1 Review of Linear Algebra (in finite dimensions)

Linear Algebra is a study of linear maps on vector spaces. This chapter will review linear algebra, which is required for this course. We will also discuss Singular Value Decomposition and Principal Component Analysis.

1.1 Vector Space

Definition 1.1 (Vector Space)**.** *A vector space V over $\mathbb{R}$ is a set together with two operations:*

(a) Addition: $V \times V \to V$, denoted as $(v, w) \mapsto v + w$; (N.b. Applying induction to this axiom allows us to add any finite number of vectors, but it does not allow us to add infinitely many vectors, unless we further define a notion of convergence by introducing a norm-induced topological structure on V.)

(b) Scalar multiplication: $\mathbb{R} \times V \to V$, denoted as $(a, v) \mapsto av$;

such that the following conditions hold

(i) identity of addition: $\exists\, 0 \in V$, such that $0 + v = v + 0 = v$, $\forall v \in V$;

(ii) associativity of addition: $u + (v + w) = (u + v) + w$, $\forall u, w, v \in V$;

(iii) additive inverse: $\forall v \in V, \exists (-v) \in V$, such that $v + (-v) = 0$;

(iv) commutativity of addition: $v + w = w + v$, $\forall v, w \in V$;

(v) associativity of multiplication: $(ab)v = a(bv)$, $\forall a, b \in \mathbb{R}, v \in V$;

(vi) identity of scalar multiplication: $1v = v$, $\forall v \in V$;

(vii) distributive laws:

$$\begin{aligned} (a+b)v &= av + bv \\ a(v+w) &= av + aw \end{aligned}$$

$\forall a, b \in \mathbb{R}$ and $v, w \in V$.

REMARK 1.1. *We will mostly consider vector spaces over $\mathbb{R}$ in this course. From now on, when I say "vector space", I will implicitly mean "vector space over $\mathbb{R}$". All matrices will be assumed to be real, unless stated otherwise.*

1.1.1 Basis and Spanning Set

Definition 1.2 (Spanning Set)**.** *Let V be a vector space. A set $\{v_1, v_2, \ldots, v_n\} \subset V$ is called a spanning set of V iff $Span\{v_1, v_2, \ldots, v_n\} = V$. Note that v_i's in this case may be linearly dependent, in which case the linear combination $v = \sum_{i=1}^{n} a_i v_i$ for some $v \in V$ may not be unique.*

Definition 1.3 (Basis)**.** *Let V be a vector space. A set $\{e_1, e_2, \ldots, e_n\} \subset V$ is called a basis of V iff every $v \in V$ can be written* ***uniquely*** *as $v = \sum_{i=1}^{n} a_i e_i$, $a_i \in \mathbb{R}$. Alternatively, $\{e_1, e_2, \ldots, e_n\}$ is a basis of V if $Span\{e_1, e_2, \ldots, e_n\} = V$ and $e_1, e_2, \ldots, e_n$ are linearly independent.*

EXERCISE 1.1. *Show that the two definitions of basis are equivalent.*

We will see later that the columns of a kernel matrix may be linearly dependent and thus form a spanning set, not a basis, of a vector subspace. More generally, we will see that the kernel feature map forms the spanning set of a pre-Hilbert space, the completion of which yields the reproducing kernel Hilbert space (RKHS).

Definition 1.4 (Dimension)**.** *The dimension $\dim(V)$ of a vector space V is the number of elements in its basis.*

REMARK 1.2. *It is crucial to note at this point that the addition operation in the definition of a vector space provides a rule for adding two vectors – and, by induction, a finitely many of them – but NOT infinitely many vectors. Hence, in infinite dimensions, a set $\{e_\omega\}_{\omega \in \Omega}$, where Ω is either countably infinite or uncountable, is a vector space basis iff any vector can be written as a unique finite linear combination of e_ω's. Equivalently, $\{e_\omega\}_{\omega \in \Omega}$ is a vector space basis of V iff it finitely spans V and any finite linear combination $\sum_{i=1}^{n} a_i e_{\omega_i} = 0$ implies $a_i = 0$, for $i = 1, \ldots, n$.*

Definition 1.5 (Hamel Basis)**.** *A vector space basis of an infinite dimensional vector space is called a Hamel basis.*

REMARK 1.3. *Zorn's lemma implies that any infinite dimensional vector space has a Hamel basis, but there is no explicit construction when the basis is not countable, as is the case for an infinite dimensional Hilbert space. So, Hamel basis is not very useful in explicit calculations, and that's why many of you haven't seen it in previous Physics courses. It is, however, important to note the distinction between a Hamel basis and an orthonormal basis (a.k.a. Hilbert space basis).*

REMARK 1.4. *On the interval $[0, L]$, you can expand a periodic square-integrable continuous function as an infinite Fourier series in $\sin(n\pi x/L)$ and $\cos(n\pi x/L)$. Note that the sines and cosines do not form a vector space basis. Confusingly, they are said to form an orthonormal basis.*

REMARK 1.5. *In general, the dimension of a vector space can be finite, infinite and uncountable, or infinite and countable. For example, an infinite dimensional Hilbert space has an uncountable basis, but a separable Hilbert space has a dense subspace with a countably infinite basis. In this chapter, we will consider only finite dimensional vector spaces. We will study infinite dimensional Hilbert spaces in subsequent chapters.*

1.1.2 Function Space

For understanding Hilbert space, it is instructive to view $\mathbb{R}^n$ as a finite dimensional function space.

Definition 1.6 (Function Space). *Let X denote a finite set $\{x_1, \ldots, x_n\}$ of elements. Then, the set $\mathbb{R}^X$ of all functions $f : X \to \mathbb{R}$ is called a function space.*

EXERCISE 1.2. *Show that $\mathbb{R}^X$ is a vector space and that $\mathbb{R}^X \simeq \mathbb{R}^n$. Note that the isomorphism map $f \mapsto (f(x_1), f(x_2), \ldots, f(x_n))$ is just the evaluation of f at all elements.*

Thus, $\mathbb{R}^n$ is a vector space of all functions from n distinct elements or points to $\mathbb{R}$. This definition generalizes to the case when X is not a finite set. For example, in Physics, we typically deal with $X = \mathbb{R}^3$ or $\mathbb{R}^4$. In that case, $\mathbb{R}^X$ is too large, and we often focus on a subspace by imposing further constraints, such as the square integrability or differentiability.

1.1.3 Subspace

A subset W of a vector space V is called a subspace if W itself is a vector space. Because W inherits algebraic properties from V, we only need to check three things to ensure that W is a vector space:

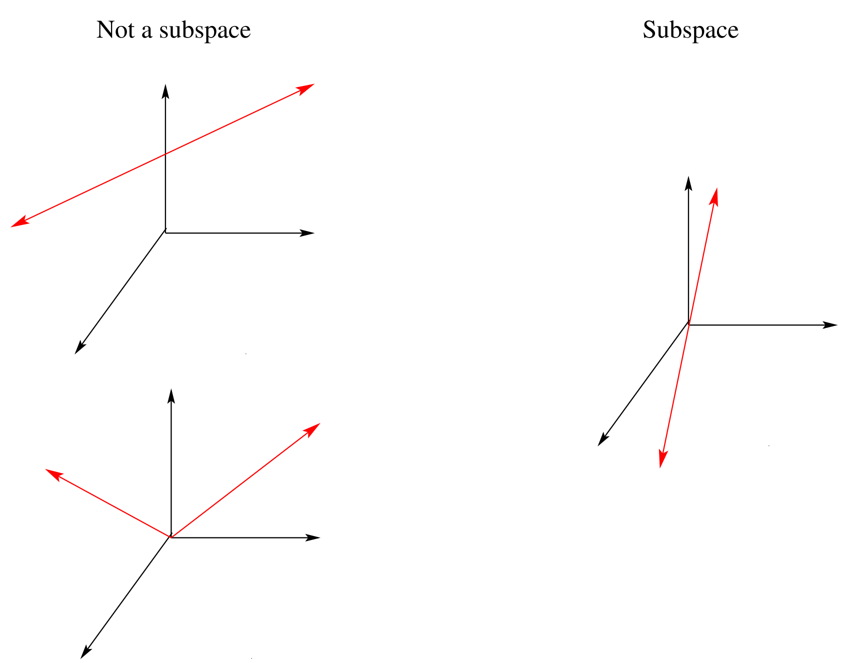
Definition 1.7 (Subspace). *A subset $W \subseteq V$ of a vector space V is called a subspace if the following conditions hold*

(i) identity of addition: $0 \in W$;

(ii) closure under addition: $\forall w_1, w_2 \in W$, we have $w_1 + w_2 \in W$.

