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:

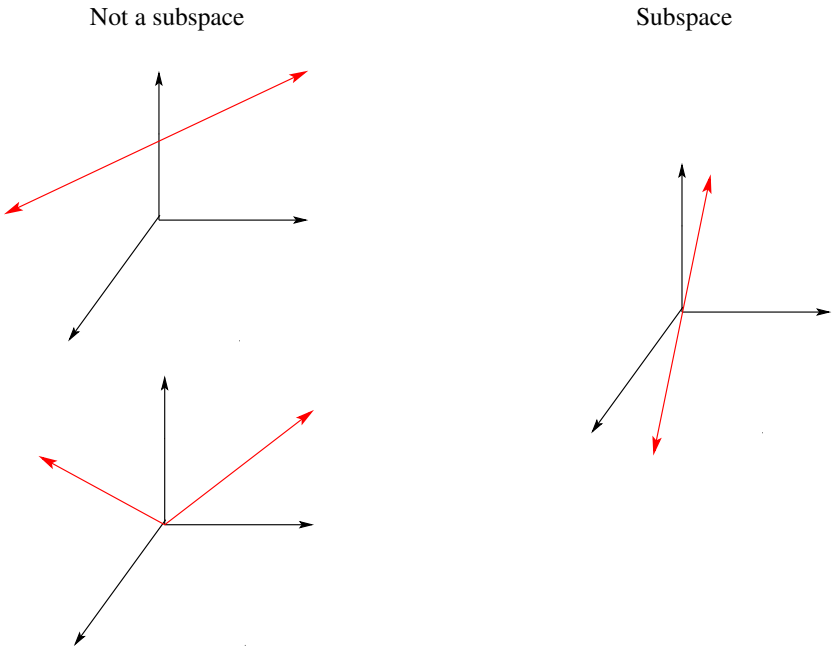
Definition 1.7 (Subspace)**.** *A subset $W \subseteq V$ of a vector space V is called a subspace if the following conditions hold*

(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 $T(e_i)$. □

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 $aw \in T(U)$. □

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. □

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

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

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. □

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 $\{e_1, \ldots, e_n\}$ is an orthonormal basis of V and $\{\tilde{e}_1, \ldots, \tilde{e}_m\}$ 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 $\{e_1^*, \ldots, e_n^*\}$ and $\{\tilde{e}_1^*, \ldots, \tilde{e}_m^*\}$. 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,* $\text{coker}(T) = \ker(T^t)$ *and* $\ker(T) = \text{coker}(T^t)$.

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

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)**.**

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

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*

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

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,*

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

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

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

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 an 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 $M^t M$ and $M M^t$; 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*

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

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.*

1.9 Positive Semi-definite Matrices

We will later see that a kernel κ is by definition any finitely positive semi-definite function on data space. That is, given any finite set of data points $x_1, \ldots, x_n$, $(\kappa(x_i, x_j))$ is a positive semi-definite matrix, which we now define:

Definition 1.29 (Positive Semi-definite and Positive Definite Matrices)**.** *A symmetric matrix $M \in \mathbb{R}^{n\times n}$ is called positive semi-definite if $v^t M v \geq 0$, $\forall v \in \mathbb{R}^n$. M is called positive definite if $v^t M v > 0$, $\forall v \in \mathbb{R}^n \setminus \{0\}$.*

Note that any matrix M can be written as a sum of symmetric and antisymmetric parts:

$$M = \frac{M + M^t}{2} + \frac{M - M^t}{2}.$$

Because $v^t(M - M^t)v \equiv 0$ for any matrix M and any vector v, positive semi-definiteness is defined only for symmetric matrices.

REMARK 1.19. *Positive (semi)-definite matrices should become your best friends for the following reasons:*

- *They arise as similarity measures, a.k.a. kernels, in machine learning.*
- *As we will discuss later, to minimize a multivariate function $f : \mathbb{R}^n \to \mathbb{R}$, we often apply an iterative algorithm that requires Taylor expanding f to the second order and descending in some direction using the Hessian matrix of f. We need this Hessian matrix to be positive definite in order for the algorithm to converge.*
- *Positive definite matrices also arise in solving a linear system of equations of the form $Qx = b$, where $Q \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}$ are given, and one needs to solve for $x \in \mathbb{R}^n$. When the dimension of Q is very large, directly inverting Q may be too slow and numerically unstable, but a matrix decomposition associated with positive definite matrices can readily yield solutions without matrix inversion (see Remark 1.21).*
- *They arise in the least-squares approximation of linear regression parameters (to be discussed in Section 3):*

  $$\hat{\beta} = \arg\min \|X\beta - Y\|_2^2 \;\Rightarrow (X^t X)\hat{\beta} = X^t Y,$$

  where the rows of matrices X and Y are observed predictive and response variables, respectively, and β is a vector of regression coefficients.
- *They arise in the computation of the Singular Value Decomposition (SVD) of a matrix (Section 1.10).*

There are in fact many equivalent ways of defining the positive semi-definiteness and positive definiteness properties. Let us first state these equivalent conditions for positive definite matrices. In order to understand one of these conditions, let us recall the definition of a minor of a matrix:

Definition 1.30 (Minor)**.** *Let M be an $m \times n$ matrix. A minor of order k, for $k \leq min(m, n)$, is the determinant of a submatrix $M_{I,J}$ of M obtained by taking k rows indexed by $I \subseteq \{1, \ldots, m\}$ and k columns indexed by $J \subseteq \{1, \ldots, n\}$. If $I = J$, the corresponding minor is called a principal minor. A principal minor of order k is called a leading principal minor if $I = J = \{1, 2, \ldots, k\}$, i.e. the corresponding submatrix, simply denoted M_I, consists of the first k rows and first k columns of M.*

We can now state the theorem characterizing positive definite matrices.

Theorem 1.17 (Positive Definiteness)**.** *Let $M \in \mathbb{R}^{n \times n}$ be a symmetric matrix. The following conditions are equivalent:*

1. *$v^t M v > 0$, $\forall v \in \mathbb{R}^n \setminus \{0\}$.*
2. *All eigenvalues of M are positive.*
3. *$M = LL^t$ for some invertible matrix $L \in \mathbb{R}^{n \times n}$.*
4. *All leading principal minors of M are positive; this condition is known as Sylvester's criterion.*
5. *There exists a unique lower triangular matrix $L \in \mathbb{R}^{n \times n}$ with positive diagonal entries such that $M = LL^t$. (This decomposition is known as the Cholesky decomposition, and the algorithm for computing L terminates successfully iff M is positive definite, yielding a quick method for testing whether a given matrix is positive definite).*

Proof. We will prove $(1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 1)$.

$(1 \Rightarrow 2)$ Let v be an eigenvector of M corresponding to its smallest eigenvalue $\lambda_{\min}$. Then, $v^t M v = \lambda_{\min} v^t v = \lambda_{\min} \|v\|_2^2 > 0$. Since $\|v\|_2^2 > 0$, we must have $\lambda_{\min} > 0$.

$(2 \Rightarrow 3)$ Consider the eigen-decomposition of M as $M = UDU^t$, where U is the matrix of eigenvectors along the columns, and D is a diagonal matrix of eigenvalues. Since all eigenvalues are positive, we can take $D = D^{\frac{1}{2}} D^{\frac{1}{2}}$, where $D^{\frac{1}{2}}$ is a diagonal matrix with the positive square roots of the eigenvalues along the diagonal. Thus, we can write $M = UD^{\frac{1}{2}}(UD^{\frac{1}{2}})^t$.

$(3 \Rightarrow 4)$ Let L_k be the submatrix of L containing the first k rows of L. Since L is invertible, the rows of L_k are linearly independent; thus, $\text{coker}(L_k) = \ker(L_k^t) = \{0\}$. Let $M_k = L_k L_k^t$ denote the submatrix of M consisting of the first k rows and k columns of M. Then, Theorem A.2 by Cauchy-Binet implies that $\det(M_k) \geq 0$. To see that the inequality must be strict, suppose $\det(M_k) = 0$, in which case M_k has a zero eigenvalue with a corresponding eigenvector, say $v \neq 0$. But, $v^t M_k v = 0 \Rightarrow \|L_k^t v\|_2^2 = 0 \Rightarrow L_k^t v = 0$, contradicting the fact that $\ker(L_k^t) = \{0\}$.

$(4 \Rightarrow 5)$ We will use induction on the dimension n. For $n = 1$, $M_{11} > 0$ imples that we can take $\sqrt{M_{11}}$ to be a positive real number; thus, we have the desired result $M = (\sqrt{M_{11}})(\sqrt{M_{11}})^t$. Now, assume that $(4 \Rightarrow 5)$ for some fixed positive integer n. For any

$(n+1) \times (n+1)$ matrix M satisfying the condition (4), we can express it in the following block form:

$$M = \begin{pmatrix} M_n & v \\ v^t & \alpha \end{pmatrix},$$

where M_n is an $n \times n$ matrix, $v \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$. By assumption, all leading principal minors of M are positive, including all leading principal minors of M_n. In particular, $\det(M_n) > 0$, and M_n is thus invertible. The block determinant formula (Theorem A.3) then implies that

$$0 < \det(M) = \det(M_n) \det(\alpha - v^t M_n^{-1} v) \Rightarrow \alpha - v^t M_n^{-1} v > 0.$$

By the induction hypothesis, there exists a lower triangular matrix L_n with positive diagonal entries, such that $M_n = L_n L_n^t$. Then,

$$\begin{pmatrix} L_n & 0 \\ v^t {L_n^t}^{-1} & \sqrt{\alpha - v^t M_n^{-1} v} \end{pmatrix} \begin{pmatrix} L_n^t & L_n^{-1} v \\ 0 & \sqrt{\alpha - v^t M_n^{-1} v} \end{pmatrix} = M,$$

where we have used the fact that $\alpha - v^t M_n^{-1} v$ is positive and thus has a real positive square root. Proof of the uniqueness of L is left as an exercise.

(You will show another construction of the Cholesky decomposition in Homework, where the diagonal entries of L will arise as the positive square roots of ratios of successive leading principal minors.)

