Mathematical Foundations of Machine Learning

Week 2: Spectral Theory

Ayberk Zeytin

Outline

1. Recap — Linear Algebra Basics
2. Cauchy-Schwarz Inequality
3. Gram-Schmidt Process
4. Norms and Dual Norms
5. Eigenvalues and eigenvectors
6. Diagonalization
7. Spectral theorem
8. Geometric interpretation
9. Schur decomposition

Recap — Linear Algebra Basics

Last week, we established the mathematical foundations for machine learning by revisiting the essential concepts of linear algebra: vector spaces, inner products, norms, orthogonality, and projections. These tools form the backbone of the geometry of data and algorithms.

Vector Spaces

A *vector space* V over $\mathbb{R}$ is a set with two operations:

- vector addition $u + v$,
- scalar multiplication αv for $\alpha \in \mathbb{R}$,

satisfying the familiar axioms (closure, associativity, commutativity, distributivity, neutral element, inverse element).

Examples.

1. The Euclidean space $\mathbb{R}^n$ with coordinate-wise addition and scalar multiplication. For instance, $x = (x_1, \dots, x_n)^\top$ is a typical element.
2. The set of polynomials of degree $\leq d$, written as $p(t) = a_0 + a_1 t + \cdots + a_d t^d$, identified with the vector $(a_0, a_1, \dots, a_d)^\top$.

Bases and Dimension

A *basis* of V is a linearly independent spanning set. All bases of a finite-dimensional vector space have the same cardinality, called the *dimension*.

$$\dim \mathbb{R}^n = n, \quad \dim\{p(t) : \deg(p) \leq d\} = d + 1.$$

Inner Product Spaces

An *inner product* on V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ satisfying:

$$\langle x, y\rangle = \langle y, x\rangle, \quad \langle \alpha x + y, z\rangle = \alpha\langle x, z\rangle + \langle y, z\rangle, \quad \langle x, x\rangle \geq 0,\ \langle x, x\rangle = 0 \iff x = 0.$$

Examples.

- In $\mathbb{R}^n$: $\langle x, y\rangle = \sum_{i=1}^n x_i y_i = x^\top y$.
- In the function space $C([a,b])$: $\langle f, g\rangle = \int_a^b f(x)g(x)\,dx$.

Norms and Length

The *norm* of $v \in V$ is defined by $\|v\| = \sqrt{\langle v, v\rangle}$. This generalizes the notion of length.

Examples.

- In $\mathbb{R}^4$, for $v = (1,2,3,4)^\top$: $\|v\| = \sqrt{1^2 + 2^2 + 3^2 + 4^2} = \sqrt{30}$.
- In $C([0,1])$, for $f(x) = x$: $\|f\| = \sqrt{\int_0^1 x^2 dx} = \sqrt{\frac{1}{3}}$.

Angles and Orthogonality

Given two nonzero vectors u, v, the angle θ between them is defined by

$$\cos(\theta) = \frac{\langle u, v\rangle}{\|u\|\|v\|}.$$

Examples.

- In $\mathbb{R}^4$, for $u = (1,0,1,0)^\top$ and $v = (1,1,0,1)^\top$, we obtain $\theta = \pi/4$.
- In $C([0,1])$, for $f(x) = x$, $g(x) = 1$, one computes $\theta = \pi/6$.

Two vectors are *orthogonal* if $\langle u, v\rangle = 0$. Examples include $u = (1,2,3,4)^\top$ and $v = (2,-1,-2,1)^\top$ in $\mathbb{R}^4$, and $\sin(x)$ and $\cos(x)$ on $[-\pi,\pi]$.

Projections

The projection of v onto u is given by

$$\text{proj}_u v = \frac{\langle v, u\rangle}{\langle u, u\rangle} u.$$

Examples.

- Projecting $v = (1,2,3,4)^\top$ onto $u = (1,0,0,0)^\top$ yields $(1,0,0,0)^\top$.
- In $C([0,\pi])$, projecting $g(x) = x$ onto $f(x) = \sin(x)$ gives $2\sin(x)$.