(iii) closure under scalar multiplication: $\forall a \in \mathbb{R}, w \in W$, we have $aw \in W$.

Example 1.1. *(Subspace or not?)*



EXERCISE 1.3. *Show that the intersection of two subspaces of a vector space is a subspace.*

1.1.4 Inner Product, Norm, and Metric

To provide a vector space with geometry, we need to impose an extra structure that allows us to measure angles between vectors. This structure is the inner product, which is also known as a dot product in Euclidean geometry:

Definition 1.8 (Inner Product)**.** *Let V be a vector space. A binary map $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ is called an inner product if the following conditions hold*

(i) symmetry: $\langle v, w\rangle = \langle w, v\rangle$, $\forall v, w \in V$;

(ii) linearity: $\langle au + bv, w\rangle = a\langle u, w\rangle + b\langle v, w\rangle$, $\forall a, b \in \mathbb{R}, u, v, w \in V$;

(iii) positive definiteness: $\langle v, v\rangle \geq 0$, $\forall v \in V$, and $\langle v, v\rangle = 0$ iff $v = 0$.

A vector space endowed with an inner product is called an inner product space.

REMARK 1.6. *Note that the geometry of a Hilbert space is directly related to physical quantities and plays a critical role in quantum mechanics. For example, the expectation value of an observable A in a state ψ is just $\langle\psi, A\psi\rangle$, and the transition probability between two states ψ and ϕ is $|\langle\phi,\psi\rangle|^2 = \cos^2\theta$, where θ is the angle between the two states. Thus, all predictions of quantum mechanics can be phrased in terms of the underlying geometry of the Hilbert space describing a physical system.*

We say that two vectors v and w are *orthogonal* or perpendicular if $\langle v, w\rangle = 0$. Inner product also allows us to compute the length or norm of a vector as

Definition 1.9 (Norm Induced by Inner Product)**.** *Let $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ be an inner product defined on a vector space V. Then, $\forall v \in V$, we define the induced norm of v to be $\|v\| = \sqrt{\langle v, v\rangle}$.*

More formally, a vector norm is defined as follows:

Definition 1.10 (Norm)**.** *Let V be a vector space. A function $\|\cdot\| : V \to \mathbb{R}$ is called a norm if the following conditions hold*

(i) homogeneity: $\|cv\| = |c|\|v\|$, $\forall v \in V, c \in \mathbb{R}$;

(ii) non-negativity: $\|v\| \geq 0, \forall v \in V$, and $\|v\| = 0$ iff $v = 0$;

(iii) triangle inequality: $\|v + w\| \leq \|v\| + \|w\|$, $\forall v, w \in V$.

A vector space endowed with a norm is called a normed vector space.

REMARK 1.7. *In statistical learning, we are often interested in measuring the magnitude of vectors and matrices, because we need to learn*

1. *how errors propagate in solving linear equations*
2. *how to best fit a model to noisy data*
3. *how to best approximate matrices with a reduced number of degrees of freedom*

4. how to perform dimensional reduction of high-dimensional data

which can be phrased as minimizing a loss function involving the norm of vectors and matrices.

A norm $\|\cdot\|$ on vector space V allows us to define open balls $B_{v_0}(\epsilon) = \{v \in V | \|v - v_0\| < \epsilon\}$ around any point $v_0 \in V$ and thus define open sets, providing V with a topological structure. Topology is needed to study the question of continuity of maps between vector spaces and convergence of sequences, as described in subsequent sections.

Even though any inner product yields a norm, not every norm arises from an inner product:

EXERCISE 1.4. *Prove that a norm $\|\cdot\|$ defines an inner product on a real vector space via the polarization formula*

$$\langle v, w \rangle := \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2\right)$$

if and only if it satisfies the parallelogram law

$$\|v+w\|^2 + \|v-w\|^2 = 2(\|v\|^2 + \|w\|^2).$$

(Hint: Proving the "only if" part in this exercise is easy; impose bilinearity on $\langle v+w, v+w\rangle$. Proving the "if" part is usually done by proving that $\langle v, qw\rangle = q\langle v, w\rangle$ for any rational number q and then extending to $\mathbb{R}$ via a continuity argument.)

EXERCISE 1.5 (ℓ_p-norm)**.** *Assume that $x \in \mathbb{R}^n$ has components $x = (x_1, x_2, \ldots, x_n)$ in the standard basis. When $n > 1$, show that the ℓ_p-norm*

$$\|x\|_p := \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, & \text{if } 1 \le p < \infty, \\ \max_i\{|x_i|\}, & \text{if } p = \infty \end{cases} \tag{1.1}$$

arises from an inner product only when $p = 2$. (Hint: Take two standard basis elements and show that the parallelogram law holds only for $p = 2$.)

REMARK 1.8. *The fact that $\|\cdot\|_p$ is indeed a norm is proved in Chapter A.1.*

REMARK 1.9. *Note that ℓ_p for $p < 1$ is not a norm, because it violates the convexity condition of a norm, as we shall see later in Example 7.3 in Chapter 7.*

Among all basis, orthonormal bases are usually the most convenient ones.

Definition 1.11 (Orthonormal Basis)**.** *Let $\{e_1, e_2, \ldots, e_n\}$ be a basis of an inner product space $(V, \langle\cdot,\cdot\rangle)$. We say that the basis $\{e_1, e_2, \ldots, e_n\}$ is orthonormal if $\langle e_i, e_j\rangle = 0$, for $i \neq j$, and $\|e_i\| := \sqrt{\langle e_i, e_i\rangle} = 1$, for $i = 1, \ldots, n$. Hence, in an orthonormal basis, the elements are pair-wise orthogonal and normalized to have a unit length.*

Not every basis is orthonormal, but we can always construct an orthonormal basis from a given basis:

Theorem 1.1 (Gram-Schmidt Process). *Let $\{e_1, e_2, \ldots, e_n\}$ be a basis of an inner product space $(V, \langle\cdot,\cdot\rangle)$. Define*

$$\begin{aligned}
\tilde{e}_1 &= \frac{e_1}{\|e_1\|} \\
\tilde{e}_2 &= \frac{e_2 - \langle e_2, \tilde{e}_1\rangle \tilde{e}_1}{\|e_2 - \langle e_2, \tilde{e}_1\rangle \tilde{e}_1\|} \\
\tilde{e}_3 &= \frac{e_3 - \langle e_3, \tilde{e}_1\rangle \tilde{e}_1 - \langle e_3, \tilde{e}_2\rangle \tilde{e}_2}{\|e_3 - \langle e_3, \tilde{e}_1\rangle \tilde{e}_1 - \langle e_3, \tilde{e}_2\rangle \tilde{e}_2\|} \\
&\vdots \\
\tilde{e}_n &= \frac{e_n - \sum_{i=1}^{n-1} \langle e_n, \tilde{e}_i\rangle \tilde{e}_i}{\|e_n - \sum_{i=1}^{n-1} \langle e_n, \tilde{e}_i\rangle \tilde{e}_i\|}.
\end{aligned}$$

Then, $\{\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n\}$ is an orthonormal basis of V.