$(5 \Rightarrow 1)$ Since a triangular matrix with positive diagonal entries is invertible, we have $L^t v \neq 0, \forall v \in \mathbb{R}^n \setminus \{0\}$. Hence,

$$v^t M v = v^t L L^t v = \|L^t v\|_2^2 > 0, \forall v \in \mathbb{R}^n \setminus \{0\}.$$

□

Similarly, we have the following equivalent conditions defining positive semi-definite matrices:

Theorem 1.18 (Positive Semi-definiteness)**.** *Let $M \in \mathbb{R}^{n \times n}$ be a symmetric matrix. The following conditions are equivalent:*

1. $v^t M v \geq 0$, $\forall v \in \mathbb{R}^n$
2. *All eigenvalues of M are ≥ 0.*
3. *$M = LL^t$ for some matrix $L \in \mathbb{R}^{n \times n}$, not necessarily invertible.*
4. *All principal minors, not just leading principal minors, of M are non-negative.*
5. *There exists a lower triangular matrix $L \in \mathbb{R}^{n \times n}$ with non-negative diagonal entries such that $M = LL^t$. (L may not be unique in this case).*

Proof. $5 \Rightarrow 1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4$ can be proven using the same approaches as in Theorem 1.17. To prove $4 \Rightarrow 5$, we will again use induction on the dimension n, with some important

modifications. For $n = 1$, $M_{11} \geq 0$ imples that we can take $\sqrt{M_{11}}$ to be a non-negative real number; thus, we have the desired result $M = (\sqrt{M_{11}})(\sqrt{M_{11}})^t$. Now, assume that $4 \Rightarrow 5$ for some fixed positive integer n; thus, via the cyclic implication of the above conditions, all $n \times n$ matrices satisfying condition (4) are positive semi-definite. For any $(n+1) \times (n+1)$ matrix M satisfying the condition (4), we can express it in the following block form:

$$M = \begin{pmatrix} M_n & v \\ v^t & \alpha \end{pmatrix},$$

where M_n is an $n \times n$ matrix, $v \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}$. By assumption, all principal minors of M are non-negative, including all principal minors of M_n. By the induction hypothesis, there thus exists a lower triangular matrix L_n with non-negative diagonal entries, such that $M_n = L_n L_n^t$, and M_n is positive semi-definite.

If $\alpha = 0$, then we must have $v = 0$; otherwise, a 2×2 principal submatrix indexed by the last row and a non-zero entry of v would have negative determinant. In this case, we thus have

$$M = \begin{pmatrix} L_n & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} L_n^t & 0 \\ 0 & 0 \end{pmatrix},$$

proving the theorem.

If $\alpha \neq 0$, then we must have $\alpha > 0$, since the principal minor containing only the last diagonal entry of M is assumed to be non-negative. We can thus apply the block determinant formula (Theorem A.3) to get

$$\det(M) = \alpha \det(M_n - \alpha^{-1} v v^t) \geq 0 \Rightarrow \det(M_n - \alpha^{-1} v v^t) \geq 0.$$

Thus, the non-negativity of $\det(M)$ implies the non-negativity of the determinant of the Schur complement $M/\alpha \equiv M_n - \alpha^{-1} v v^t$. We will now show that all other principal minors of M/α are also non-negative. Let $S \subseteq \{1, \ldots, n\}$ be any non-empty subset. Let M_S denote the principal submatrix of M with the columns and rows indexed by S, and so forth. Then, Theorem A.3 again implies that

$$\det(M_{S \cup \{n+1\}}) = \alpha \det\left[(M/\alpha)_S\right].$$

Since M is assumed to satisfy condition (4) and $\alpha > 0$, we see that $0 \leq \det(M_{S \cup \{n+1\}}) \Rightarrow 0 \leq \det\left[(M/\alpha)_S\right]$ for any index set $S \subseteq \{1, \ldots, n\}$. That is, the Schur complement M/α is an $n \times n$ symmetric matrix satisfying condition (4); by the induction hypothesis, we thus see that v and α cannot be arbitrary, but must satisfy the condition that M/α is positive semi-definite. Now, we will show that $v \perp \ker(M_n)$. Suppose there exists $w \in \ker(M_n)$, such that $v^t w \neq 0$. Then,

$$w^t (M/\alpha) w = w^t M_n w - \alpha^{-1} (w^t v)^2 = -\alpha^{-1} (w^t v)^2 < 0,$$

contradicting the fact the M/α is positive semi-definite. To find a constraint on α, consider

$$0 \leq (M_n^+ v)^t (M/\alpha)_S (M_n^+ v) = v^t M_n^+ M_n M_n^+ v - \alpha^{-1} (v^t M_n^+ v)^2 = v^t M_n^+ v - \alpha^{-1} (v^t M_n^+ v)^2$$

where M_n^+ is the Moore-Penrose pseudo-inverse (Theorem A.9). Since $\alpha > 0$, we must thus have

$$\alpha(v^t M_n^+ v) \geq (v^t M_n^+ v)^2.$$

Note that if $v = 0$, then there is no further restriction on α. If $v \neq 0$, then it can be shown that $v^t M_n^+ v > 0$ (see Section 1.10), in which case the constraint becomes

$$\alpha \geq v^t M_n^+ v.$$

Using these two constrains on v and α, we can now complete the proof of $4 \Rightarrow 5$. Consider the product

$$\begin{pmatrix} L_n & 0 \\ u^t & \beta \end{pmatrix} \begin{pmatrix} L_n^t & u \\ 0 & \beta \end{pmatrix} = \begin{pmatrix} L_n L_n^t & L_n u \\ u^t L_n^t & u^t u + \beta^2 \end{pmatrix}.$$

In order for this product to be equal to M, we need

$$L_n u = v \qquad \text{and} \qquad u^t u + \beta^2 = \alpha.$$

Since $v \perp \ker(M_n)$, we see that $v \in \mathrm{Im}(M_n) = \mathrm{Im}(L_n)$; hence, a solution u exists, and the minimum norm solution is $u_* = L_n^+ v$; but, the solution is not unique, since for any $w \in \ker(L_n)$, we have $L_n(u_* + w) = L_n u_* = v$. Hence, a generic solution is

$$\boxed{u = L_n^+ v + w}$$

for some $w \in \ker(L_n)$. As for β, we need to have

$$\beta^2 = \alpha - u^t u = \alpha - v^t L_n^{+t} L_n^+ v - \|w\|_2^2 = \alpha - v^t M_n^+ v - \|w\|_2^2,$$

where we have used the fact that $\ker(L_n) = \ker(L_n^{+t})$. Since we have already shown that $\alpha - v^t M_n^+ v \geq 0$, we can always tune w such that $\alpha - v^t M_n^+ v - \|w\|_2^2 \geq 0$, and we can thus obtain the positive square root

$$\boxed{\beta = \sqrt{\alpha - v^t M_n^+ v - \|w\|_2^2}}.$$

□

REMARK 1.20 (Variation of Condition (3))**.** *For any $m \times k$ matrix M, MM^t and $M^t M$ are positive semi-definite, since for all $v \in \mathbb{R}^m$, we have $v^t M M^t v = \|M^t v\|_2^2 \geq 0$ and $v^t M^t M v = \|Mv\|_2^2 \geq 0$. Both MM^t and $M^t M$ are examples of the so-called Gram matrix, whose elements arise from taking pairwise dot-products of the columns of M^t and M, respectively.*

More generally, we define:

Definition 1.31 (Gram Matrix)**.** *The Gram matrix G of n vectors $x_1, \ldots, x_n$ in an inner product space $(V, \langle , \rangle)$ is defined as $G_{ij} = \langle x_i, x_j \rangle, i, j = 1, \ldots n$. By the same reasoning as in Remark 1.20, the Gram matrix is always positive semi-definite.*

The following properties immediately follows from condition 4 of Theorem 1.18, but we can also directly prove them:

Corollary 1.7. *If $M = (m_{ij})$ is positive semi-definite, then all diagonal elements are ≥ 0. If M is positive definite, then the inequality $\geq$ is strict (i.e. $>$).*

Proof. Taking $e_i = (\delta_{ji})$, we see that $0 \leq e_i^t M e_i = m_{ii}$. □

Corollary 1.8. *If $M = (m_{ij})$ is positive semi-definite and $m_{ii} = 0$ for some i, then $m_{ij} = m_{ji} = 0$ for all j.*

Proof. Exercise. □

REMARK 1.21 (Cholesky Decomposition for Solving Linear Equations)**.** *Suppose you need to solve for x in the linear equation*

$$Qx = b,$$

where Q is a symmetric positive definite matrix. When the dimension of Q is very large, directly inverting Q may not be a good approach. Fortunately, we know that Q has the Cholesky decomposition $Q = LL^t$; thus, we can instead look for y that satisfies $Ly = b$ and then look for x that satisfies $L^t x = y$. Both of these equations involving the triangular matrices L and L^t can be solved using simple forward and back substitution methods, respectively, without inverting these matrices.

If Q is invertible but not positive definite, then we can instead solve the equation

$$Q^t Q x = Q^t b$$

for x using the Cholesky decomposition of $M \equiv Q^t Q$. The obtained solution x must also satisfy $Qx = b$, since Q – thus, also Q^t – is assumed to be invertible.

Definition 1.32 (Diagonally Dominant Matrix)**.** *An $n \times n$ matrix M is called diagonally dominant if $|m_{ii}| \geq \sum_{j \neq i} |m_{ij}|$ for all $i = 1, \ldots, n$.*

Example 1.13 (Graph Laplacian)**.** *Let Γ be an undirected graph with n nodes, and denote the edge weight between the i-th and j-th nodes as $w_{ij} \geq 0$. By convention, let $w_{ii} = 0$, i.e. no self-edges. The degree δ_i of node i is the sum over the weights of all edges connected to node i; i.e. $\delta_i = \sum_{j=1}^n w_{ij}$. Let $W = (w_{ij})$ be the matrix of edge weights and $D = \mathrm{diag}(\delta_1, \ldots, \delta_n)$. Then, the so-called unnormalized graph Laplacian $L = D - W$ is a diagonally dominant matrix.*

Theorem 1.19. *A diagonally dominant symmetric matrix M with non-negative diagonal elements is positive semi-definite.*

Proof. The stated conditions imply that each Gershgorin disc $D_i = \{z \in \mathbb{C} | \, |z - m_{ii}| \leq \sum_{j \neq i} |m_{ij}|\}$ consists of points that have non-negative real values. The Gershgorin Circle Theorem (Theorem 1.10) thus implies that all eigenvalues of M are non-negative. □

Corollary 1.9. *The unnormalized graph Laplacian of any undirected graph is positive semi-definite.*

Proof. In Example 1.13, we saw that an unnormalized graph Laplacian is diagonally dominant. Since its diagonal elements are non-negative, it is positive semi-definite by Theorem 1.19. □

1.10 Singular Value Decomposition

For a general rectangular matrix, we cannot talk about its eigenvalues and eigenvectors, because it will map a vector space to another vector space of a different dimension. But, for any real $m \times n$ matrix M, we can construct two real symmetric matrices: M^tM and MM^t. Because M^tM is symmetric, it always has n real eigenvalues $\lambda_1, \ldots, \lambda_n$ and n orthonormal real eigenvectors $v_1, v_2, \ldots, v_n$. Likewise, MM^t has m real eigenvalues $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_m$ and m orthonormal real eigenvectors $w_1, w_2, \ldots, w_m$. We will show that

- the eigenvalues of M^tM and MM^t cannot be negative. (This condition will allow us to take square roots of non-zero eigenvalues).
- the set of positive eigenvalues of M^tM is identical to the set of positive eigenvalues of MM^t. (This condition will allow us to consider only the eigenvalues of M^tM and not worry about MM^t).
- there is a one-to-one correspondence between the eigenvectors of M^tM with non-zero eigenvalues and the eigenvectors of MM^t with non-zero eigenvalues. (This condition will allow us to reconstruct w's from v's).