Key Facts

- **Pythagorean Theorem:** If $u \perp v$, then $\|u+v\|^2 = \|u\|^2 + \|v\|^2$. Example: $f(x) = \sin(x)$ and $g(x) = \cos(x)$ on $[-\pi,\pi]$.
- **Triangle Inequality:** $\|u+v\| \le \|u\| + \|v\|$. Example: in $\mathbb{R}^2$, $\|(1,0) + (0,1)\| = \sqrt{2} \le 2$.

1 Cauchy-Schwarz Inequality

Definition 1.1. We say that a vector space V is an *inner product space* if it admits an inner product $\langle\cdot,\cdot\rangle$.

Example 1.1. ►The pair $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ where $\langle x, y\rangle := x^\top y$, is an inner product space.
►For any $a, b \in \mathbb{R}$ with $a < b$, the pair $(C([a,b]), \langle\cdot,\cdot\rangle)$; where $\langle f, g\rangle := \int_a^b f(x)g(x)\,dx$ is an inner product space.

Theorem 1.1 (Cauchy–Schwarz)**.** *For all x, y in an inner product space $(V, \langle\cdot,\cdot\rangle)$,*

$$|\langle x, y\rangle| \le \|x\| \cdot \|y\|.$$

Moreover, equality holds if and only if x and y are linearly dependent.

Proof. If $y = 0$, the statement is immediate. Assume $y \neq 0$. For any $\lambda \in \mathbb{R}$,

$$\begin{aligned} 0 \leq \|x - \lambda y\|^2 &= \langle x - \lambda y,\, x - \lambda y\rangle \\ &= \langle x, x\rangle - \lambda\langle y, x\rangle - \lambda\langle x, y\rangle + \lambda^2\langle y, y\rangle \\ &= \langle x, x\rangle - 2\lambda\langle x, y\rangle + \lambda^2\langle y, y\rangle. \end{aligned}$$

Choose

$$\lambda = \frac{\langle y, x\rangle}{\|y\|^2} \quad \text{(note that } \|y\|^2 = \langle y, y\rangle > 0, \text{ as } y \neq 0).$$

Substituting this choice gives

$$0 \leq \|x\|^2 - \frac{|\langle y, x\rangle|^2}{\|y\|^2} = \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2},$$

whence

$$|\langle x, y\rangle|^2 \leq \|x\|^2\,\|y\|^2,$$

and taking square roots yields the inequality.

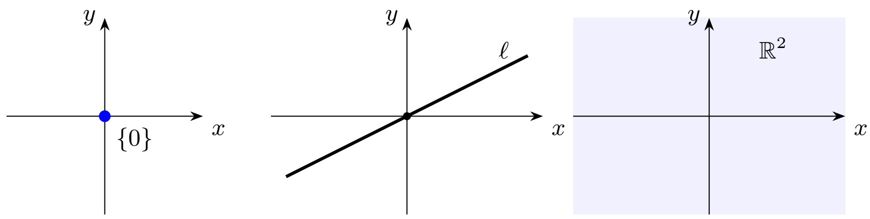
For the equality case, observe that the inequality above becomes an equality if and only if $\|x - \lambda y\|^2 = 0$ for $\lambda = \langle y, x\rangle/\|y\|^2$, i.e., if and only if $x = \lambda y$. Thus equality holds exactly when x and y are linearly dependent. $\square$

Definition 1.2. Let V be a real vector space. A subspace of W of V is a non-empty subset of V which itself is a vector space with respect to the addition and scalar multiplication inherited from V[1].

[1] When a structure in mathematics is defined - such as a group, ring, vector space,... - a substructure is defined in this manner; i.e. a subset which itself carries the same structure.

Example 1.2. Subspaces of $\mathbb{R}^2$ are

- $\{0\}$
- *lines* through the origin $(0,0)^\top$
- $\mathbb{R}^2$ itself



(i) $\{0\} \subset \mathbb{R}^2$ (ii) line through $(0,0)^\top$ (iii) $\mathbb{R}^2$

Figure 1: All linear subspaces of $\mathbb{R}^2$: the zero subspace, any line through the origin, and the whole space.

nasıl line`lar?