Proof. Exercise. □

If we want to just measure the difference between vectors, then we can drop the homogeneity condition and require

Definition 1.12 (Metric). *Let V be a vector space. A function $d : V \times V \to \mathbb{R}$ is called a metric if the following conditions hold*

(i) symmetry: $d(v, w) = d(w, v), \forall v, w \in V$

(ii) non-negativity: $d(v, w) \geq 0, \forall v, w \in V$, and $d(v, w) = 0$ iff $v = w$;

(iii) triangle inequality: $d(v, w) \leq d(v, u) + d(w, u)$, $\forall u, v, w \in V$.

A vector space endowed with a metric is called a metric vector space.

EXERCISE 1.6. *Show that every norm induces a metric.*

EXERCISE 1.7. *Even though every norm gives rise to a metric, not every metric arises from a norm. Let d be any metric on a vector space V. Show that $\tilde{d}(v, w) = d(v, w)/(1 + d(v, w))$ defines a new metric that cannot arise from a norm.*

1.2 Linear Map

Throughout this subsection, let V and W be vector spaces.

Definition 1.13 (Linear Map). *A map $T : V \to W$ is called linear if $\forall v, w \in V$ and $a \in \mathbb{R}$,*

(i) $T(v + w) = T(v) + T(w)$, and

(ii) $T(av) = aT(v)$.

From this definition, it follows that

Corollary 1.1. *Let $T : V \to W$ be a linear map. Then, $\forall a, b \in \mathbb{R}$, $v, w \in V$,*

$$T(av + bw) = aT(v) + bT(w).$$

REMARK 1.10. *In English, the definition means that a linear map preserves the algebraic structure of vector space. That is, the image of a sum is the sum of the images; and, the image of a scalar multiple is the scalar multiple of the image.*

When the co-domain $W = \mathbb{R}$, we have a special name:

Definition 1.14 (Functional)**.** *A linear map $f : V \to \mathbb{R}$ is called a functional on V.*

Example 1.2 (Operators in Quantum Mechanics)**.** *In quantum mechanics, operators (a.k.a. observables) are linear maps on infinite dimensional Hilbert spaces. Hilbert spaces are quantum mechanical analogues of the classical phase space. An intuitive dictionary of correspondence between classical and quantum mechanics is:*

$$\left\{\begin{array}{c} \textbf{Classical Mechanics} \\ \textit{Phase space } T^*X \\ \textit{State } (x,p) \in T^*X \\ \textit{Physical observables are functions on } T^*X \\ \textit{Hamiltonian dynamics} \end{array}\right\} \Rightarrow \left\{\begin{array}{c} \textbf{Quantum Mechanics} \\ \textit{Hilbert space } \mathcal{H} \subset \{f : X \to \mathbb{C}\} \\ \textit{Wave function } \psi \in \mathcal{H} \\ \textit{Self-adjoint operators on } \mathcal{H} \\ \textit{Schrödinger time evolution} \end{array}\right\}$$

A linear map is nice, because we only need to specify its action on the basis of V in order to completely determine how it acts on the entire space V. That is,

Theorem 1.2. *Let $T : V \to W$ be a linear map and $\{e_1, e_2, \ldots, e_n\}$ a basis of V. Then, the action of T on V is uniquely determined by its action on the basis $\{e_1, e_2, \ldots, e_n\}$.*

Proof. Since $\{e_1, e_2, \ldots, e_n\}$ is a basis of V, for all $v \in V$, there exists a unique set of numbers $a_i \in \mathbb{R}$, $i = 1, \ldots, n$, such that $v = \sum_{i=1}^n a_i e_i$. But, by linearity of T, we have

$$T(v) = T\left(\sum_{i=1}^n a_i e_i\right) = \sum_{i=1}^n a_i T(e_i).$$

Hence, $\forall v \in V$, $T(v)$ is uniquely determined by its values $T(e_i)$. $\square$

A linear map maps subspaces to subspaces.

Theorem 1.3. *Let $T : V \to W$ be a linear map, and U a subspace of V. Then,*

(a) $T(0) = 0$.

(b) *$T(U)$ is a subspace of W.*

Proof. (a) $T(0) = T(-1 \cdot 0) = -T(0) \Rightarrow T(0) = 0$. (b) Since U is a subspace, $0 \in U$. Since $T(0) = 0$, $0 \in T(U)$. To check the closure under addition, suppose $w_1, w_2 \in T(U)$; then, $\exists v_1, v_2 \in U$ such that $T(v_1) = w_1$ and $T(v_2) = w_2$. Hence, $w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2)$. Since U is a subspace, $v_1 + v_2 \in U$, implying that $w_1 + w_2 \in T(U)$. To check the closure under scalar multiplication, suppose $w = T(v)$ for some $v \in U$; then, $\forall a \in \mathbb{R}$, $aw = aT(v) = T(av)$. Since U is a subspace, $av \in U$, and thus $aw \in T(U)$. $\square$

Definition 1.15. *Let $T : V \to W$ be a linear map, and $\langle \cdot, \cdot \rangle$ an inner product on W. Then, we define*

(a) (Kernel) $\ker(T) = \{v \in V \mid T(v) = 0\}$,

(b) (Image) $\mathrm{Im}(T) = \{T(v) \mid v \in V\}$,

(c) (Cokernel) $\mathrm{coker}(T) = \{w \in W \mid \langle w, T(v) \rangle = 0, \forall v \in V\}$. *(In general,* $\mathrm{coker}(T) = W/Im(T)$*, and the definition given here can be thought of as defining a dual space of this quotient space).*

Thus, $\mathrm{coker}(T)$ is defined here to be the orthogonal complement of $\mathrm{Im}(T)$.

Example 1.3 (Forgetful-z map)**.** *Consider the map $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x, y, z) = (x, y)$. Then,* $\ker(T)$ *is the entire z-axis, $Im(T) = \mathbb{R}^2$, and $coker(T) = \{0\}$.*

Importantly, the kernel, image, and cokernel of a linear map are not just subsets, but they actually form subspaces.

Theorem 1.4. *Let $T : V \to W$ be a linear map. Then,*

(a) $\ker(T)$ *is a subspace of V,*

(b) $\mathrm{Im}(T)$ *is a subspace of W,*

(c) $\mathrm{coker}(T)$ *is a subspace of W.*

Proof. Exercise. $\square$

Example 1.4. *Let $T : V \to W$ be a linear map. Suppose $w = T(v_0)$. Then, the set of all solutions to the equation $T(v) = w$ is* $\{v_0 + v_1 \mid v_1 \in \ker(T)\}$.

Definition 1.16 (Rank)**.** *The rank of a linear map $T : V \to W$ is the dimension of $Im(T)$.*

Theorem 1.5 (Dimension Theorem)**.** *Let $T : V \to W$ be a linear map. Then,*

(a) $\dim(V) = \dim(\ker(T)) + \mathrm{rank}$.

(b) $\dim(W) = \dim(\mathrm{coker}(T)) + \mathrm{rank}$.