These properties will allow us to decompose any $m \times n$ matrix M as a sum of products of w's and v^t's, with coefficients given by the square root of the eigenvalues.

Theorem 1.20. *M^tM and MM^t are positive semi-definite, i.e. their eigenvalues are non-negative.*

Proof. If M is a square matrix, then Theorem 1.18 implies that M^tM and MM^t are positive semi-definite and have non-negative eigenvalues. For rectangular M, we similarly note that for all $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^m$,

$$v^tM^tMv = \|Mv\|_2^2 \geq 0 \quad \text{and} \quad w^tMM^tw = \|M^tw\|_2^2 \geq 0.$$

Hence, both M^tM and MM^t are positive semi-definite, and Theorem 1.18 implies that they have non-negative eigenvalues. □

Theorem 1.21. *The set of non-zero eigenvalues of M^tM is identical to the set of non-zero eigenvalues of MM^t, with the same multiplicity.*

Proof. (Note that since the eigenvalues of M^tM and MM^t are non-negative, all non-zero eigenvalues are positive.) Suppose M^tM has k positive eigenvalues and MM^t has ℓ positive eigenvalues. We will show that $k = \ell$ and that the eigenvalues are the same. Suppose $v_1, \ldots, v_k$ are eigenvectors of M^tM with positive eigenvalues $\lambda_1, \ldots, \lambda_k$, respectively. Then, we can see that the vectors defined by $w_i = Mv_i$ are eigenvectors of MM^t with eigenvalues λ_i: $MM^t(Mv_i) = M(M^tM)v_i = M(\lambda v_i) = \lambda_i(Mv_i)$. Note that from the proof of Theorem 1.20, we know that $\|Mv_i\|^2 = \lambda_i\|v_i\|^2 > 0$, since λ_i is assumed to be positive; hence, Mv_i is non-zero, and it is indeed an eigenvector of MM^t with an eigenvalue λ_i. To see that $Mv_1, Mv_2, \ldots Mv_k$ are linearly independent, note that $\langle Mv_i, Mv_j \rangle = \langle v_i, M^tMv_j \rangle =$

$\lambda_j \langle v_i, v_j \rangle = 0$ for $i \neq j$, since v_i are orthonormal. Thus, $Mv_1, \ldots, Mv_k$ themselves are also orthogonal (but, not normalized to 1, as we will see below). Hence, we see

$$\left\{ \begin{array}{c} \text{orthogonal eigenvectors} \\ v_1, \ldots, v_k \\ \text{of } M^tM \text{ with positive eigenvalues} \\ \lambda_1, \ldots, \lambda_k \end{array} \right\} \Rightarrow \left\{ \begin{array}{c} \text{orthogonal eigenvectors} \\ Mv_1, \ldots, Mv_k \\ \text{of } MM^t \text{ with positive eigenvalues} \\ \lambda_1, \ldots, \lambda_k \end{array} \right\}$$

Thus, we see that $k \leq \ell$. Conversely, suppose $w_1, \ldots, w_\ell$ are eigenvectors of MM^t with positive eigenvalues $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_\ell$, respectively. Then, a similar arguments shows that

$$\left\{ \begin{array}{c} \text{orthogonal eigenvectors} \\ w_1, \ldots, w_\ell \\ \text{of } MM^t \text{ with positive eigenvalues} \\ \tilde{\lambda}_1, \ldots, \tilde{\lambda}_\ell \end{array} \right\} \Rightarrow \left\{ \begin{array}{c} \text{orthogonal eigenvectors} \\ M^tw_1, \ldots, M^tw_\ell \\ \text{of } M^tM \text{ with positive eigenvalues} \\ \tilde{\lambda}_1, \ldots, \tilde{\lambda}_\ell \end{array} \right\}$$

Thus, we see that $\ell \leq k$. Combining the two results, we see that $k = \ell$ and that the positive eigenvalues are the same between M^tM and MM^t. □

Since M^tM and MM^t have the same set of non-zero eigenvalues, we see that

Corollary 1.10. *M, M^t, M^tM and MM^t have the same rank.*

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ be the positive eigenvalues of M^tM. Since these values are positive, we can take square roots that are real:

Definition 1.33 (Singular Values and Singular Vectors)**.** *The square roots of the positive eigenvalues of M^tM are called the singular values of M. The corresponding orthonormal eigenvectors of M^tM are called the right singular vectors. That is, if $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$ are the positive eigenvalues of M^tM, then the singular values are $\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \cdots, \sigma_k = \sqrt{\lambda_k}$. Since square root is a monotonic function, we have $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$.*

From the proof of Theorem 1.21, we know that we can reconstruct all eigenvectors of MM^t with positive eigenvalues from the orthonormal eigenvectors $v_i, \ldots, v_k$ of M^tM with positive eigenvalues by taking $Mv_1, \ldots, Mv_k$. We have already seen that $Mv_1, \ldots, Mv_k$ are orthogonal, so we just need to know the norm of Mv_i.

Theorem 1.22. *If v is a unit eigenvector of M^tM with eigenvalue $\sigma^2 > 0$, then $w = Mv/\sigma$ is a unit eigenvector of MM^t with eigenvalue σ^2.*

Proof. We already know from Theorem 1.21 that Mv is an eigenvector of MM^t with eigenvalue σ^2. The norm of Mv is:

$$\|Mv\|^2 = \langle Mv, Mv \rangle = \langle v, M^tMv \rangle = \sigma^2 \langle v, v \rangle = \sigma^2,$$

where we have used the fact that v is a unit vector. Thus, Mv/σ is a unit vector. □

To summarize, for any real $m \times n$ matrix M,

- M^tM and MM^t have non-negative eigenvalues.
- M^tM and MM^t have identical sets of positive eigenvalues, including multiplicity.
- If $\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_k^2 > 0$ are the positive eigenvalues of M^tM with orthonormal eigenvectors $v_1, \ldots, v_k$, then $w_i := Mv_i/\sigma_i$ are the orthonormal eigenvectors of MM^t with eigenvalues σ_i^2.

Theorem 1.23 (Singular Value Decomposition). *Let $M : V \to W$ be a real $m \times n$ matrix with singular values $\sigma_1 \geq \cdots \geq \sigma_k$ and corresponding right singular vectors $v_1, \ldots, v_k$. Then, we have the Singular Value Decomposition*

$$M = (w_1 w_2 \cdots w_k) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{pmatrix} \begin{pmatrix} v_1^t \\ v_2^t \\ \vdots \\ v_k^t \end{pmatrix}, \tag{1.6}$$

where $w_i = Mv_i/\sigma_i$, which we call the left singular vectors.

Proof. Choose as a basis of V, $\{v_1, \ldots, v_k, v_{k+1}, \ldots, v_n\}$, where $v_{k+1}, \ldots, v_n$ are the null eigenvectors of M^tM. Similarly, choose as a basis of W, $\{w_1, \ldots, w_k, w_{k+1}, \ldots, w_m\}$, where $w_{k+1}, \ldots, w_m$ are the null eigenvectors of MM^t. Then, in these bases, one can easily show that the left and right hand sides of (1.6) are identical maps from V to W. □

An equivalent way of writing (1.6) is

$$\boxed{M = \sum_{i=1}^{k} \sigma_i w_i v_i^t}. \tag{1.7}$$

(N.B. Note the similarity between (1.7) and (1.5). Thus, SVD of a rectangular matrix is a generalization of diagonalizing a symmetric matrix.)

REMARK 1.22. *A geometric interpretation of SVD of $M \in \mathbb{R}^{m \times n}$ is mapping orthonomal axes of spheres in $\mathbb{R}^n$ to the principal axes of (collapsed) ellipsoids in $\mathbb{R}^m$.*

REMARK 1.23. *If $M \in \mathbb{R}^{m \times n}$ is a data matrix of m i.i.d. samples of n mean-centered features, then the first right singular vector v_1 of M is the first principal component, and the k-th entry of $\sigma_1 w_1$ is the coordinate of the k-th row of M projected onto v_1. More generally, $\sigma_i w_i$ yields the coordinates of the rows of M projected onto the i-th principal component v_i.*

EXERCISE 1.14. *Show that SVD can be used to construct the polar decompositions $M = PU = \tilde{U}\tilde{P}$, where $P, \tilde{P}$ are positive semi-definite matrices and $U, \tilde{U}$ are orthogonal matrices.*

1.10.1 Symmetric Matrices: Relation Between Eigenvalues and Singular Values

If M is symmetric, then $M^tM = M^2$. Clearly, all eigenvectors of M are also the right singular vectors. Hence, the eigenvalues of M^tM are just squares of the eigenvalues of M. Since the singular values are defined to be the square roots of the eigenvalues of M^tM, we thus see that the singular values of a symmetric matrix M are just absolute values of the eigenvalues of M. Thus, the spectral radius $\rho(M)$ is equal to the largest singular value σ_1. The left singular vectors are

$$w_i = \frac{Mv_i}{\sigma_i} = \pm v_i,$$

where we take the $+$ sign if the eigenvalue of v_i with respect to M (not M^tM) is positive, or the $-$ sign, otherwise.

1.10.2 Normal Matrices: Relation Between Eigenvalues and Singular Values

More generally, a real normal matrix $M \in \mathbb{R}^{n\times n}$ can be diagonalized using a unitary matrix U of eigenvectors as

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

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix of the eigenvalues of M. Note that the eigenvalues λ_i may be complex; in fact, all eigenvalues of a normal matrix are real iff the matrix is also symmetric. Let us assume that we have ordered the eigenvalues, such that

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Since, $M^\dagger = M^t$, this decomposition yields

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

implying that the singular values σ_i of M satisfy

$$\sigma_i = |\lambda_i|.$$

If a singular value σ_i is nondegenerate, then it implies that the corresponding λ_i was real, in which case we can choose the corresponding eigenvector to be purely real; this eigenvector of M is also the right singular vector for σ_i. If σ_i is degenerate, then the right singular vectors corresponding to σ_i can be chosen as any orthogonal unit linear combinations of the real and imaginary parts of the corresponding eigenvectors of M that have absolute eigenvalue equal to σ_i.