Exercise 1.1. What are subspaces of $\mathbb{R}^3$ and $\mathbb{R}^4$?

Definition 1.3 (Span)**.** Let V be a vector space over $\mathbb{R}$ and let $S \subseteq V$. The *span* of S, denoted $\operatorname{span}(S)$, is the set of all finite linear combinations of vectors from S:

$$\operatorname{span}(S) = \left\{ \sum_{i=1}^{k} \alpha_i v_i \;\middle|\; k \in \mathbb{N},\ v_i \in S,\ \alpha_i \in \mathbb{R} \right\}.$$

Equivalently, $\operatorname{span}(S)$ is the smallest subspace of V that contains S.

Example 1.3. Let

$$S = \{e_1 = (1,0,0,0)^\top,\ e_2 = (0,1,0,0)^\top,\ u = (1,1,0,0)^\top\} \subset \mathbb{R}^4.$$

Then

$$\operatorname{span}(S) = \left\{\alpha e_1 + \beta e_2 + \gamma u : \alpha, \beta, \gamma \in \mathbb{R}\right\} = \{(x_1, x_2, 0, 0)^\top : x_1, x_2 \in \mathbb{R}\}.$$

Thus $\operatorname{span}(S)$ is the 2-dimensional subspace of $\mathbb{R}^4$ consisting of all vectors with third and fourth coordinates zero, i.e. the (x_1, x_2)-plane inside $\mathbb{R}^4$.

Example 1.4. Let $V = C([0,1])$ be the vector space of continuous real-valued functions on $[0,1]$, and let

$$S = \{1,\ t,\ t^2\}.$$

Then

$$\operatorname{span}(S) = \{\alpha + \beta t + \gamma t^2 : \alpha, \beta, \gamma \in \mathbb{R}\},$$

the subspace of polynomials of degree at most 2. The set $\{1, t, t^2\}$ is a basis, so this subspace is 3-dimensional.

Definition 1.4. The *dimension* of a vector (sub)space W is defined as the cardinality of the smallest set $\mathcal{B}$ for which $W = \operatorname{span}(\mathcal{B})$. In this case, the set $\mathcal{B}$ is called a *spanning set* for W.

Example 1.5. Consider the subspace

$$W = \{(x, y, 0)^\top : x, y \in \mathbb{R}\} \subset \mathbb{R}^3.$$

A spanning set is

$$\mathcal{B} = \{(1,0,0)^\top, (0,1,0)^\top\}.$$

These two vectors are linearly independent and span W, so $\mathcal{B}$ is a basis of W. Therefore,

$$\dim(W) = 2.$$

Geometrically, W is the xy-plane inside $\mathbb{R}^3$.

Example 1.6. Let $V = C([0,1])$, the vector space of continuous real-valued functions on $[0,1]$. Consider the subspace

$$\mathcal{P} = \{p(t) : p \text{ is a polynomial in } t\}.$$

We know that $\{1, t, t^2, \ldots, t^n\}$ spans the polynomials of degree at most n. However, no finite set of functions can span all of $\mathcal{P}$, since for each n there are polynomials of higher degree. Hence $\mathcal{P}$ has an infinite basis, and therefore

$$\dim(\mathcal{P}) = \infty.$$

2 Gram-Schmidt Process

Definition 2.1 (Orthogonal complement)**.** For a subspace $W \subseteq V$, the **orthogonal complement** is $W^{\perp} = \{x : \langle x, w\rangle = 0 \ \forall w \in W\}$.

Example 2.1. Let $V = \mathbb{R}^2$ with the standard dot product. Consider a subspace W spanned by the vector $w_1 = (2, 1)^{\top}$. This subspace is a line passing through the origin.