1.2.1 Continuous Function

In this subsection, we will briefly review the notion of a continuous function between two normed vector spaces.

Definition 1.17 (Continuous Function). *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces. A function $f : V \to W$ is called continuous at x if $\forall \epsilon > 0$, $\exists\, \delta(\epsilon, x) > 0$, such that $\|y - x\|_V < \delta \Rightarrow \|f(y) - f(x)\|_W < \epsilon$. We say that f is continuous on V if it is continuous at every point of V. If δ depends only on ϵ and not on the point, then f is said to be uniformly continuous.*

For a linear map, we have

Theorem 1.6 (Continuous Linear Map). *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces. A linear map $T : V \to W$ is uniformly continuous on V iff it is continuous at a point in V.*

Proof. If T is continuous on V, then it is, by definition, continuous at any given point. Now suppose that T is continuous at a point $x \in V$. Then, for any $\epsilon > 0$, there exists $\delta(\epsilon, x) > 0$ such that

$$\forall\, y \in V \text{ satisfying } \|y - x\|_V < \delta,\ \|T(y) - T(x)\|_W < \epsilon.$$

But, by linearity of T, this condition implies that

$$\forall\, v \in V \text{ satisfying } \|v\|_V < \delta,\ \|T(v)\|_W < \epsilon.$$

In particular, for any other point $z \in V$, notice that

$$\forall\, \Delta z \in V \text{ satisfying } \|\Delta z\|_V < \delta,\ \|T(z + \Delta z) - T(z)\|_W = \|T(\Delta z)\|_W < \epsilon.$$

Since δ does not depend on z, T is uniformly continuous. $\square$

A strong form of uniform continuity is the Lipschitz continuity:

Definition 1.18 (Lipschitz Continuous Function). *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces. A function $f : V \to W$ is said to be Lipschitz continuous if there exists a constant $C > 0$, such that*

$$\|f(x) - f(y)\|_W \leq C\|x - y\|_V$$

for any $x, y \in V$.

Choosing $\delta(\epsilon) = \epsilon/C$ shows that a Lipschitz continuous function is continuous on V. We will see in Section 1.15 that a linear map on a finite dimensional vector space is always Lipschitz continuous.

1.3 Matrix Representation

Let $T : V \to W$ be a linear map. Recall from Theorem 1.2 that a linear map is uniquely determined by how it acts on the basis vectors. Let $E = \{e_1, \ldots, e_n\}$ and $\tilde{E} = \{\tilde{e}_1, \ldots, \tilde{e}_m\}$

be bases of V and W, respectively. Since for all $i = 1, \dots n$, $T(e_i) \in W$, and since $\tilde{E}$ is a basis of W, there exist real numbers T_{ji}, $j = 1, \dots, m$, such that we can expand $T(e_i)$ as

$$T(e_i) = \sum_{j=1}^{m} T_{ji} \tilde{e}_j.$$

Since E is a basis of V, for any $v \in V$, we can write $v = \sum_{i=1}^{n} a_i e_i$. Then, by linearity, we have

$$T(v) = T\left(\sum_{i=1}^{n} a_i e_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} T_{ji} a_i \tilde{e}_j = \sum_{j=1}^{m}\left(\sum_{i=1}^{n} T_{ji} a_i\right)\tilde{e}_j. \tag{1.2}$$

In the basis E, the matrix representation of a vector $v = \sum_{i=1}^{n} a_i e_i \in V$ is written as a column

$$v = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Similarly, in the basis $\tilde{E}$, the matrix representation of a vector $w = \sum_{i=1}^{m} b_i \tilde{e}_i \in W$ is written as a column

$$w = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

In this notation, we can rewrite (1.2) as

$$T\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} T_{1i} a_i \\ \sum_{i=1}^{n} T_{2i} a_i \\ \vdots \\ \sum_{i=1}^{n} T_{mi} a_i \end{pmatrix} := \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1n} \\ T_{21} & T_{22} & \cdots & T_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ T_{m1} & T_{m2} & \cdots & T_{mn} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Definition 1.19. *We will sometimes use the notation $M \in \mathbb{R}^{m \times n}$ to indicate that M is a real $m \times n$ matrix.*

Definition 1.20 (Dual Vector Space and Dual Basis)**.** *Let V be a vector space. A dual vector space V^* of V is the vector space of all linear functionals $f : V \to \mathbb{R}$. If $\{e_1, \dots, e_n\}$ is a basis of V, then the dual basis $\{e_1^*, \dots, e_n^*\}$ of V^* is defined by $e_i^*(e_j) = \delta_{ij}$.*

Definition 1.21 (Transpose Map)**.** *Let $T : V \to W$ be a linear map. The transpose map $T^t : W^* \to V^*$ of T is defined by $T^t(w^*) = w^* \circ T$.*

EXERCISE 1.8. *If T has a matrix representation $(T)_{ij}$ in some fixed bases of V and W,* *then T^t has a matrix representation $(T^t)_{ij} = (T)_{ji}$ in the corresponding dual bases. (Note that this is just a matrix transpose.)*

If $\{e_1, \dots, e_n\}$ is an orthonormal basis of V and $\{\tilde{e}_1, \dots, \tilde{e}_m\}$ is an orthonormal basis of W with respect to their inner products $\langle \cdot, \cdot \rangle_V$ and $\langle \cdot, \cdot \rangle_W$, respectively, then the maps $e_i \mapsto e_i^* := \langle e_i, \cdot \rangle_V$ and $\tilde{e}_i \mapsto \tilde{e}_i^* := \langle \tilde{e}_i, \cdot \rangle_W$ define the dual bases $\{e_1^*, \dots, e_n^*\}$ and $\{\tilde{e}_1^*, \dots, \tilde{e}_m^*\}$. By linearity of the inner product, this representation extends to the entire vector space:

$$\sum_i \alpha_i e_i \overset{\text{1-to-1}}{\longleftrightarrow} \sum_i \alpha_i e_i^* \quad \text{and} \quad \sum_i \beta_i \tilde{e}_i \overset{\text{1-to-1}}{\longleftrightarrow} \sum_i \beta_i \tilde{e}_i^*$$

Thus, we will prove later that these maps induce the isomorphisms $V \simeq V^*$ and $W \simeq W^*$; we will see that these isomorphisms are examples of the Riesz Representation Theorem for finite dimensional vector spaces, and v_i is the unique representer of the dual element v_i^*. Thus, in an elementary linear algebra course, one typically learns that $T^t : W \to V$, defined by the matrix transpose of T. We will also use this idea in this course.

Given this information, suppose $T : V \to W$ is a linear map from inner product space $(V, \langle, \rangle_V)$ to $(W, \langle, \rangle_W)$. Then, for any $v \in V$ and $w \in W$, we can find unique dual elements $v^* \in V^*$ and $w^* \in W^*$ represented by v and w as

$$v^* = \langle v, \cdot \rangle_V \quad \text{and} \quad w^* = \langle w, \cdot \rangle_W.$$

Thus, for all $v \in V$ and $w \in W$

$$\langle w, Tv \rangle_W = w^*(Tv) = (T^t w^*)v = \langle T^t w, v \rangle_V,$$

where $T^t w$ is, by definition, the unique representer of $T^t w^* \in V^*$ in V.

EXERCISE 1.9. *Show that in matrix form, $T^t w$ is indeed the transpose of the matrix of T multiplied by the column vector w.*

Theorem 1.7. *Let $T : V \to W$ be a linear map, and T^t its transpose. Then,* $\operatorname{coker}(T) = \ker(T^t)$ *and* $\ker(T) = \operatorname{coker}(T^t)$.