1.11 Conjugate Directions

As we will discuss later, positive definite matrices allow us to define special directions that facilitate optimization and other numerical techniques. Intuitively, given a positive definite matrix $M \in \mathbb{R}^{n\times n}$, let us choose a basis consisting of its orthonormal eigenvectors v_i with corresponding eigenvalues λ_i, $i = 1, \ldots, n$, and inversely scale each eigen-direction v_i by $\sqrt{\lambda_i}$.

That is, under this transformation, an arbitrary vector $u = \sum_{i=1}^n \alpha_i v_i$ gets mapped as

$$u = \sum_{i=1}^n \alpha_i v_i \mapsto \tilde{u} = \sum_{i=1}^n \beta_i v_i, \text{ where } \beta_i \equiv \frac{\alpha_i}{\sqrt{\lambda_i}}.$$

A sphere of radius R defined by $\sum_{i=1}^n \alpha_i^2 = R^2$ thus gets mapped to an ellipsoid

$$\tilde{u}^t M \tilde{u} = \sum_{i=1}^n \lambda_i \beta_i^2 = R^2.$$

Images of orthogonal directions under this transformation are conjugate directions:

Definition 1.34 (Conjugate Vectors)**.** *Two vectors $p, q \in \mathbb{R}^n$ are said to be conjugate with respect to (w.r.t.) a positive definite $n \times n$ matrix M if $p^t M q = 0$, i.e. they are orthogonal w.r.t. the inner product $\langle p, q \rangle_M \equiv p^t M q$.*

EXERCISE 1.15. *Show that $\langle , \rangle_M$ defined above indeed satisfies the symmetry, linearity, and positive definiteness properties of an inner product.*

Theorem 1.24. *Non-zero vectors $p_1, \ldots, p_k \in \mathbb{R}^n$ that are mutually conjugate w.r.t. a positive definite matrix M are linearly independent.*

Proof. Suppose $\sum_{i=1}^k \alpha_i p_i = 0$ for some coefficients $\alpha_i \in \mathbb{R}, i = 1, \ldots, k$. Then, for any fixed $j \in \{1, \ldots, k\}$, we have

$$0 = \langle p_j, \sum_{i=1}^k \alpha_i p_i \rangle_M = \sum_{i=1}^k \alpha_i \langle p_j, p_i \rangle_M = \alpha_j \langle p_j, p_j \rangle_M,$$

where we have used the fact that $\langle p_j, p_i \rangle_M = 0$ for $i \neq j$. Since $p_j \neq 0$ and M is positive definite, we must have $\langle p_j, p_j \rangle_M > 0$ and, thus, $\alpha_j = 0$. Since j was arbitrary, we must have $\alpha_j = 0 \ \forall j \in \{1, \ldots, n\}$.

□

In particular, when $k = n$, we have

Corollary 1.11. *If n non-zero vectors $p_1, \ldots, p_n$ in $\mathbb{R}^n$ are mutually conjugate w.r.t. a positive definite matrix M, then they form a basis of $\mathbb{R}^n$.*

How does one obtain conjugate vectors? The following theorem tells us how to turn any set of linearly independent vectors into a set of conjugate vectors.

Theorem 1.25. *Given a set $\{u_1, \ldots, u_k\} \in \mathbb{R}^n$ of linearly independent vectors and a positive definite $n \times n$ matrix M, the vectors $p_i, i = 1, \ldots, k$, defined by*

$$p_1 = u_1 \text{ and } p_i = u_i - \sum_{j=1}^{i-1} \frac{\langle p_j, u_i \rangle_M}{\langle p_j, p_j \rangle_M} p_j, \ 2 \leq i \leq k, \tag{1.8}$$

are mutually conjugate w.r.t. M.

Proof. Note that each p_i is a linear combination of $u_1, \ldots, u_i$ with the coefficient of u_i being 1; linear independence of $u_1, \ldots, u_k$ thus implies that $p_i \neq 0$. As a result, $\langle p_i, p_i \rangle_M \neq 0, \forall i = 1, \ldots, k$, and the coefficients in (1.8) are thus well defined. The shown orthogonalization is just the general Gram-Schmidt process (Theorem 1.1) without the normalization step. □

REMARK 1.24. *Note that the Gram-Schmidt process (1.8) depends on the ordering of the elements $u_1, \ldots, u_k$.*

Let us now see why conjugate directions may be useful.

Example 1.14. *Suppose you need to solve $Qx = b$, where $Q \in \mathbb{R}^{n \times n}$ is a positive definite matrix. Pick some set of n conjugate directions $\{p_1, \ldots, p_n\}$ w.r.t. Q. Then, since $\{p_1, \ldots, p_n\}$ forms a basis, the solution x_* can be expanded as*

$$x_* = \sum_{i=1}^{n} \tilde{\alpha}_{*i} \, p_i \,, \quad \tilde{\alpha}_{*i} \in \mathbb{R}, \tag{1.9}$$

and the problem reduces to finding the expansion coefficients $\tilde{\alpha}_{*i}$. Plugging in this expression into $Qx_* = b$ and multiplying both sides by p_k^t for each fixed k, we get

$$\sum_{i=1}^{n} \tilde{\alpha}_{*i} \, p_k^t Q p_i = p_k^t b \;\; \Rightarrow \;\; \tilde{\alpha}_{*k} = \frac{p_k^t b}{\langle p_k, p_k \rangle_Q}, \tag{1.10}$$

where we have used the fact that $\langle p_k, p_i \rangle_Q = 0$, for $i \neq k$.

1.11.1 Iterative Conjugate Direction Method

The formula (1.10) provides an easy method for finding the solution to $Qx = b$ when Q is positive definite. Even though (1.10) is valid for any choice of conjugate directions, not all choices are equally useful, especially when the dimension n is very large. For example, when $n \gg 1$, obtaining conjugate directions from an arbitrary basis via (1.8) requires computing and storing n^2 entries; furthermore, most of n coefficients in (1.10) may be non-zero and may need to be computed. It would thus be useful to tailor the conjugate directions to the given problem $Qx = b$, so that we only need a subset of conjugate directions and expansion coefficients to obtain either the exact solution or its accurate approximation.

We will construct such informative directions in Section 7 by noticing that the solution x_* to $Qx = b$ also solves the minimization problem

$$x_* = \arg\min f(x), \;\; \text{where } f(x) = \left(\frac{1}{2} x^t Q x - x^t b \right), \tag{1.11}$$

and performing iterative descent along conjugate directions adjusted from the gradient $-\nabla f$; note that the gradient $-\nabla f$ itself represents the steepest descent direction at each iteration, which can be undesirable when the condition number, defined to be the ratio of the maximum and minimum singular values of Q, is large (see Section A.3). This approach of modifying the gradient descent direction via the Gram-Schmidt orthogonalization (1.8), using Q as

the positive definite matrix M, is called the Conjugate Gradient Descent method. We will develop this approach in Section 7.

For now, let us show that the solution (1.9) to the problem $Qx = b$ using any set of mutually conjugate directions w.r.t. Q can be derived as an iterative solution to a quadratic optimization problem (1.11), variations of which have many applications in training machine learning models. First, note that we can reparametrize the solution x_* in (1.9) around any fixed $x_1 \in \mathbb{R}^n$ as

$$x_* = x_1 + \sum_{i=1}^{n} \alpha_{*i}\, p_i\,, \;\; \alpha_{*i} \in \mathbb{R},$$

where

$$\boxed{\alpha_{*k} = -\frac{p_k^t g_1}{\langle p_k, p_k \rangle_Q}}\,, \tag{1.12}$$

where $g_1 \equiv \nabla f(x_1) = Qx_1 - b$ is the gradient of f at x_1.

EXERCISE 1.16. *Derive this expression* (1.12) *by appropriately modifying* (1.10).

We claim that we can think of x_* as being the termination point, $x_* = x_{n+1}$, of the following iterative updates:

$$x_{k+1} = x_k + \alpha_{*k}\, p_k,$$

where

$$\alpha_{*k} = \arg\min_{\alpha_k} f(x_k + \alpha_k\, p_k).$$

That is, α_{*k} is the optimal value that minimizes $f(x)$ in the line search along the affine subspace

$$x_k + \text{Span}\{p_k\} \equiv \{x \in \mathbb{R}^n \mid x = x_k + \alpha_k\, p_k,\;\; \alpha_k \in \mathbb{R}\}.$$

The following lemma will be useful in the discussion:

Lemma 1.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic function with a Hessian matrix Q. Then, $\forall x, y \in \mathbb{R}^n$,*

$$\nabla f(x) - \nabla f(y) = Q(x - y),$$

where the gradients $\nabla f(x)$ and $\nabla f(y)$ are represented as column vectors of partial derivatives of f at x and y, respectively, and $(x - y)$ is a column vector. Using the iterative coordinate update $x_{k+1} = x_k + \alpha_{*k}\, p_k$ and the notation $g_k = \nabla f(x_k)$, we thus have the following iterative gradient update:

$$g_{k+1} = g_k + \alpha_{*k} Q p_k.$$

Proof. Exercise. □

To minimize $f(x)$ iteratively, starting from x_1 and iteratively descending along the directions $p_1, \ldots, p_n$, we need to update $x_{k+1} = x_k + \alpha_{*k}\, p_k$ from x_k by computing the directional derivative

$$\frac{df(x_k + \alpha_k\, p_k)}{d\alpha_k} = p_k^t(Q(x_k + \alpha_k\, p_k) - b) = p_k^t(g_k + \alpha_k Q p_k)\,,$$

where $g_k \equiv \nabla f(x_k) = Qx_k - b$ is the gradient of f at x_k, which is also the residual in attempting to solve $Qx = b$. Solving for α_{*k} at which $\frac{df(x_k+\alpha_k p_k)}{d\alpha_k}\Big|_{\alpha_{*k}} = 0$ yields

$$\boxed{\alpha_{*k} = -\frac{p_k^t g_k}{\langle p_k, p_k \rangle_Q}}, \tag{1.13}$$

which agrees with (1.12), since Lemma 1.1 implies that

$$p_k^t g_k = p_k^t \left(g_1 + \sum_{i=1}^{k-1} \alpha_{*i}\, Q p_i \right) = p_k^t g_1,$$

where the last equality follows from the conjugate direction condition $\langle p_k, p_i \rangle_Q = 0, i \neq k$. Thus, $g_{n+1} = Qx_{n+1} - b = 0$, and the iterative optimization terminates.