To find the orthogonal complement $W^{\perp}$, we need to find all vectors $x = (x_1, x_2)^{\top}$ in $\mathbb{R}^2$ such that their dot product with any vector in W is zero. Since W is spanned by w_1, it is sufficient to ensure that $\langle x, w_1\rangle = 0$.

$$\langle (x_1, x_2)^{\top}, (2, 1)^{\top}\rangle = 2x_1 + 1x_2 = 0$$

This equation defines a line in $\mathbb{R}^2$ that passes through the origin and is perpendicular to the vector $(2, 1)^{\top}$. We can express this as $x_2 = -2x_1$.

Thus, $W^{\perp}$ is the subspace spanned by a vector satisfying this equation, for example, the vector $(1, -2)\top$.

$$W^{\perp} = \text{span}\{(1, -2)^{\top}\}$$

Geometrically, W is a line, and its orthogonal complement $W^{\perp}$ is the line perpendicular to it, also passing through the origin.

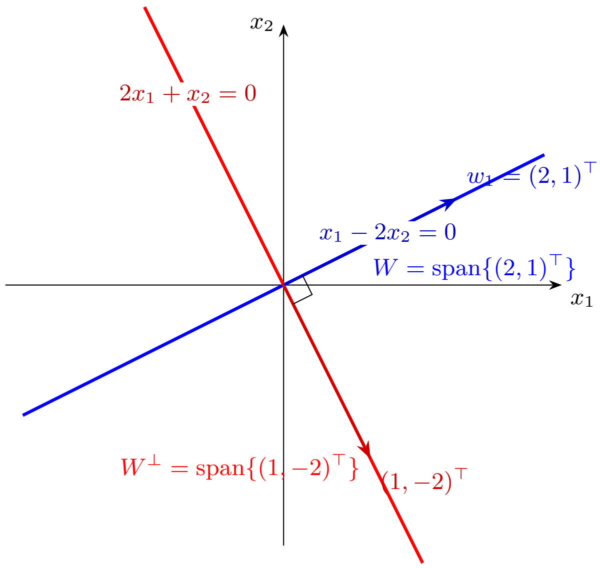


Figure 2: A subspace $W = \text{span}\{(2, 1)^{\top}\}$ and its orthogonal complement $W^{\perp} = \text{span}\{(1, -2)^{\top}\}$ in $\mathbb{R}^2$.

Example 2.2. Let $V = \mathbb{R}^3$ with the standard inner product. Consider a subspace W which is a plane defined by the equation $x_1 - 2x_2 + 3x_3 = 0$. This plane passes through the origin.

The normal vector to this plane is $n = (1, -2, 3)^{\top}$. The plane W consists of all vectors w that are orthogonal to this normal vector.

To find the orthogonal complement $W^{\perp}$, we need to find all vectors $x = (x_1, x_2, x_3)^{\top}$ that are orthogonal to every vector in W. By definition, this means $W^{\perp}$ is the set of all vectors orthogonal to the entire subspace W.

$$W^{\perp} = \{x \in \mathbb{R}^3 \mid \langle x, w\rangle = 0 \quad \forall w \in W\}$$

From the equation of the plane, we know that any vector x in $W^{\perp}$ must be parallel to the normal vector $n = (1, -2, 3)^{\top}$.

$$W^{\perp} = \text{span}\{(1, -2, 3)^{\top}\}$$

Geometrically, W is a plane through the origin, and its orthogonal complement $W^{\perp}$ is the line that is perpendicular to the plane and passes through the origin. This line is precisely the span of the plane's normal vector.

Theorem 2.1 (Projection Theorem)**.** *Let W be a finite-dimensional subspace of an inner product space V. For any $x \in V$ there exist unique $p \in W$ and $r \in W^{\perp}$ such that $x = p + r$. The vector p is the **orthogonal projection** of x onto W.*

Example 2.3. Let $V = \mathbb{R}^3$ with the standard dot product and let

$$W = \text{span}\{v_1, v_2\} \quad \text{where} \quad v_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \ v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

$x \notin W$

Given $x = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$, the orthogonal projection of x onto W can be computed via the normal equations or the matrix formula $p = \cancel{A(A^\top} A)^{-1}A^\top x$, where $A = [\,v_1\; v_2\,]$. A straightforward calculation yields

$$p = \begin{pmatrix} \frac{26}{9} \\ \frac{20}{9} \\ \frac{7}{9} \end{pmatrix}, \qquad r = x - p = \begin{pmatrix} \frac{1}{9} \\ -\frac{2}{9} \\ \frac{2}{9} \end{pmatrix}.$$

One checks that $r \perp v_1$ and $r \perp v_2$ (equivalently $A^\top r = \mathbf{0}$), hence $r \in W^\perp$ and $x = p + r$ with $p \in W$.