Proof. We will first show that $coker(T) \subseteq \ker(T^t)$ and then that $\ker(T^t) \subseteq coker(T)$, which together will imply that $\ker(T^t) = coker(T)$. Suppose $w \in coker(T)$. Then, by definition, $\langle w, T(v) \rangle = 0, \forall v \in V$. But, since $\langle T^t(w), v \rangle = \langle w, T(v) \rangle$, we have $\langle T^t(w), v \rangle = 0, \forall v \in V \Rightarrow T^t(w) = 0 \Rightarrow coker(T) \subseteq \ker(T^t)$. Conversely, suppose $w \in \ker(T^t)$. Then, $T^t(w) = 0 \Rightarrow \langle T^t(w), v \rangle = \langle w, T(v) \rangle = 0, \forall v \in V \Rightarrow w \in coker(T)$. Thus, $\ker(T^t) \subseteq coker(T)$, and $\ker(T^t) = coker(T)$. Since $(T^t)^t = T$, we have $\ker(T) = \ker((T^t)^t) = coker(T^t)$. □

EXERCISE 1.10 (Rank Revisited)**.** *Show that the rank of a matrix M is the dimension of the column span of M. Show that the rank is also equal to the dimension of the row span of M.*

1.4 Determinants

To describe eigenvalues as solutions to a characteristic polynomial in the next section, we will need to know how to compute the determinant of a matrix. To define the determinant, we first need to define the antisymmetric tensor. Let us first discuss how to permute objects. Let $X = (1, 2, \ldots, n)$ be an ordered set. A permutation of X is a rearrangement of the elements in X. For example, if $X = (1, 2, 3)$, the possible permutations of X are $(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2)$, and $(3, 2, 1)$. In general, if X has n elements, then there will be $n!$ possible permutations of X.

Definition 1.22 (Transposition). *A transposition of a set is a permutation that swaps only two elements in the set.*

Example 1.5. *The transpositions of* $X = (1, 2, 3)$ *are* $(1, 3, 2), (3, 2, 1), (2, 1, 3)$.

Definition 1.23 (Parity of Permutation). *A permutation of* X *is called an even permutation if it is a composition of an even number of transpositions. A permutation of* X *is called an odd permutation if it is a composition of an odd number of transpositions.*

Example 1.6. *To check the parity of* $(3, 1, 2)$, *we note that* $(3, 1, 2)$ *is obtained from* $(1, 2, 3)$ *by the following transpositions:* $(1, 2, 3) \to (2, 1, 3) \to (3, 1, 2)$, *where we swaped the first and the second elements in the first transposition, and we swapped the first and the third elements in the second transposition. Thus,* $(3, 1, 2)$ *is an even permutation of* $(1, 2, 3)$.

Example 1.7. $(3, 1, 4, 2)$ *is an odd permutation of* $(1, 2, 3, 4)$, *because we can obtain it from:* $(1, 2, 3, 4) \to (1, 4, 3, 2) \to (1, 3, 4, 2) \to (3, 1, 4, 2)$.

Definition 1.24 (Antisymmetric Tensor).

$$\epsilon^{i_1, i_2, \ldots, i_n} = \begin{cases} 1, & \textit{if } (i_1, i_2, \ldots, i_n) \textit{ is an even permutation of } (1, 2, \ldots, n) \\ -1, & \textit{if } (i_1, i_2, \ldots, i_n) \textit{ is an odd permutation of } (1, 2, \ldots, n) \\ 0, & \textit{otherwise.} \end{cases}$$

REMARK 1.11. *The antisymmetric tensor is 0 if some of the indices are repeated. For example,* $\epsilon^{1,1,2,3} = \epsilon^{1,2,3,3} = \epsilon^{1,2,2,3} = 0$

We can now define:

Definition 1.25 (Determinant). *The determinant* $\det(M)$ *of an* $n \times n$ *matrix* $M = (M_{ij})$ *is*

$$\det(M) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} \epsilon^{i_1, i_2, \ldots, i_n} M_{1i_1} M_{2i_2} \cdots M_{ni_n}.$$

We also denote the determinant as $|M|$.

REMARK 1.12. *In English, it means, "choose one element from each row, such that the columns of the elements do not overlap, multiply them together, give a sign based on the parity of the permutation* $(i_1, \ldots, i_n)$, *and sum over all possible such products"*

Example 1.8. *Let* M *be a* 2×2 *matrix Then,*

$$\det(M) = \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \epsilon^{i_1, i_2} M_{1i_1} M_{2i_2} = \epsilon^{12} M_{11} M_{22} + \epsilon^{21} M_{12} M_{21} = M_{11} M_{22} - M_{12} M_{21}.$$

That is,

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

1.5 Eigenvalues and Eigenvectors

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. One very useful way of understanding how T acts on $\mathbb{R}^n$ is to figure out whether the action of T on certain vectors is to just scale them, i.e.

$$Tv = \lambda v, \text{ where } v \in \mathbb{R}^n, \lambda \in \mathbb{R}.$$

It turns out this condition is too strict for most matrices, so we need to relax the condition to

$$Tv = \lambda v, \text{ where } v \in \mathbb{C}^n, \lambda \in \mathbb{C}. \tag{1.3}$$

If this relation holds for a non-zero vector v, then we call λ an eigenvalue of T, and v its corresponding eigenvector.

EXERCISE 1.11. *Suppose λ is an eigevenvalue of $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{C}^{2\times 2}$.*

- *Show that $\begin{pmatrix} b \\ \lambda - a \end{pmatrix}$ and $\begin{pmatrix} \lambda - d \\ c \end{pmatrix}$, if non-zero, are eigenvectors of A corresponding to the eigenvalue λ.*
- *When both $\begin{pmatrix} b \\ \lambda - a \end{pmatrix}$ and $\begin{pmatrix} \lambda - d \\ c \end{pmatrix}$ are zero vectors, what are the eigenvectors? In this case, is λ degenerate or non-degenerate?*

The best case scenario would be when you are able to find n real eigenvectors, $v_1, \ldots, v_n$, that are linearly independent, i.e. they form a basis of $\mathbb{R}^n$. In that case, you can express any vector $v \in \mathbb{R}^n$ as a linear combination $v = \sum_{i=1}^n \alpha_i v_i, \alpha_i \in \mathbb{R}$, and the action of T on v can be simply expressed as

$$T(v) = T\left(\sum_{i=1}^n \alpha_i v_i\right) = \sum_{i=1}^n \alpha_i T(v_i) = \sum_{i=1}^n \alpha_i \lambda_i \, v_i,$$

i.e. T scales the eigenvectors by their corresponding real eigenvalues. We will see in the next subsection that symmetric matrices satisfy these desired properties and, more generally, that normal matrices satisfy equivalent properties in complex vector spaces.

Unfortunately, not all $n \times n$ matrices will have n eigenvectors that are linearly independent. Some eigenvectors and eigenvalues may even be complex. But, you will always be able to find n eigenvalues, which may be complex and have multiplicity; and, for each distinct eigenvalue, you can always find at least one corresponding eigen-direction in $\mathbb{C}^n$. Let us rewrite (1.3) as

$$Tv - \lambda v = (T - \lambda I)v = 0,$$

where I is an $n \times n$ matrix. If $T - \lambda I$ were invertible, then multiplying the above equation by $(T - \lambda I)^{-1}$ will yield $v = 0$. Hence, in order for a non-zero solution v to exist, $T - \lambda I$ cannot have an inverse. But, $T - \lambda I$ does not have an inverse if and only if

$$\det(T - \lambda I) = 0. \tag{1.4}$$

From the definition of the determinant, we see that (1.4) is an n-th degree polynomial in λ, called the characteristic polynomial. Thus, the Fundamental Theorem of Algebra guarantees that there exist n solutions to (1.4), where some of the solutions may be complex or repeated.