REMARK 1.25. *We can also derive* (1.13) *as solutions to the optimization problem* (1.11) *by rewriting the cost function* $f(x)$ *as*

$$\begin{aligned} f(x) &= \sum_{k=1}^{n} \left[\frac{1}{2}(x_1 + \alpha_k p_k)^t Q (x_1 + \alpha_k p_k) - (x_1 + \alpha_k p_k)^t b \right] + c_1 \\ &= \sum_{k=1}^{n} \left[\frac{1}{2} \langle p_k, p_k \rangle_Q \, (\alpha_k - \alpha_{*k})^2 \right] + c_2 \ , \end{aligned}$$

where c_1 *and* c_2 *are constants, clearly showing that the minimization of* $f(x)$ *decouples into a set of independent univariate minimization along different conjugate directions.*

These results prove the first part of the following important theorem for quadratic function optimization:

Theorem 1.26. *The iterative conjugate direction method for a quadratic function* $f : \mathbb{R}^n \to \mathbb{R}$ *with a positive definite Hessian terminates in at most* n *exact line searches along directions* $p_1, \ldots, p_n$ *conjugate w.r.t. to the Hessian. Furthermore,* x_{k+1} *minimizes* $f(x)$ *in the affine subspace*

$$x_1 + \mathrm{Span}\{p_1, \ldots, p_k\} \equiv \{x \in \mathbb{R}^n \mid x = x_1 + \sum_{i=1}^{k} \alpha_i\, p_i,\ \alpha_i \in \mathbb{R}\}.$$

To prove the second part of the theorem, we need the following lemma:

Lemma 1.2. *In the iterative conjugate direction method using directions* $p_1, \ldots, p_n$ *conjugate w.r.t. the positive definite Hessian of a quadratic function* $f : \mathbb{R}^n \to \mathbb{R}$*, we have for all* $1 \leq k \leq n$*,*

$$p_i^t\, g_{k+1} = 0,\ 1 \leq i \leq k,$$

where $g_{k+1} = \nabla f(x_{k+1})$*. That is* g_{k+1} *is orthogonal to all preceding conjugate directions* $p_1, \ldots, p_k$*.*

Proof. Let Q denote the Hessian of f. For $i = k$, note that $p_k^t g_{k+1} = \frac{df(x_k+\alpha_k p_k)}{d\alpha_k}|_{\alpha_{*k}} = 0$. For $i < k$, rewriting

$$p_i^t g_{k+1} = \sum_{j=i+1}^{k} p_i^t (g_{j+1} - g_j) + p_i^t g_{i+1}$$

and noting that Lemma 1.1 implies that

$$g_{j+1} - g_j = \alpha_{*j} Q p_j \,,$$

we get

$$p_i^t g_{k+1} = \sum_{j=i+1}^{k} \alpha_{*j} p_i^t Q p_j + p_i^t g_{i+1} = 0 \,,$$

where we have used the fact that $\langle p_i, p_j \rangle_Q = 0, i \neq j$, and $p_i^t g_{i+1} = 0$ from the first part of the proof. □

Proof of Theorem 1.26. The first part of the Theorem was already proved by the explicit calculation of the line-search optimal values α_{*k} that agree with the exact solution (1.12) to $Qx = b$. It also follows from Lemma (1.2), which implies that g_{n+1} is orthogonal to all conjugate directions (which span $\mathbb{R}^n$) and, thus, must be 0. To prove the second part, note that the directional derivative of f at x_{k+1} along any vector $\sum_{i=1}^{k} \gamma_i p_i$ in the affine subspace is

$$\left(\sum_{i=1}^{k} \gamma_i p_i \right)^t g_{k+1} = 0 \quad \text{(by Lemma 1.2)}.$$

This calculation shows that x_{k+1} is a critical point of f restricted to the affine subspace. Since the Hessian of f restricted to the affine subspace is also positive definite (why?), we see that x_{k+1} is in fact a global minimum of f in the affine subspace. □

REMARK 1.26. *All conjugate direction methods rely on Theorem 1.26, even though they may differ in choosing the algorithm-specific conjugate directions.*

1.12 Projection Operators and Distance between Subspaces

Definition 1.35 (Projection)**.** *Let V be a vector space. A linear map $P : V \to V$ is called a projection map if it is idempotent, i.e. $P^2 = P$.*

The range of P, or the column span of P in some matrix representation of P, will define a subspace of V invariant under P.

EXERCISE 1.17. *Show that if $P : V \to V$ is a projection on vector space V, then so is $I - P$, where I is the identity map on V. Show that* $\ker(P) = \mathrm{Im}(I - P)$ *and* $\ker(I - P) = \mathrm{Im}(P)$.

Example 1.15. $P = \begin{pmatrix} 1 & \alpha \\ 0 & 0 \end{pmatrix}$ *is a projection of $\mathbb{R}^2$ onto the x-axis for any $\alpha \in \mathbb{R}$.*

The above example shows that projection operators onto the same subspace may not be unique. To find one-to-one correspondence between subspaces and project operators, we need to impose the following orthogonality constraint:

Definition 1.36 (Orthogonal Projection)**.** *Let V be a vector space with an inner product $\langle \cdot, \cdot \rangle$. A linear map $P : V \to V$ is called an orthogonal projection if it is idempotent and $\forall v \in V, \langle v - P(v), P(v) \rangle = 0$.*

Note in the above definition that since $(I - P)Pw = 0$ for any $w \in V$, we have $0 = \langle (I-P)(v+Pw), P(v+Pw) \rangle = \langle (I-P)v, P(v+w) \rangle = \langle (I-P)v, Pw \rangle$. Hence, a projection P is an orthogonal projection if $\langle (I - P)v, Pw \rangle = 0$ for all $v, w \in V$. If P is an orthogonal projection onto $S \subset V$, then $I - P$ projects onto the orthogonal complement of S.

EXERCISE 1.18. *Show that a projection P is orthogonal iff its matrix representation is symmetric in an orthonormal basis.*

Theorem 1.27. *The orthogonal projection map onto a subspace is unique.*

Proof. Suppose $P_1, P_2 \in \mathbb{R}^{n \times n}$ are orthogonal projection matrices onto the same subspace $S \subseteq \mathbb{R}^n$. Then, $\forall v \in \mathbb{R}^n$,

$$\begin{aligned} \|(P_1 - P_2)v\|_2^2 &= v^t(P_1 - P_2)^2 v = v^t(P_1^2 - P_1P_2 - P_2P_1 + P_2^2)v \\ &= v^t(P_1 - P_1P_2 - P_2P_1 + P_2)v = v^tP_1(I - P_2)v + v^tP_2(I - P_1)v = 0 \end{aligned}$$

where the last equality follows from the fact that $I - P_1$ and $I - P_2$ are projection operators onto the orthogonal complement $S^\perp$. Since the only vector with a zero norm is the zero vector, we have $\forall v \in \mathbb{R}^n$

$$(P_1 - P_2)v = 0 \Rightarrow P_1v = P_2v \Rightarrow P_1 = P_2.$$

□

Conversely, given a m-dimensional subspace $S \subseteq \mathbb{R}^n$, choose an orthonomal basis $u_1, \ldots, u_m$ of S and organize them as the columns of a matrix $U = (u_1 \cdots u_m)$. Then, $P \equiv UU^t$ is the unique orthogonal projection operator onto S. There is thus one-to-one correspondence between subspaces and orthogonal project operators. This correspondence allows us to measure the distance between subspaces S_1 and S_2 of equal dimension, with orthogonal projection operators P_1 and P_2, respectively, as

$$d(S_1, S_2) = \|P_1 - P_2\|, \tag{1.14}$$

where $\|P_1 - P_2\|$ is some norm on the space of matrices. We will learn in Section 1.15.3 how to compute this distance explicitly.

REMARK 1.27. *This definition doesn't work if S_1 and S_2 have different dimensions. For example, if S_1 is a subspace of S_2, then we would like to think of their distance to be 0; but, the right-hand side of (1.14) is non-zero.*

1.13 Power Method for Approximating the Dominant Eigenvalue and Eigenvector

Finding the exact eigenvalues of a large matrix can be difficult, because one cannot usually find the roots of its characteristic polynomial algebraically. However, the largest eigenvalue of certain diagonalizable matrices can be easily obtained via a stochastic algorithm.

Definition 1.37 (Dominant Eigenvalue)**.** *An eigenvalue λ of a square matrix M is called the dominant eigenvalue of M if, for any other eigenvalue μ of M, $|\lambda| > |\mu|$. An eigenvector corresponding to λ is called a dominant eigenvector.*

Theorem 1.28. *Let $M \in \mathbb{R}^{n\times n}$ be diagonalizable over $\mathbb{R}$ and with a dominant eigenvalue. Then, there exists $v \in \mathbb{R}^n$, such that $v_k \equiv M^k v$ approximates a dominant eigenvector for sufficiently large k. Hence, the dominant eigenvalue is approximated by the Rayleigh quotient*

$$\frac{v_k^t M v_k}{v_k^t v_k}.$$

Proof. Since M is diagonalizable, it has n linearly independent eigenvectors $u_1, \ldots, u_n \in \mathbb{R}^n$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Without loss of generality, assume that λ_1 is the dominant eigenvalue. Since the eigenvectors form a basis, any vector $v \in \mathbb{R}^n$ can be expanded as

$$v = \sum_{i=1}^{n} c_i\, u_i, \quad c_i \in \mathbb{R}.$$

Then, for any vector v such that $c_1 \neq 0$ in the above expansion, we have

$$M^k v = \sum_{i=1}^{n} \lambda_i^k c_i\, u_i = \lambda_1^k \left(c_1 u_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k c_2 u_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^k c_n u_n \right).$$

But, since λ_1 is the dominant eigenvalue, we have $|\lambda_i/\lambda_1|^k \to 0$, for $i > 1$, as $k \to \infty$. Hence, for sufficiently large k, $M^k v$ becomes approximately proportional to u_1. □

REMARK 1.28. *What it means for the power k to be "sufficiently large" depends on the magnitude of the remaining eigenvalues. That is, in order for the power method to work well, the second largest eigenvalue should be small compared to the dominant eigenvalue.*