Geometric picture. The vector x decomposes into p (lying in the plane W) and r (perpendicular to W).

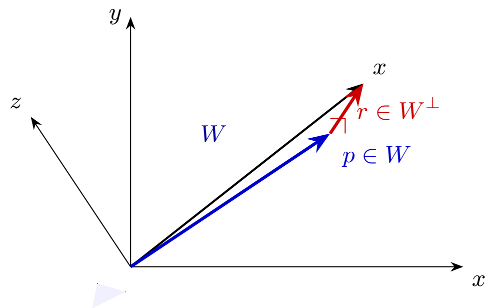


Figure 3: Decomposition $x = p + r$ with $p \in W$ and $r \perp W$ (schematic).

Example 2.4. Let $V = C([0,1])$ with the inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$, and let

$$W = \operatorname{span}\{1, t\}.$$

Given $x(t) = t^2$, the orthogonal projection $p(t) \in W$ has the form $p(t) = a + bt$ determined by

$$\boxed{\langle x - p,\, 1\rangle = 0, \qquad \langle x - p,\, t\rangle = 0,}$$

i.e. the normal equations

$$\begin{pmatrix} \langle 1,1\rangle & \langle 1,t\rangle \\ \langle t,1\rangle & \langle t,t\rangle \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \langle x,1\rangle \\ \langle x,t\rangle \end{pmatrix}.$$

Evaluating the integrals on $[0,1]$,

$$\langle 1,1\rangle = 1, \quad \langle 1,t\rangle = \frac{1}{2}, \quad \langle t,t\rangle = \frac{1}{3}, \quad \langle x,1\rangle = \frac{1}{3}, \quad \langle x,t\rangle = \frac{1}{4},$$

solving gives

$$a = -\frac{1}{6}, \qquad b = 1, \qquad \text{so} \quad p(t) = t - \frac{1}{6}.$$

Thus $x(t) = t^2 = p(t) + r(t)$ with $r(t) = t^2 - \left(t - \frac{1}{6}\right)$ satisfying $r \perp W$.

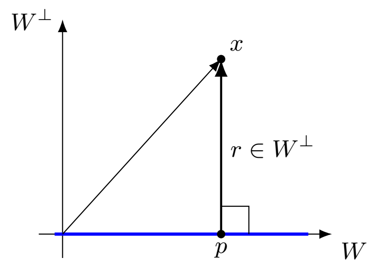


Figure 4: Geometric view of orthogonal projection.

Gram-Schmidt Process

The **Gram-Schmidt process** is an algorithm for constructing an **orthonormal basis** (or **orthogonal**) from an arbitrary basis (or simply a set of linearly independent vectors) for an inner product space, such as $\mathbb{R}^n$ with the standard dot product.

Indeed, given a basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ for a subspace W of $\mathbb{R}^n$, the process yields an **orthonormal basis** $\mathcal{O} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ for W, such that:

1. **Orthogonal:** $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for all $i \neq j$.
2. **Normal (Unit Length):** $||\mathbf{u}_i|| = 1$ for all i. — optional

Crucially, for each $j = 1, \dots, k$, the set $\{\mathbf{u}_1, \dots, \mathbf{u}_j\}$ is an orthonormal basis for the subspace spanned by $\{\mathbf{v}_1, \dots, \mathbf{v}_j\}$.

The process proceeds inductively. We first construct an orthogonal basis $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k\}$ and then normalize each vector to obtain the orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$.