EXERCISE 1.12. *Show that if an eigenvalue of a real $n \times n$ matrix is real, then a corresponding eigenvector can be also chosen to have only real components.*

EXERCISE 1.13. *Let T be an $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. Factorize the characteristic polynomial $\det(T - \lambda I)$ to show that $\det(T) = \prod_{i=1}^{n} \lambda_i$ and $\mathrm{tr}(T) = \sum_{i=1}^{n} \lambda_i$.*

Example 1.9. *Let T be 2×2 matrix given by*

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then, we want to solve for λ that satisfies

$$det(T - \lambda I) = \begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = (a - \lambda)(d - \lambda) - cb = 0$$

Using the quadratic formula, we see that the two eigenvalues are

$$\lambda_{\pm} = \frac{a + d \pm \sqrt{(a-d)^2 + 4bc}}{2}.$$

The corresponding eigenvectors are

$$v_{\pm} = \begin{pmatrix} \frac{(a-d) \pm \sqrt{(a-d)^2 + 4bc}}{2c} \\ 1 \end{pmatrix}.$$

REMARK 1.13. *There also exist algebraic formulae for finding the roots of degree 3 and degree 4 polynomials, but not for degree 5 and higher; the latter fact is a famous result of Galois theory. Thus, in practice, most eigenvalue problems must be solved via iterative numerical algorithms, as we will describe in subsequent sections.*

As already mentioned, a very useful way of understanding a matrix, if possible, would be to find an appropriate basis on which the matrix acts by simply scaling each basis element. That is, in that particular basis, if it exists, the matrix representation would be diagonal. The following theorem tells us when we can find such a basis:

Theorem 1.8. *A matrix $M \in \mathbb{C}^{n\times n}$ is diagonalizable, i.e. there exists an invertible $U \in \mathbb{C}^{n\times n}$ such that $U^{-1}MU = D$ for some diagonal matrix D, iff M has n linearly independent eigenvectors.*

Proof. ($\Rightarrow$) Rewriting $M = UDU^{-1}$, we see that the i-th column of U is an eigenvector of M with eigenvalue D_{ii}. Because U is invertible, the eigenvectors are linearly independent.

($\Leftarrow$) Suppose v_i, $i = 1, \dots, n$, are linearly independent eigenvectors of M with corresponding eigenvalues λ_i. Taking $U = (v_1 \cdots v_n)$, we see that

$$MU = U\Lambda, \quad \text{where } \Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n).$$

Hence,

$$U^{-1}MU = U^{-1}U\Lambda = \Lambda.$$

□

REMARK 1.14. *In physics and other branches of science, we are often interested finding a transformation matrix U that is either orthogonal or unitary, which preserves the ℓ_2-norm of vectors. The orthogonality or unitarity of U implies that the eigenvectors can be chosen to form an orthonormal basis. We will precisely characterize the matrices that possess this critical property in Section 1.8.*

So, when are eigenvectors linearly independent? The following theorem provides a sufficient condition:

Theorem 1.9. *Eigenvectors corresponding to distinct eigenvalues are linearly independent over $\mathbb{C}$.*

Proof. Exercise. □

An immediate consequence of this theorem is:

Corollary 1.2. *A complex matrix $n\times n$ with n distinct eigenvalues has n linearly independent eigenvectors and can be diagonalized.*

The contrapositive of the above statement is:

Corollary 1.3. *A nondiagonalizable matrix must have at least one degenerate eigenvalue.*

Example 1.10. *An example of an $n \times n$ matrix that does not have enough eigenvectors is*

$$M = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

which has 1 along the diagonal and just above the diagonal, and 0 elsewhere. The eigenvalue of M is 1 with algebraic multiplicity n, and its only eigenvector is $(1\ 0\ 0 \cdots\ 0)^t$. A more general example is

$$M = \lambda I + N,\ \ \lambda \neq 0,$$

where I is the $n \times n$ identity matrix, and N is any $n \times n$ nilponent matrix, which by definition means that its eigenvalue is 0 with multiplicity n. The eigenvalue of M is λ with multiplicity n, and the dimension of the eigenspace is equal to the dimension of $\ker(N)$. Thus, if N is not a zero matrix, then M must have fewer than n linearly independent eigenvectors.

REMARK 1.15. *But, the converse is not necessarily true; that is, not all matrices with degenerate eigenvalues are nondiagonalizable. For example, in quantum mechanics, you learned that Hermitian matrices are always diagonalizable, even if they have degenerate eigenvalues. In Section 1.8, we will see that symmetric and Hermitian matrices are examples of a more general class of matrices called normal matrices that possess a special symmetry property.*

Let us end this section with the following definition of spectral radius, which provides a lower bound of any matrix norm measuring the magnitude of matrices, as discussed in Section 1.15:

Definition 1.26 (Spectral Radius)**.** *Let T be an $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$. The spectral radius $\rho(T)$ of T is the largest absolute value of its eigenvalues; i.e.*

$$\rho(T) = \max_i \{|\lambda_i|\}.$$

REMARK 1.16. *The spectral radius itself cannot be a norm, because the only eigenvalue of any non-zero nilpotent matrix is 0, so its spectral radius is 0, but the matrix itself is not zero.*

1.6 Bounds on Eigenvalues

Galois proved that there exist no algebraic solutions to general polynomial equations of degree ≥ 5. Thus, finding the eigenvalues of a large matrix may not be so easy. The following simple yet amazing theorem, however, provides useful hints as to where the eigenvalues may reside on the complex plane. We will later apply this theorem to prove that the graph Laplacian is positive semi-definite.

Theorem 1.10 (Gershgorin Circle Theorem)**.** *Let $M = (m_{ij})$ be an $n \times n$ complex matrix. Let $R_i = \sum_{j \neq i} |m_{ij}|$ be the sum of absolute values of all off-diagonal elements in the i-th row, and let $D_i = \{z \in \mathbb{C} | \, |z - m_{ii}| \leq R_i\}$ be the disc of radius R_i centered at the diagonal element m_{ii} in the complex plane. Then, all eigenvalues of M lie in the union $\cup_{i=1}^n D_i$.*

Proof. Let λ be an eigenvalue of M with a eigenvector $v = (v_1, \ldots, v_n)^t$. Let $i = \arg\max_j |v_j|$. Then, $u = v/v_i$ is still an eigenvector of M with eigenvalue λ, and all of its components have magnitude less than or equal to 1; in particular, the i-th component of u is 1. The i-th row of the equation $Mu = \lambda u$ then yields

$$\sum_{j \neq i} m_{ij} u_j + m_{ii} = \lambda \,.$$

Hence, $|\lambda - m_{ii}| = |\sum_{j\neq i} m_{ij}u_j| \leq \sum_{j\neq i} |m_{ij}u_j|$, by the triangle inequality. But, $|u_j| \leq 1$, so we have $|\lambda - m_{ii}| \leq \sum_{j\neq i} |m_{ij}| = R_i$, and λ is thus contained in D_i. □

An application of this theorem provides an upper bound on the spectral radius:

Corollary 1.4. *For any complex $n \times n$ matrix $M = (m_{ij})$,*

$$\rho(M) \leq \max_i \left\{ \sum_{j=1}^{n} |m_{ij}| \right\} \equiv \|M\|_\infty.$$