REMARK 1.29. *In practice, one iteratively multiplies v by M and normalizes the resulting vector to have unit norm after each step to assess convergence. See Example 1.16.*

REMARK 1.30. *If the initial vector v is chosen such that $c_1 = 0$, then the algorithm will fail.*

REMARK 1.31. *Even though the power method involves taking increasing powers of matrix M, each iteration step involves only one matrix-vector multiplication, instead of matrix-matrix multiplications which can be much more CPU and memory intensive for large matrices.*

Even though Theorem 1.28 states a sufficient condition, it turns out that it is actually not necessary for a matrix to be diagonalizable in order for the algorithm to work. For instance, consider the defective matrix

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

which has an eigenvalue of 1 with multiplicity 2, but has only one eigenvector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. The number of times an eigenvalue appears as a root of the characteristic polynomial is called the algebraic multiplicity of the eigenvalue, while the dimension of the eigenspace of an eigenvalue is called the geometric multiplicity. The above matrix thus has 1 as an eigenvalue with algebraic multiplicity 2, but geometric multiplicity 1. Even though it has only one eigenvector, the k-th power of M is

$$M^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix};$$

hence, for any nonzero vector $v = \begin{pmatrix} a \\ b \end{pmatrix}$, we have

$$M^k v = \begin{pmatrix} a + kb \\ b \end{pmatrix}.$$

If $b = 0$, then v is already an eigenvector of M. If $b \neq 0$, then as $k \to \infty$,

$$\frac{b}{a + kb} = \frac{1}{a/b + k} \to 0,$$

and the power iteration will always converge to an eigenvector proportional to $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. In general, the power method will work for a real matrix with real eigenvalues when there is a unique dominant eigenvalue, which may have algebraic multiplicity greater than 1.

Example 1.16 (Google's PageRank Algorithm)**.** *Google's page rank algorithm published in 1999 finds the dominant eigenvector of a stochastic matrix M, whose columns are transition probability vectors on a Markov chain, by using the power method. This dominant vector is the stationary distribution of M, ranking the considered webpages by the corresponding entry values that represent the probability that each website will be visited by users in the world.*

In greater detail, suppose there are n websites to be modeled. Let ℓ_j denote the number of outgoing links that the j-th website contains. Then, the transition probability of going from

j-th website to i-th website is assumed to be

$$L_{ij} = \begin{cases} \frac{1}{\ell_j}, & \textit{if the i-th website is linked from the j-th website} \\ 0, & \textit{otherwise} \end{cases}.$$

The resulting matrix $L = (L_{ij})$ is a column-stochastic matrix, i.e. the entries are all non-negative and the column-sum of each column is 1. As is, L is not guaranteed to have a single dominant eigenvalue. However, we can turn L into an irreducible aperiodic matrix M, for which the Perron-Frobenius theorem guarantees the existence of a unique dominant eigenvalue with multiplicity 1. This step involves mixing each column of L with a random jumping probability distribution

$$q = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix},$$

where q_i represents some kind of unbiased probability that a web surfer will jump to the i-th website; e.g. $q_i = 1/n$.

The j-th columns of the final PageRank matrix M is

$$M_{:j} = \begin{cases} q, & \textit{if } L_{:j} \textit{ is a 0 vector} \\ (1-\pi)L_{:j} + \pi q, & \textit{otherwise} \end{cases}$$

where π, $0 < \pi < 1$, is a small mixing coefficient. Once an appropriate q is chosen, the Perron-Frobenius theorem guarantees that M has a unique dominant eigenvalue of 1 with multiplicity 1, and a corresponding eigenvector can be chosen to consist of all positive entries. One can thus normalize the dominant eigenvector $p = (p_1, \ldots, p_n)^t$, $p_i > 0\ \forall i$, such that

$$\|p\|_1 \equiv \sum_{i=1}^{n} p_i = 1.$$

We can interpret p has a probability distribution of site visit over the set of n websites, and rank the websites in decreasing order of p_i values.

To calculate the dominant vector p, one applies the power method as follows

---

PageRank Pseudocode

```
Given:  PageRank matrix M ∈ ℝ^{n×n}.
        N_max = maximum number of allowed iterations
        ϵ = stopping criterion

function PageRank_Power_Method(M, N_max, ϵ):
        Initialize p_0 ∈ ℝ^n_{>0}, such that ‖p_0‖_1 = 1
        For k = 1, ..., N_max:
```

$z = Mp_{k-1}$
$p_k = z/\|z\|_1$
`If` $\|p_k - p_{k-1}\|_1 < \epsilon$: `break`
Return p_k.

1.14 Orthogonal Iteration and QR Iteration

Let $M \in \mathbb{R}^{n\times n}$ have eigenvalues $\lambda_1, \ldots, \lambda_n$, ordered by their absolute values:

$$|\lambda_1| \geq \cdots \geq |\lambda_\ell| > |\lambda_{\ell+1}| \geq \cdots \geq |\lambda_n|,$$

where we have assumed that $|\lambda_\ell| > |\lambda_{\ell+1}|$ to guarantee convergence in the ensuing discussion. The power method for computing the dominant eigenvalue and eigenvector can be generalized to yield an ℓ-dimensional dominant invariant subspace $S \subseteq \mathbb{R}^n$ of M. Corresponding dominant eigenvalues and eigenvectors can be retrieved by attempting to diagonalize M in this restricted invariant subspace S.

To gain intuition into the algorithm, let us suppose we are interested in the first two eigenvalues, assuming $|\lambda_1| > |\lambda_2| > |\lambda_3|$, and choose two random orthonormal vectors $p_0, q_0 \in \mathbb{R}^n$. As before, we will iteratively act on p_0 with M until $M^k p_0$ is roughly aligned with the first eigenvector. Likewise, if we just keep acting on q_0 iteratively with M, then $M^k q_0$ will also become aligned with the first eigenvector, losing information about the second eigenvalue. To avoid this redundancy and loss of information, after each operation Mq_k, we orthonormalize it w.r.t. Mp_k using the Gram-Schmidt process. If indeed $|\lambda_1| > |\lambda_2| > |\lambda_3|$, then the orthonormalized version of Mq_k will roughly reside in the 2-dimensional subspace spanned by the first two eigenvectors, since all other directions are suppressed by at least $|\lambda_3/\lambda_2|^k$. If $|\lambda_1| = |\lambda_2| > |\lambda_3|$, and the eigenspace corresponding to λ_1 and λ_2 is 2-dimensional, then the algorithm still converges to the target eigenspace, but it is possible that neither iteration vector individually converges to a fixed vector (Homework).

If $\lambda_1 = \lambda_2$ and the eigenspace is 1-dimensional, then as long as $|\lambda_2| > |\lambda_3|$, the algorithm will still converge to a 2-dimensional invariant subspace and the eigenvalue can be recovered; however, M will have only one eigenvector in this invariant subspace. For example, consider again

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

In this case, the first iteration vector will converge to $(\pm 1, 0)$, while the second iteration vector converges to $(0, \pm 1)$.

The general orthogonal iteration algorithm is below:

`Orthogonal Iteration Pseudocode`

`Given:` $M \in \mathbb{R}^{n\times n}$.
$N_{\max}$ `= maximum number of allowed iterations.`
ℓ `= dimension of the desired dominant invariant subspace of` M.

ϵ = stopping criterion.

function Orthogonal_Iteration_Method($M, N_{\max}, \ell,\ \epsilon$):
Initialize $Q_0 \in \mathbb{R}^{n\times\ell}$ with orthonormal columns
For $k = 1, \ldots, N_{\max}$:
$Z_k = MQ_{k-1}$
$Q_kR_k = Z_k$ (Gram-Schmidt or other thin QR factorization)
{Eigenvalues of $Q_k^tMQ_k$} $= \{\lambda_1^{(k)}, \ldots, \lambda_\ell^{(k)}\}$
If $\|Q_kQ_k^t - Q_{k-1}Q_{k-1}^t\| < \epsilon$: break
Return $Q_k, \{\lambda_1^{(k)}, \ldots, \lambda_\ell^{(k)}\}$

REMARK 1.32. *In the thin QR factorization $Z_k = Q_kR_k$, $Q_k \in \mathbb{R}^{n\times\ell}$ has orthonormal columns and $R_k \in \mathbb{R}^{\ell\times\ell}$ is upper triangular with positive diagonal entries. There are many numerical methods for computing this QR factorization, and the Gram-Schmidt process is just one of them.*

The next iteration we will discuss is called the QR iteration, and its theoretical formulation is in essence a slight modification of the orthogonal iteration; however, a major advantage of the QR iteration over the orthogonal iteration is that the QR iteration is accompanied by fast implementation algorithms.

To understand how the QR iteration works, let us consider the case $\ell = n$ and the orthogonal matrices $Q_k \in \mathbb{R}^{n\times n}$ obtained via the orthogonal iteration (see the pseudocode), and define

$$T_{k-1} = Q_{k-1}^tMQ_{k-1}.$$

We previously denoted MQ_{k-1} as Z_k, and then factorized it as $Z_k = Q_kR_k$. Thus, we can rewrite T_{k-1} as

$$T_{k-1} = Q_{k-1}^tQ_kR_k = (Q_{k-1}^tQ_k)R_k \equiv U_kR_k$$

Acting on the right-side of the equation $Z_k = Q_kR_k$ with Q_k^t, we get

$$Q_k^tZ_k = Q_k^tMQ_{k-1} = R_k$$

since the columns of Q_k are orthonormal. Hence,

$$T_k = Q_k^tMQ_k = (Q_k^tMQ_{k-1})Q_{k-1}^tQ_k = R_kU_k.$$

Hence, T_k is obtained from T_{k-1} by simply switching the order of the orthogonal matrix U_k and upper triangular matrix R_k. If we did not compute Q_{k+1} in the orthogonal iteration, but want to get T_{k+1} from T_k, then we just perform QR factorization of T_k and then switch the order of the orthogonal and upper triangular factors again. This observation thus suggests the following algorithm:

QR Iteration Pseudocode

Given: $M \in \mathbb{R}^{n\times n}$.