Step 1: Constructing an Orthogonal Basis $\{\mathbf{w}_i\}$

w_1

$r = v_2 - f$

1. **First Vector:** The first vector is simply the first vector of the original set.

$$\mathbf{w}_1 = \mathbf{v}_1$$

2. **Second Vector:** The second vector is obtained by subtracting the component of $\mathbf{v}_2$ that lies in the direction of $\mathbf{w}_1$. This is the projection of $\mathbf{v}_2$ onto $\mathbf{w}_1$.

w_2

$$\mathbf{w}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{w}_1}\mathbf{v}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1$$

3. **Third Vector:** The third vector is $\mathbf{v}_3$ minus its projections onto the previously found orthogonal vectors $\mathbf{w}_1$ and $\mathbf{w}_2$.

$$\mathbf{w}_3 = \mathbf{v}_3 - \text{proj}_{\mathbf{w}_1}\mathbf{v}_3 - \text{proj}_{\mathbf{w}_2}\mathbf{v}_3 = \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1 - \frac{\mathbf{v}_3 \cdot \mathbf{w}_2}{\mathbf{w}_2 \cdot \mathbf{w}_2}\mathbf{w}_2$$

4. **General k-th Vector:** For the k-th vector, subtract its projections onto all previously found orthogonal vectors $\mathbf{w}_1, \dots, \mathbf{w}_{k-1}$.

$$\mathbf{w}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \text{proj}_{\mathbf{w}_j}\mathbf{v}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \frac{\mathbf{v}_k \cdot \mathbf{w}_j}{\mathbf{w}_j \cdot \mathbf{w}_j}\mathbf{w}_j$$

Step 2: Normalization Finally, normalize each orthogonal vector $\mathbf{w}_i$ to obtain the orthonormal basis vectors $\mathbf{u}_i$:

$$\mathbf{u}_i = \frac{1}{||\mathbf{w}_i||}\mathbf{w}_i \quad \text{for } i = 1, \dots, k$$

Example 2.5. Let $W = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ where $\mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$.

Step 1a: Find $\mathbf{w}_1$

$$\mathbf{w}_1 = \mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

Step 1b: Find $\mathbf{w}_2$

$$\begin{aligned}
\mathbf{w}_2 &= \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1 \\
\mathbf{v}_2 \cdot \mathbf{w}_1 &= (2)(3) + (2)(1) = 8 \\
\mathbf{w}_1 \cdot \mathbf{w}_1 &= (3)(3) + (1)(1) = 10 \\
\mathbf{w}_2 &= \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \frac{8}{10}\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \frac{4}{5}\begin{pmatrix} 3 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} 2 - 12/5 \\ 2 - 4/5 \end{pmatrix} = \begin{pmatrix} 10/5 - 12/5 \\ 10/5 - 4/5 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 6/5 \end{pmatrix}
\end{aligned}$$

To simplify the basis, we can use $\mathbf{w}_2' = 5\mathbf{w}_2 = \begin{pmatrix} -2 \\ 6 \end{pmatrix}$ (or even $\begin{pmatrix} -1 \\ 3 \end{pmatrix}$), as multiplying by a scalar preserves orthogonality. We will continue with $\mathbf{w}_2 = \begin{pmatrix} -2/5 \\ 6/5 \end{pmatrix}$ for correctness in the next step. The orthogonal basis is $\left\{\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} -2/5 \\ 6/5 \end{pmatrix}\right\}$.

Step 2: Normalization (Find $\mathbf{u}_1$ and $\mathbf{u}_2$)

$$||\mathbf{w}_1|| = \sqrt{3^2 + 1^2} = \sqrt{10}$$
$$\mathbf{u}_1 = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{pmatrix}$$

$$||\mathbf{w}_2|| = \sqrt{(-2/5)^2 + (6/5)^2} = \sqrt{4/25 + 36/25} = \sqrt{40/25} = \frac{2\sqrt{10}}{5}$$
$$\begin{aligned} \mathbf{u}_2 &= \frac{1}{2\sqrt{10}/5} \begin{pmatrix} -2/5 \\ 6/5 \end{pmatrix} = \frac{5}{2\sqrt{10}} \begin{pmatrix} -2/5 \\ 6/5 \end{pmatrix} \\ &= \begin{pmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \end{pmatrix} \end{aligned}$$

The orthonormal basis is $\mathcal{O} = \left\{ \begin{pmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{pmatrix}, \begin{pmatrix} -1/\sqrt{10} \\ 3/\sqrt{10} \end{pmatrix} \right\}$.