Proof. Exercise. □

Gershgorin also proved the following variation in his 1932 paper:

Theorem 1.11. *Let $M \in \mathbb{C}^n$ be a complex matrix, and let D_i denote the i-th Gershgorin disk, as defined above. Let $\Gamma = \cup_{i=1}^n D_i$. If Γ_S is a maximally connected component of Γ consisting of $|S|$ Gershgorin disks indexed by $S \subset \{1, \ldots, n\}$, then Γ_S contains exactly $|S|$ eigenvalues of M. In particular, an isolated Gershgorin disk contains precisely one eigenvalue.*

Sketch of Proof. Let $M(t)$, $0 \leq t \leq 1$, be a one-parameter family of matrices where we scale the off-diagonal elements of M by t. Then, the radius of each Gershgorin disk is scaled by t. Furthermore, $M(1) = M$ and $M(0) = \text{diag}(m_{11}, \ldots, m_{nn})$. The eigenvalues of $M(0)$ are just the diagonal entries. As t increases towards 1, these eigenvalues may move continuously in the complex plane, but they must still reside within some Gershgorin disk and cannot jump between unconnected disks. Thus, the eigenvalues starting at $m_{i,i}$, $i \in S$, must stay within the connected component Γ_S as $t \to 1$.

□

1.7 Symmetric Matrices

Symmetric matrices are the nicest matrices with the right properties to allow us to perform calculations and factorizations.

Definition 1.27 (Symmetric Matrix)**.** *Let M be an $n \times n$ matrix. M is said to be symmetric if $M^t = M$.*

REMARK 1.17. *Clearly, a matrix has to be a square matrix in order for it to be symmetric. However, given any matrix M, we will later construct two symmetric matrices M^tM and MM^t; the spectral analysis of these two matrices will allow us to factorize M in terms of their eigenvectors and square roots of their eigenvalues. This factorization is called Singular Value Decomposition (SVD). Before we study SVD, we first need to understand the basic properties of symmetric matrices.*

Theorem 1.12. *Let M be a real $n \times n$ symmetric matrix. Then,*

1. *All eigenvalues of M are real.*

2. *Eigenvectors of M corresponding to distinct eigenvalues are mutually orthogonal.*

3. M has n orthonormal eigenvectors that form a basis of $\mathbb{R}^n$.

Proof. 1. Suppose λ is an eigenvalue with eigenvector v. Then, $v^{*t}Mv = \lambda\|v\|^2$, where v^* is complex conjugate of v. Taking complex conjugate and then transpose of this equation, we get $v^{*t}M^{*t}v = \lambda^*\|v\|^2$. But, since M is real and symmetric, $M^{*t} = M$. We thus have $\lambda\|v\|^2 = \lambda^*\|v\|^2$. Since an eigenvector has a non-zero norm, we thus see that $\lambda = \lambda^*$, i.e. λ is real.

2. Let $\lambda_1 \neq \lambda_2$ be eigenvalues with eigenvectors v_1 and v_2, respectively. Then, $\langle v_1, Mv_2\rangle = \lambda_2\langle v_1, v_2\rangle$. But, $\langle v_1, Mv_2\rangle = \langle M^t v_1, v_2\rangle = \langle Mv_1, v_2\rangle = \lambda_1\langle v_1, v_2\rangle$. Combining the two equations, we get $(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle$. Since $\lambda_1 \neq \lambda_2$, we must have $\langle v_1, v_2\rangle = 0$.

3. If all eigenvalues are distinct, then this statement follows from the second property. Let v be an eigenvector of M with a degenerate eigenvalue λ. Then, M preserves the orthogonal complement of v in $\mathbb{R}^n$. That is, if $w \in v^\perp$, then $Mw \in v^\perp$. Hence, M restricted to $v^\perp$ must have an eigenvector in $v^\perp$ with an eigenvalue λ. Repeating this process yields an orthogonal decomposition of $\mathbb{R}^n$ into eigenvector subspaces of M. □

Theorem 1.13 (Diagonalization of a Symmetric Matrix)**.** *Let S be a symmetric matrix with n eigenvalues $\lambda_1, \ldots, \lambda_n$ and n orthonormal eigenvectors $v_1, v_2, \ldots, v_n$. Then, we can express S as*

$$S = TDT^t = \sum_{i=1}^{n} \lambda_i v_i v_i^t \tag{1.5}$$

where $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix with eigenvalues along the diagonal and 0 in off-diagonal entries, and

$$T = (v_1, v_2, \ldots, v_n)$$

has the corresponding orthonormal eigenvectors as columns. (T *is thus an element of the orthogonal group* $O(n) \equiv \{M \in \mathbb{R}^{n\times n} \,|\, MM^t = M^tM = I\}$.)

Proof. The left and right hand sides act the same way on the basis $\{v_1, v_2, \ldots, v_n\}$. Thus, they are identical as linear operators on the entire vectors space. □

1.8 Spectral Theorem for Normal Matrices

Real symmetric and complex Hermitian matrices are special examples of a more general class of complex matrices called normal matrices that are unitarily diagonalizable. We will discuss these matrices here as maps on $\mathbb{C}^n$ endowed with the standard Euclidean inner product and norm:

$$\forall v, w \in \mathbb{C}^n, \langle v, w\rangle = v^{*t}w \quad \text{and} \quad \|v\|_2^2 = \langle v, v\rangle = v^{*t}v.$$

Definition 1.28. *A complex square matrix M is called normal if it commutes with its adjoint $M^\dagger$ (complex conjugate of M^t), i.e. $MM^\dagger = M^\dagger M$. For a real square matrix M, this condition becomes $MM^t = M^tM$.*

This seemingly simply property has many profound consequences and utility. First, let us prove that if M is normal, then Mv and $M^\dagger v$ have the same ℓ_2-norm:

Theorem 1.14. *Let $M \in \mathbb{C}^{n\times n}$ be a normal matrix. Then,*

$$\forall v \in \mathbb{C}^n, \ \ \|Mv\|_2 = \|M^\dagger v\|_2.$$

Proof. By the definition of ℓ_2-norm, we have $\forall v \in \mathbb{C}^n$,

$$\|Mv\|_2^2 = \langle Mv, Mv\rangle = \langle M^\dagger Mv, v\rangle = \langle MM^\dagger v, v\rangle = \langle M^\dagger v, M^\dagger v\rangle = \|M^\dagger v\|_2^2.$$

□

An immediate consequence is:

Corollary 1.5. *If $M \in \mathbb{C}^{n\times n}$ is normal, then*

$$\ker(M) = \ker(M^\dagger).$$

Thus, the zero eigenvectors of M are shared with $M^\dagger$, and vice versa.