$N_{\max}$ = maximum number of allowed iterations.
ϵ = stopping criterion.

function QR_Iteration_Method($M, N_{\max}, \epsilon$):
Initialize an orthogonal matrix $U_0 \in \mathbb{R}^{n\times n}$
$T_0 = U_0^t M U_0$
For $k = 1, \ldots, N_{\max}$:
$T_{k-1} = U_k R_k$ (QR factorization)
$T_k = R_k U_k$
If $\|T_k - T_{k-1}\| < \epsilon$: break
Return $T_k, \{\text{Eigenvalues of } T_k\} = \{\lambda_1^{(k)}, \ldots, \lambda_n^{(k)}\}$

REMARK 1.33. *A key step in an efficient implementation of the QR iteration is to choose U_0 so that the only non-zero entries of T_0 are in the upper triangular and the first subdiagonal locations, yielding the so-called Hessenberg matrix form. There exist fast QR factorization methods for Hessenberg matrices.*

REMARK 1.34. *If the eigenvalues of M are all real, then T_k will converge to an upper triangular matrix with the eigenvalues along the diagonal, yielding the Schur decomposition. If some of the eigenvalues of M are complex, then T_k will converge to a block triangular matrix with the block diagonal terms being either a real eigenvalue of M or a 2×2 matrix having complex conjugate eigenvalues that are also eigenvalues of M.*

1.15 Vector Space of Matrices

An $m \times n$ matrix M can be viewed as an element in $\mathbb{R}^{mn}$, i.e. the entries of M provide the coordinates of M in $\mathbb{R}^{mn}$. It is easy to show that the space $\mathcal{M}_{m\times n}$ of all $m \times n$ matrices is a vector space under matrix addition and scalar multiplication. If we do not put any constraint on the matrices, e.g. being symmetric or orthogonal, then this vector space is isomorphic to $\mathbb{R}^{mn}$. We would like to endow this vector space with geometry by defining an inner product and also with topology by defining a norm.

Definition 1.38 (Frobenius Inner Product)**.** *The map $\langle \cdot, \cdot \rangle_F : \mathcal{M}_{m\times n} \times \mathcal{M}_{m\times n} \to \mathbb{R}$ defined by $\langle M, N \rangle_F = \operatorname{tr}(M^t N)$ is called the Frobenius inner product.*

EXERCISE 1.19. *Show that the Frobenius inner product is indeed an inner product on the vector space $\mathcal{M}_{m\times n}$.*

(N.B. In order for the Frobenius inner product to make sense, we need to assume that the matrices are represented in orthonormal bases.) It doesn't matter which orthonormal bases, because the inner product is invariant under any orthogonal transformation of domain and codomain bases. (Check this statement.)

1.15.1 Frobenius Norm

As seen before, any inner product yields a norm:

Definition 1.39 (Frobenius Norm). *The map* $\|\cdot\|_F : \mathcal{M}_{m\times n} \to \mathbb{R}$ *defined by* $\|M\|_F = \sqrt{\langle M, M\rangle_F} = \sqrt{\mathrm{tr}(M^tM)}$ *is called the Frobenius norm.*

It is sometimes useful to think of the squared Frobenius norm of $M \in \mathcal{M}_{m\times n}$ as being the sum of squared ℓ_2-norm of all rows or columns of M. That is,

$$\|M\|_F^2 = \sum_{i=1}^{m} \|m_{i,:}\|_2^2 = \sum_{j=1}^{n} \|m_{:,j}\|_2^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} m_{ij}^2,$$

where $m_{i,:}$ and $m_{:,j}$ denote the i-th row and j-th column of M, respectively. Another way of understanding the Frobenius norm is to note that

$$\|M\|_F^2 = \mathrm{tr}(M^tM) = \sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k} \sigma_i^2,$$

where λ_i are the positive eigenvalues of M^tM, $\sigma_i = \sqrt{\lambda_i}$ are the singular values of M and k is the rank of M.

REMARK 1.35. *Note that for any* $M \in \mathbb{R}^{m\times n}$ *the Frobenius norm is invariant under* $M \mapsto U_1MU_2$*, where* $U_1 \in O(m)$ *and* $U_2 \in O(n)$ *are any orthogonal matrices. The Frobenius norm is thus said to be unitarily invariant.*

There are other ways of defining a norm on $\mathcal{M}_{m\times n}$, as we will see shortly.

1.15.2 Rank-1 Matrix Approximation

Given a matrix $M \in \mathcal{M}_{m\times n}$, which may be contaminated by noise that artificially increases the rank, we often want to obtain a simplified matrix A that captures bulk of the structural properties of M. One way of phrasing this problem is to find $A \in \mathcal{M}_{m\times n}$ having some fixed small rank, such that $\|M - A\|_F$ is minimized. When we restrict A to be of rank one, this problem is equivalent to finding unit vectors $w \in \mathbb{R}^m, v \in \mathbb{R}^n$ and a positive number σ such that

$$\|M - \sigma wv^t\|_F$$

is minimized, which amounts to finding the best rank-1 SVD approximation. Similar to the power method for finding eigenvalues, described in Section 1.13, we have an iterative algorithm for finding the dominant singular value and corresponding singular vectors:

`Rank-1 SVD Power Method Pseudocode`

`Given` $M \in \mathcal{M}_{m\times n}$.
function `PowerMethod(`M`):`
    `Initialize a unit vector` $v \in \mathbb{R}^n$
    While `stopping criteria and maximum iteration not reached:`

$w := Mv/\|Mv\|$
$v := M^t w/\|M^t w\|$
$\sigma := w^t Mv$
Return σ, w, v.

REMARK 1.36. *This iterative algorithm will fail when the initial guess v is orthogonal to the top right singular vector. This issue can be easily addressed by trying multiple runs with different randomized initial v.*

REMARK 1.37. *Stopping criteria may involve assessing the changes in σ and the singular vectors and checking whether these changes are sufficiently small.*

REMARK 1.38. *We will encounter a generalization of this iterative method when we discuss the CP decomposition of tensors later in the course.*

1.15.3 Operator Norm

Definition 1.40 (Operator Norm)**.** *Let $M : V \to W$ be a linear operator, where V and W are normed vector spaces with norms $\|\cdot\|_V$ and $\|\cdot\|_W$, respectively. The induced operator norm of M is*

$$\|M\|_{V,W} = \sup_{v \in V \setminus \{0\}} \frac{\|Mv\|_W}{\|v\|_V}.$$

When V and W are finite dimensional, the operator norm is also called the matrix norm. Note that the operator norm depends on the norms on the domain V and codomain W. We make this dependence explicit by indexing the operator norm with V, W. This definition immediately implies that

$$\boxed{\|Mv\|_W \leq \|M\|_{V,W}\|v\|_V, \, \forall v \in V.}$$

Definition 1.41 (Bounded Operator)**.** *A linear operator $M : V \to W$ from normed vector space V to normed vector space W is called bounded if $\|M\|_{V,W} < \infty$. Equivalently, M is bounded if there exists a finite constant $C \geq 0$ such that $\forall v \in V, \|Mv\|_W \leq C\|v\|_V$. The operator norm $\|M\|_{V,W}$ is the infimum of these constants C. Yet another equivalent definition is that M is bounded if it maps the unit sphere in V to a bounded set in W.*

We have previously seen that a linear map is uniformly continuous iff it is continuous at a point. We will prove here that a linear map is uniformly continuous iff it is bounded.

Theorem 1.29. *Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed vector spaces, and let $M : V \to W$ be a linear map. Then, the following statements are equivalent:*

1. *M is bounded.*
2. *M is uniformly continuous on V.*
3. *M is continuous on V.*

4. *M is continuous at a point in V.*

Proof. (1⇒2) If M is bounded, then $\forall x, y \in V$, $\|M(x-y)\|_W \leq \|M\|_{V,W}\|x-y\|_V$, where $\|M\|_{V,W} < \infty$. Hence, M is Lipschitz continuous and thus uniformly continuous.
(2⇒3, 3⇒4) By definition. See also Theorem 1.6.
(4⇒1) Assume M is continuous at $x \in V$. Then, there exists a constant $\delta > 0$ such that

$$\|y - x\|_V \leq \delta \Rightarrow \|M(y-x)\|_W \leq 1.$$

For any $v \in V \setminus \{0\}$,

$$\|M(v)\|_W = \|v\|_V \|M(v/\|v\|_V)\|_W = \frac{\|v\|_V}{\delta}\|M(v\delta/\|v\|_V)\|_W.$$

But, since $\|v\delta/\|v\|_V\|_V \leq \delta$, we have $\|M(v\delta/\|v\|_V)\|_W \leq 1$ and, thus,

$$\|M(v)\|_W \leq \delta^{-1}\|v\|_V. \tag{1.15}$$

Since $\|M\|_{V,W}$ provides the smallest such bound (1.15), $\|M\|_{V,W} \leq \delta^{-1} \leq \infty$. □

EXERCISE 1.20. *Let V be an inner product space. Use the Cauchy-Schwarz inequality, $|\langle\phi,\psi\rangle| \leq \|\phi\|\|\psi\|$, to show that $\langle\phi,\cdot\rangle : V \to \mathbb{C}$ and $\phi\langle\psi,\cdot\rangle : V \to V$ are bounded and, thus, continuous linear maps. In quantum mechanics, these maps are usually written as $\langle\phi|$ and $|\phi\rangle\langle\psi|$.*

You will show in Problem Set 1 that any finite dimensional matrix is bounded and, thus, continuous in ℓ_2-norm of V and W. An operator acting on an infinite dimensional Hilbert space, however, is not necessarily bounded. In particular, many operators in quantum mechanics, such as the position and momentum operators are unbounded. This distinction will play a critical role in defining a reproducing kernel Hilbert space (RKHS) later.