The process is a sequence of orthogonal projections and subtractions.

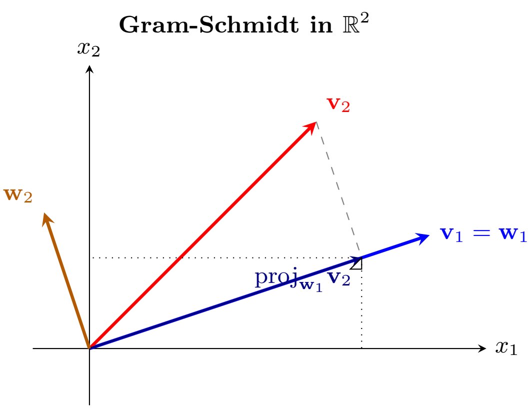


Figure 5: The geometry of finding $\mathbf{w}_2$ by subtracting the projection of $\mathbf{v}_2$ onto $\mathbf{w}_1$.

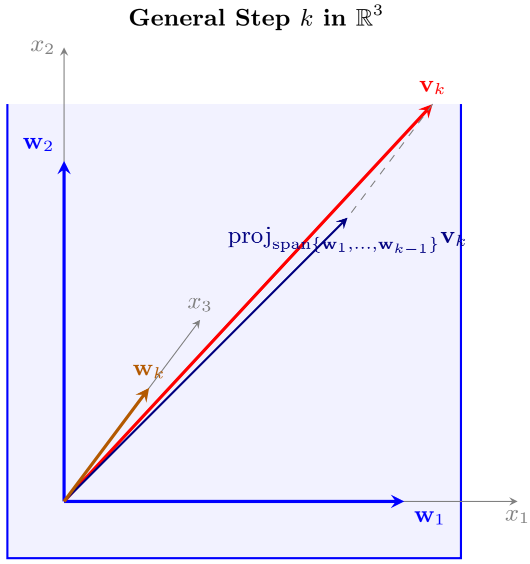


Figure 6: The general step for $\mathbf{w}_k$ in $\mathbb{R}^3$: $\mathbf{v}_k$ minus its projection onto the previously spanned subspace.

$||\lambda v|| = \lambda ||v||$

3 Norms and Dual Norms

Definition 3.1. A function $\|\cdot\| : \mathbb{R}^n \to [0, \infty)$ is a **norm** if it is positive-definite, absolutely homogeneous, and satisfies the triangle inequality.

Example 3.1 (Vector norms)**.** $\|x\|_1 = \sum_i |x_i|$, $\quad \|x\|_2 = (\sum_i x_i^2)^{1/2}$, $\quad \|x\|_\infty = \max_i |x_i|$.

Definition 3.2. The **dual norm** is $\|y\|_* = \sup_{\|x\| \le 1} \langle y, x \rangle$.

Theorem 3.1 (Hölder's inequality)**.** *For any $x, y \in \mathbb{R}^n$,*

$$|\langle x, y \rangle| \le \|x\| \cdot \|y\|_*.$$

Example 3.2 (Dual pairs)**.** (ℓ_1, ℓ_∞), $\quad (\ell_2, \ell_2)$. More generally, (ℓ_p, ℓ_q) with $\frac{1}{p} + \frac{1}{q} = 1$ for $1 < p < \infty$.