Proof. Exercise. □

In fact, all other eigenvectors are also shared:

Corollary 1.6. *If $M \in \mathbb{C}^{n\times n}$ is normal and $v \in \mathbb{C}^n$ is an eigenvector of M with eigenvalue $\lambda \in \mathbb{C}$, then $v \in \mathbb{C}^n$ is also an eigenvector of $M^\dagger$ with eigenvalue λ^**

Proof. Let I denote the $n \times n$ identity matrix. First, note that if M is normal, then for any $\lambda \in \mathbb{C}$, $T_\lambda \equiv M - \lambda I$ is also normal, as

$$\begin{aligned} T_\lambda^\dagger T_\lambda &= (M-\lambda I)^\dagger (M-\lambda I) = (M^\dagger - \lambda^* I)(M-\lambda I) = M^\dagger M - \lambda^* M - \lambda M^\dagger - |\lambda|^2 I \\ &= MM^\dagger - \lambda^* M - \lambda M^\dagger - |\lambda|^2 I = (M-\lambda I)(M-\lambda I)^\dagger = T_\lambda T_\lambda^\dagger. \end{aligned}$$

Hence, Theorem 1.14 implies that for any eigenvector v of M with eigenvalue λ, we have

$$0 = \|T_\lambda v\|_2 = \|T_\lambda^\dagger v\|_2,$$

where we have used the fact that $T_\lambda v = 0$. Since the only vector with zero norm is the zero vector, we thus have $T_\lambda^\dagger v = (M^\dagger - \lambda^* I)v = 0$. □

These results inform the following critical property that a normal matrix preserves a subspace that is the orthogonal complement of an eigensubspace; this is the key property that guarantees the diagonalizability of normal matrices:

Theorem 1.15. *Let $M \in \mathbb{C}^{n\times n}$ be a normal matrix. If $v \in \mathbb{C}^n$ is an eigenvector of M, then $\forall w \in \mathbb{C}^n$ orthogonal to v, i.e. $\langle w, v\rangle = 0$, we have*

$$\langle Mw, v\rangle = 0 \ \ \textit{and} \ \ \langle M^\dagger w, v\rangle = 0.$$

Proof. Using the definition of adjoint, we have

$$\langle Mw, v\rangle = \langle w, M^\dagger v\rangle.$$

But, Corollary 1.6 implies that $M^\dagger v = \lambda^* v$, where λ is the eigenvalue corresponding to v. We thus have

$$\langle Mw, v\rangle = \lambda^* \langle w, v\rangle = 0.$$

Similarly,

$$\langle M^\dagger w, v\rangle = \langle w, Mv\rangle = \lambda\langle w, v\rangle = 0.$$

□

Theorem 1.16 (Spectral Theorem for Normal Matrices)**.** *A complex square matrix $M \in \mathbb{C}^{n\times n}$ is normal iff it is unitarily equivalent to a diagonal matrix, i.e. $M = U\Lambda U^\dagger$ for some unitary matrix U and a diagonal matrix Λ. (The diagonal entries of Λ are the eigenvalues of M and the columns of U are the corresponding eigenvectors.)*

Proof. ($\Leftarrow$) If $M = U\Lambda U^\dagger$, then $M^\dagger = U\Lambda^* U^\dagger$. Hence,

$$MM^\dagger = U\Lambda U^\dagger\, U\Lambda^* U^\dagger = U\Lambda\Lambda^* U^\dagger = U\Lambda^*\Lambda U^\dagger = U\Lambda^* U^\dagger\, U\Lambda U^\dagger = M^\dagger M,$$

where we have used the unitarity condition $U^\dagger U = I$ and the fact that diagonal matrices commute.
($\Rightarrow$) We will prove this part of the theorem using mathematical induction. The case of $n = 1$ is trivially satisfied, as we can just take $U = I$ and $\Lambda = (M_{11})$. Now, assume that any normal matrix $M_n \in \mathbb{C}^{n\times n}$ can be unitarily diagonalized, and we will prove the condition for dimension $n + 1$.

We proceed by recalling that a square matrix always has at least one non-zero eigenvector over the field of complex numbers; in fact, characteristic polynomials arise precisely as a condition for finding such a non-zero eigenvector. Thus, given any normal matrix $M \in \mathbb{C}^{(n+1)\times(n+1)}$, let us choose one eigenvector $v \in \mathbb{C}^{n+1} \setminus \{0\}$ of M, normalized so that $\|v\|_2 \equiv \sqrt{v^{*t}v} = 1$, and choose n orthonomal vectors $e_1, \ldots, e_n$ in the orthogonal complement of $\mathrm{Span}_{\mathbb{C}}\{v\}$. By construction, the set $\{v, e_1, \ldots, e_n\}$ is an orthonomal basis of $\mathbb{C}^{n+1}$, and the matrix

$$U = (v\, e_1\, \cdots\, e_n)$$

consisting of $v, e_1, \ldots, e_n$ along the columns is unitary. In this basis, matrix M has the form

$$U^\dagger M U = \begin{pmatrix} \lambda & w^t \\ z & M_n \end{pmatrix}$$

where $w, z \in \mathbb{C}^n$, λ is the eigenvalue corresponding to v, and $M_n \in \mathbb{C}^{n\times n}$. We claim that $z = 0$, $w = 0$, and M_n is normal. First, note that the i-th entry of z is given by $z_i = \lambda\, e_i^{*t} v$, which is 0 since v is orthogonal to e_i. To see that $w = 0$, note that for any $x \in \mathbb{C}^n$,

$$U^\dagger M U \begin{pmatrix} 0 \\ x \end{pmatrix} = \begin{pmatrix} w^t x \\ M_n x \end{pmatrix}.$$

But, Theorem 1.15 ensures that

$$U \begin{pmatrix} 0 \\ x \end{pmatrix},$$

which is a linear combination of $e_1, \ldots, e_n$, must remain orthogonal to v upon the action of M and, thus, that $w^t x = 0$. Since x was an arbitrary vector in $\mathbb{C}^n$, choosing $x = w^*$ shows that $\|w\|_2^2 = 0$, which implies that $w = 0$. To see that M_n is a normal matrix, convince yourself that any unitary transformation of a normal matrix is also normal (Exercise); as a result,

$$U^\dagger M U = \begin{pmatrix} \lambda & 0 \\ 0 & M_n \end{pmatrix}$$

is a normal matrix. That is, we must have

$$\begin{pmatrix} \lambda & 0 \\ 0 & M_n \end{pmatrix} \begin{pmatrix} \lambda^* & 0 \\ 0 & M_n^\dagger \end{pmatrix} = \begin{pmatrix} \lambda^* & 0 \\ 0 & M_n^\dagger \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & M_n \end{pmatrix},$$

from which it follows that $M_n M_n^\dagger = M_n^\dagger M_n$. The induction hypothesis now implies that there exists a unitary matrix $Q_n \in \mathbb{C}^{n\times n}$ and a diagonal matrix $\Lambda_n \in \mathbb{C}^{n\times n}$ such that $M_n = Q_n \Lambda_n Q_n^\dagger$. Thus,

$$U^\dagger M U = \begin{pmatrix} \lambda & 0 \\ 0 & Q_n \Lambda_n Q_n^\dagger \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & Q_n \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & \Lambda_n \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q_n^\dagger \end{pmatrix} \equiv Q \Lambda Q^\dagger,$$

where the unitary matrix Q and the diagonal matrix Λ are defined as

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & Q_n \end{pmatrix} \quad \text{and} \quad \Lambda = \begin{pmatrix} \lambda & 0 \\ 0 & \Lambda_n \end{pmatrix}.$$

Finally, we can express M as

$$M = (UQ)\, \Lambda\, (UQ)^\dagger,$$

and note that UQ is unitary (Exercise). □

REMARK 1.18. *In textbooks, you will often see an alternative proof of this spectral theorem by first invoking the Schur decomposition theorem, which states that any matrix is unitarily equivalent to an upper triangular matrix, and then proving that a triangular matrix commutes with its Hermitian conjugate iff it is diagonal.*

Example 1.11. *Real matrices that are orthogonal, symmetric, or antisymmetric are normal. Thus, all these matrices can be unitarily diagonalized.*

Example 1.12. *Complex matrices that are unitary, Hermitian, or antihermitian are normal. Thus, all these matrices can be unitarily diagonalized. These fundamental results form the foundation of quantum mechanics and quantum field theory.*