For now, let us familiarize ourselves with the operator norm by choosing some concrete norms on finite dimensional V and W and examining the induced operator norm. The ℓ_p-norms on V and W induce a natural matrix norm on the space of matrices. If the choice of p is the same for V and W, then we have

Definition 1.42 (Matrix ℓ_p-norm)**.** *Let V and W be vector spaces with ℓ_p norms, and let $M : V \to W$ be a linear map. Then, the matrix ℓ_p-norm of M is defined as*

$$\|M\|_p = \sup_{v \in V\setminus\{0\}} \frac{\|Mv\|_p}{\|v\|_p} = \sup_{v \in V, \|v\|_p = 1} \|Mv\|_p.$$

For some values of p, we can explicitly evaluate the norm:

1. ℓ_1-norm:

$$\|M\|_1 = \sup_{v, \|v\|_1=1} \|Mv\|_1 = \sup_{v, \|v\|_1=1} \sum_i \left|\sum_j m_{ij}v_j\right| \leq \sup_{v, \|v\|_1=1} \sum_i \left(\sum_j |m_{ij}||v_j|\right) \tag{1.16}$$

where we have used the triangle inequality. But,

$$\sup_{v,\|v\|_1=1} \sum_i \left(\sum_j |m_{ij}||v_j| \right) = \sup_{v,\|v\|_1=1} \sum_j \left(\sum_i |m_{ij}| \right) |v_j| = \sup_{v,\|v\|_1=1} \sum_j \|m_{:,j}\|_1 |v_j|.$$

Under the constraint $\|v\|_1 = 1$, we have

$$\sum_j \|m_{:,j}\|_1 |v_j| \leq \max_k \{\|m_{:,k}\|_1\} \sum_j |v_j| = \max_k \{\|m_{:,k}\|_1\},$$

with the inequality saturated when $v_j = 1$ for $j = \arg\max_k \{\|m_{:,k}\|_1\}$ and $v_j = 0$ otherwise. With this choice of v, the inequality in (1.16) is also saturated. To summarize, we have shown that

$$\boxed{\|M\|_1 = \max_j \{\|m_{:,j}\|_1\}}.$$

2. ℓ_2-norm (a.k.a. the Spectral Norm): Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$ denote the singular values of M. Like the Frobenius norm, the ℓ_2-norm $\|M\|_2$ is unitarily invariant under the transformation $M \mapsto U_1 M U_2$, where U_1 and U_2 are any orthogonal matrices that match the row and column dimensions of M. Choosing the columns of U_2 to be the orthonormal eigenvectors of $M^t M$ shows that

$$\boxed{\|M\|_2 = \sigma_1}.$$

Note that since $\|M\|_F = \sqrt{\sum_i \sigma_i^2}$, we always have $\|M\|_2 \leq \|M\|_F$. Note also that for a symmetric matrix M, $\|M\|_2 = \rho(M)$ (see Section 1.10.1).

3. ℓ_∞-norm:

$$\|M\|_\infty = \sup_{v,\|v\|_\infty=1} \|Mv\|_\infty = \sup_{v,\|v\|_\infty=1} \max_i \left\{ \left| \sum_j m_{ij} v_j \right| \right\}.$$

By the triangle inequality, we have

$$\left| \sum_j m_{ij} v_j \right| \leq \sum_j |m_{ij}||v_j| \leq \|m_{i,:}\|_1,$$

where the last inequality follows from the assumption that each $|v_j| \leq 1$. Hence,

$$\|M\|_\infty \leq \max_i \{\|m_{i,:}\|_1\}. \tag{1.17}$$

Note that (1.17) is saturated if we choose $v_j = \text{sign}(m_{kj})$, where $k = \arg\max_i \{\|m_{i,:}\|_1\}$. Thus,

$$\boxed{\|M\|_\infty = \max_i \{\|m_{i,:}\|_1\}}.$$

Example 1.17. *Let* $M = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. *Then,* $\rho(M) = \|M\|_1 = \|M\|_2 = \|M\|_\infty = 3$ *and* $\|M\|_F = \sqrt{10}$.

Example 1.18. *Let* $M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. *Then,* $\|M\|_1 = \|M\|_\infty = 2$, $\|M\|_2 = \sqrt{\frac{3+\sqrt{5}}{2}}$ *and* $\|M\|_F = \sqrt{3}$. *The spectral radius is* $\rho(M) = 1$. *Note that* $\rho(M) < \|M\|_2 < \|M\|_F < \|M\|_{1,\infty}$.

Example 1.19. *Let* $M = \begin{pmatrix} 3 & -1 \\ 0 & 1 \end{pmatrix}$. *Then,* $\|M\|_1 = 3$, $\|M\|_2 = \sqrt{\frac{11+\sqrt{85}}{2}}$, $\|M\|_\infty = 4$, *and* $\|M\|_F = \sqrt{11}$. *In this case,* $\rho(M) = \|M\|_1 < \|M\|_2 < \|M\|_F < \|M\|_\infty$.

In all these examples, we see that $\rho(M) \leq \|M\|$ for any of the matrix norm. This inequality holds in general, as the following theorem shows:

Theorem 1.30. *Let* $M : \mathbb{C}^n \to \mathbb{C}^n$ *be a linear map. Given any matrix norm* $\|M\|$, *the spectral radius of* M *satisfies*

$$\rho(M) \leq \|M^k\|^{\frac{1}{k}},$$

for any positive integer k.

Proof. Let v be an eigenvector of M corresponding to the eigenvalue λ with the largest absolute value. Then, $M^k v = \lambda^k v$. Thus,

$$\|M^k v\| = |\lambda|^k \|v\| \leq \|M^k\| \|v\|.$$

Since $\|v\| \neq 0$, we have $|\lambda| \leq \|M^k\|^{1/k}$. □

Thus, the spectral radius provides the lower bound of any matrix norm. An interesting related fact proven by Gelfand is

Theorem 1.31 (Gelfand's Spectral Radius Theorem)**.** *Let* $M : \mathbb{C}^n \to \mathbb{C}^n$ *be a linear map. For any matrix norm* $\|\cdot\|$,

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

EXERCISE 1.21. *Let* $|\lambda_{\min}|$ *and* $|\lambda_{\max}|$ *be the smallest and the largest absolute values of the eigenvalues of* $M \in \mathbb{C}^{n\times n}$, *respectively. Show that*

$$\sigma_{\min} \leq |\lambda_{\min}| \leq |\lambda_{\max}| \leq \sigma_{\max}$$

where $\sigma_{\min}$ *and* $\sigma_{\max}$ *are the smallest and the largest singular values of* M, *respectively. Here, we consider all singular values, and* $\sigma_{\min}$ *may be 0. Thus, we might have the situation*

$$\max_{i,j} \frac{|\lambda_i|}{|\lambda_j|} \ll \kappa_2(M)$$

where $\kappa_2 \equiv \sigma_{\max}/\sigma_{\min}$ is the condition number of M w.r.t. the ℓ_2-norm. Compared to singular values, eigenvalues may thus be less useful for studying the sensitivity of matrix equation solutions to perturbation (see Section A.3).

EXERCISE 1.22. *In Section 1.12, we learned that we can measure the distance between m-dimensional subspaces S_1 and S_2 in $\mathbb{R}^n$, respectively defined by two orthogonal projection operators P_1 and P_2, as*

$$d(S_1, S_2) = \|P_1 - P_2\|,$$

where $\|P_1 - P_2\|$ is some unspecified norm on the space of matrices. Two examples of widely used matrix norm for this purpose are $\frac{1}{\sqrt{2}}\|P_1 - P_2\|_F$ and $\|P_1 - P_2\|_2$. Show that

$$\frac{1}{\sqrt{2}}\|P_1 - P_2\|_F = \sqrt{m - \mathrm{tr}(P_1 P_2)}.$$

Let $U \equiv (U_1\ U_2) \in \mathbb{R}^{n\times n}$ be an orthogonal matrix where the columns of $U_1 \in \mathbb{R}^{n\times m}$ form an orthonormal basis of S_1 and the columns of $U_2 \in \mathbb{R}^{n\times(n-m)}$ form an orthonormal basis of $S_1^\perp$. Similarly, let $W \equiv (W_1\ W_2) \in \mathbb{R}^{n\times n}$ be an orthogonal matrix where the columns of $W_1 \in \mathbb{R}^{n\times m}$ form an orthonormal basis of S_2 and the columns of $W_2 \in \mathbb{R}^{n\times(n-m)}$ form an orthonormal basis of $S_2^\perp$. Show that

$$\|P_1 - P_2\|_2 = \|U_1^t W_2\|_2 = \|U_2^t W_1\|_2.$$

1.15.4 Equivalence of Norms and Continuity of Linear Maps in Finite Dimensions

A natural question arises at this point regarding which norm to use to define topology on a vector space. In finite dimensions, it turns out all norms are equivalent in the following sense:

Definition 1.43 (Equivalence of Norms)**.** *Two norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ on a vector space V are said to be equivalent if there exist constants $c > 0$ and $C > 0$ such that*

$$c\|v\|_\alpha \le \|v\|_\beta \le C\|v\|_\alpha,\ \forall v \in V.$$

EXERCISE 1.23. *Show that the above property indeed defines an equivalence relation.*

EXERCISE 1.24. *Show that if a linear map $M : V \to W$ is bounded and, thus, continuous in one set of norms on V and W, then it is also bounded and continuous in any other equivalent set of norms.*

In finite dimensions, but not in infinite dimensions, the following theorem holds:

Theorem 1.32. *All norms on a finite dimensional vector space are equivalent and define the same topology and the same notion of convergence of sequences.*

Corollary 1.12. *A linear map $M : V \to W$ of finite dimensional vector spaces V and W is always continuous and bounded.*

Proof. In Problem Set 1, you will show that M is continuous using the ℓ_2-norm on V and W. The equivalence of all norms on V and W then implies that for any norms $\|\cdot\|_V$ and $\|\cdot\|_W$, there exist constants $C > 0$ and $C' > 0$ such that

$$\|Mv\|_W \leq C\|Mv\|_2 \leq C\|M\|_2\|v\|_2 \leq CC'\|M\|_2\|v\|_V, \forall v \in V.$$

Since $\|M\|_{V,W} \leq CC'\|M\|_2 < \infty$, M is bounded and, thus, continuous with respect to $\|\cdot\|_V$ and $\|\cdot\|_W$. □

In fact, only the domain needs to be finite dimensional in order for the corollary to hold:

Corollary 1.13. *A linear map $M : V \to W$ on a finite dimensional vector space V is always continuous and bounded.*

Proof. Let $\{e_1, \ldots, e_n\}$ be the standard basis of V. For any norm $\|\cdot\|_W$ on W, we have by triangle inequality

$$\|Mv\|_W = \|M(\sum_{i=1}^{n} v_i e_i)\|_W \leq \sum_{i=1}^{n} |v_i|\|Me_i\|_W \leq A\|v\|_1,$$

where $A = \max_i \|Me_i\|_W$ is finite, because it is the maximum of a finite set of numbers. The equivalence of any other norm on V with the ℓ_1 norm now implies that M is bounded and thus Lipschitz continuous. □

Thus, linear maps on finite dimensional vector spaces are always continuous and bounded with respect to all norms. This is perhaps the reason why a typical linear algebra course does not teach these concepts. Importantly, these nice properties of linear maps in finite dimensions will no longer automatically hold in infinite dimensions. In fact, an arbitrary linear map on an infinite dimensional vector space is in general not continuous and, thus, not bounded. In certain Hilbert spaces, however, we will shortly see that

- the presence of a particular bounded linear functional will allow us to define the so-called reproducing kernel, which in turn will provide us with a feature map from data space X into a reproducing kernel Hilbert space (RKHS) of functions on X;
- the reproducing kernel crucially depends on the inner product; that is, it arises as the inner product "representer" of the bounded linear functional mentioned above.