" from each page. The quantities $r(P_i)$ form a vector $\mathbf{r}$ solves

$$\mathbf{r} = \mathbf{r}H, \quad H_{ij} = \frac{1}{d(B_i)}.$$

The matrix H is non-negative, its spectral radius $\rho(H) = \rho(H^T) = 1$ is an eigenvalue. It is practical to consider another matrix

$$H' = \alpha H + (1-\alpha)\frac{1}{n}\mathbf{1}\mathbf{1}^T$$

to enforce the uniqueness of $\mathbf{r}$.

Example 1.5.21 (Lyapunov Exponent). *The Lyapunov exponent of a sequence of random matrices is*

$$\limsup_{n\to\infty} \frac{1}{2n} \log \|S_n S_n^T\|$$

*where $S_n = A_n \cdots A_1$ is the product of random matrices. If A_i are i.i.d. random **positive** matrices of determinant 1, then $S_n S_n^T$ is still a positive matrix of determinant 1. The norm $\|S_n S_n^T\| \ge \rho(S_n S_n^T)$. Thus, the spectral radius $\rho(S_n S_n^T) > 1$, meaning the Lyapunov exponent is positive, hence can cause "chaos".*

1.6 Exercises

☞ **Problem 1.6.1.** *Prove Theorem 1.1.5.*

☞ **Problem 1.6.2.** *Prove Corollary 1.2.3.*

☞ **Problem 1.6.3.** *Let $\|\cdot\|$ be a norm on the vector space V such that*

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2, \quad \forall x, y \in V.$$

Show that $\phi(x,y) = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2)$ defines an inner product. This is known as Jordan - von Neumann Theorem.

☕ **Problem 1.6.4.** *Prove the following Hardy's inequality.*

$$\int_{\mathbb{R}^3} \frac{|u(x)|^2}{|x|^2} dx \le 4 \int_{\mathbb{R}^3} |\nabla u(x)|^2 dx, \quad u \in C_c^\infty(\mathbb{R}^3).$$

Hint: Use the idea in Lemma 1.2.2 and the divergence theorem:

$$\int_{\mathbb{R}^3} \frac{xu(x)}{|x|^2} \cdot \nabla u(x) = \int_{\mathbb{R}^3} \frac{x}{|x|^2} \cdot \nabla \frac{|u(x)|^2}{2} dx = -\int_{\mathbb{R}^3} \frac{|u(x)|^2}{2} \nabla \cdot \frac{x}{|x|^2} dx.$$

☞ **Problem 1.6.5.** *Let $A \in \mathbb{C}^{n\times n}$ with spectral radius $\rho(A) < 1$, then*

$$\lim_{n\to\infty} A^n = \mathbf{0}.$$

☞ **Problem 1.6.6.** *Let $A \in \mathbb{R}^{n\times n}$ be a positive matrix, show that*

$$\rho(A) = \lim_{n\to\infty} (\mathrm{tr}(A^n))^{1/n}.$$

☕ **Problem 1.6.7.** *Let $\Omega \subset \mathbb{R}^2$ be an open convex set with a smooth boundary and $\{x_j\}_{j=1}^n \subset \partial\Omega$ be a set of distinct points on the boundary. Consider the matrix $T = (T_{ij}) \in \mathbb{R}^{n\times n}$ that*

$$T_{ij} = \frac{\mathbf{n}(x_j)\cdot(x_j - x_i)}{|x_i - x_j|^2},$$

where $\mathbf{n}(x_j)$ denotes the outward unit normal vector at x_j. Prove that

1. *The matrix T is well-defined, that is, every entry is a finite number.*
2. *The matrix T has only one eigenvector with positive entries.*

☕ **Problem 1.6.8.** *Suppose $\{x_i\}_{i=1}^n$ is a set of real numbers. Let the matrix $K \in \mathbb{R}^{n\times n}$ be defined by*

$$K_{ij} = e^{-\lambda[\sin(x_i - x_j)]^2},$$

where $\lambda > 0$. Prove K is positive semidefinite.

Extended Reading

Horn, R. A. and Johnson, C. R. (2012). *Matrix analysis*. Cambridge university press.

Lax, P. D. (2014). *Functional analysis*. John Wiley & Sons.

Rigby, J. and Wiegold, J. (1973). Independent axioms for vector spaces. *The Mathematical Gazette*, 57(399):56–62.