4 Eigenvalues and Eigenvectors

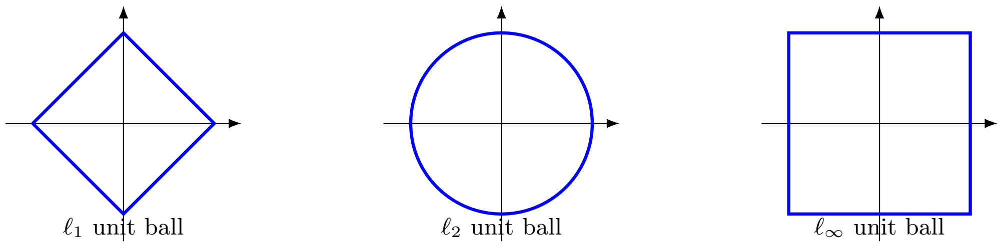
Definition 4.1. Let $A \in \mathbb{R}^{n \times n}$. A nonzero vector $v \in \mathbb{R}^n$ is an **eigenvector** of A if there exists $\lambda \in \mathbb{R}$ such that

$$Av = \lambda v.$$

The scalar λ is the corresponding **eigenvalue**.

Example 4.1. For $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$, the eigenvalues are $2, 3$ with eigenvectors $(1, 0)$ and $(0, 1)$.

Theorem 4.1. *The eigenvalues of a triangular matrix are precisely its diagonal entries.*



(a) ℓ_1 (b) ℓ_2 (c) ℓ_∞

Figure 7: Unit balls in $\mathbb{R}^2$ for common norms (geometry that drives regularization and robustness).

(a) Duality: $(\ell_1)^* = \ell_\infty$. (b) Duality: $(\ell_2)^* = \ell_2$.

Figure 8: Dual norms as polars of unit balls: $B^\circ_{\|\cdot\|} = B_{\|\cdot\|_*}$.

5 Diagonalization

Definition 5.1. A matrix A is **diagonalizable** if there exists an invertible P such that

$$P^{-1}AP = D$$

with D diagonal.

Theorem 5.1. *If A has n linearly independent eigenvectors, then A is diagonalizable.*

6 Spectral Theorem

Theorem 6.1 (Spectral Theorem)**.** *If A is a real symmetric matrix, then there exists an orthogonal matrix Q such that*

$$Q^\top AQ = \Lambda$$

where Λ is diagonal with the eigenvalues of A on its diagonal.

7 Geometric Interpretation

Eigenvectors are directions preserved by A, scaled by λ.

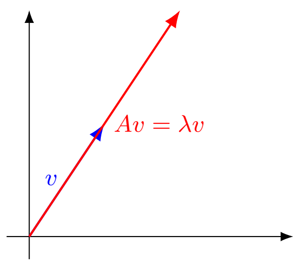


Figure 9: An eigenvector v retains its direction under A, only scaled by λ.

8 Schur Decomposition

Theorem 8.1 (Schur Decomposition)**.** *For any $A \in \mathbb{C}^{n\times n}$, there exists a unitary Q such that*

$$Q^*AQ = T$$

is upper triangular, with eigenvalues of A on the diagonal.

9 Exercises

Exercise 9.1 (Equivalence of norms in $\mathbb{R}^n$)**.** Show that $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\,\|x\|_2$ and $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\,\|x\|_\infty$.

Exercise 9.2 (Projection)**.** Compute the orthogonal projection of $(2,1,1)$ onto $\operatorname{span}\{(1,1,0),(0,1,1)\}$.

Exercise 9.3 (Dual of ℓ_1)**.** Using the definition $\|y\|_* = \sup_{\|x\|\leq 1}\langle y, x\rangle$, prove that $(\ell_1)^* = \ell_\infty$.

Exercise 9.4 (Hölder via polarity)**.** Show that $|\langle x, y\rangle| \leq 1$ for all $x \in B_{\|\cdot\|}$ and $y \in B_{\|\cdot\|_*}$, and deduce Hölder's inequality.

Exercise 9.5. Find the eigenvalues and eigenvectors of

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

Is A diagonalizable?

Exercise 9.6. Prove that if A is symmetric, then eigenvectors corresponding to distinct eigenvalues are orthogonal.

Exercise 9.7. Show that every 2×2 real matrix has at least one real eigenvalue.

Exercise 9.8. Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Diagonalize A by finding P and D.