Optimization for Machine Learning

Gouri Shanker

Contents

1 Vector Spaces

Definition 1.1 (Field)
A *field* is a nonempty set $\mathbb{F}$ together with two binary operations

$$+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}, \qquad \cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F},$$

called binary *addition* and binary *multiplication*, satisfying the following axioms.

1. **Additive structure.** The tuple $(\mathbb{F}, +)$ is an abelian group:

   (A1) $(a+b)+c = a+(b+c), \quad \forall a,b,c \in \mathbb{F}$

   (A2) $a+b = b+a, \quad \forall a,b \in \mathbb{F}$

   (A3) $\exists\, 0 \in \mathbb{F}$ such that $a+0 = a, \quad \forall a \in \mathbb{F}$

   (A4) $\forall a \in \mathbb{F},\ \exists\, (-a) \in \mathbb{F}$ such that $a + (-a) = 0$

2. **Multiplicative structure.** The tuple $(\mathbb{F} \setminus \{0\}, \cdot)$ is an abelian group:

   (M1) $(a \cdot b) \cdot c = a \cdot (b \cdot c), \quad \forall a,b,c \in \mathbb{F}$

   (M2) $a \cdot b = b \cdot a, \quad \forall a,b \in \mathbb{F}$

   (M3) $\exists\, 1 \in \mathbb{F}$, $1 \neq 0$, such that $a \cdot 1 = a, \quad \forall a \in \mathbb{F}$

   (M4) $\forall a \in \mathbb{F} \setminus \{0\},\ \exists\, a^{-1} \in \mathbb{F}$ such that $a \cdot a^{-1} = 1$

3. **Compatibility.**

   (D) $a \cdot (b+c) = a \cdot b + a \cdot c, \quad \forall a,b,c \in \mathbb{F}$

Lemma 1.2 (Uniqueness of Inverse)
Let $(\mathbb{F}, +)$ be a group and let $a \in \mathbb{F}$. If $b, c \in \mathbb{F}$ satisfy

$$a + b = 0 \qquad \text{and} \qquad a + c = 0,$$

then $b = c$.

Proof. Using associativity and the additive identity,

$$b = b + 0 = b + (a + c).$$

By associativity,

$$b + (a+c) = (b+a) + c.$$

By commutativity, $b + a = a + b = 0$, hence

$$(b+a) + c = 0 + c = c.$$

Therefore $b = c$. □

Lemma 1.3 (Uniqueness of Multiplicative Inverse)
Let $(\mathbb{F} \setminus \{0\}, \cdot)$ be a group and let $a \neq 0$. If $b, c \in \mathbb{F}$ satisfy

$$a \cdot b = 1 \qquad \text{and} \qquad a \cdot c = 1,$$

then $b = c$.

Definition 1.4 (Vector Space)
Let $\mathbb{F}$ be a field. A *vector space* over $\mathbb{F}$ is a nonempty set V together with two operations

$$+ : V \times V \to V, \qquad \cdot : \mathbb{F} \times V \to V,$$

called *vector addition* and *scalar multiplication*, such that the following axioms hold.

1. **Additive structure.** The tuple $(V, +)$ is an abelian group:

   (V1) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

   (V2) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

   (V3) $\exists\, \mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$

   (V4) $\forall \mathbf{v} \in V,\ \exists\, (-\mathbf{v}) \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$

2. **Scalar multiplication axioms.** For all $a, b \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in V$:

   (S1) $(ab)\mathbf{v} = a(b\mathbf{v})$

   (S2) $1\mathbf{v} = \mathbf{v}$

   (S3) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$

   (S4) $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$

Example $((\mathbb{F}^n, \mathbb{F}, +, \cdot))$
Let $\mathbb{F}$ be a field and $n \in \mathbb{N}$. Define

$$\mathbb{F}^n := \{(x_1, \ldots, x_n) : x_i \in \mathbb{F}\}.$$

Addition and scalar multiplication are defined componentwise:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) := (x_1 + y_1, \ldots, x_n + y_n),$$

$$a(x_1, \ldots, x_n) := (ax_1, \ldots, ax_n).$$

Then $(\mathbb{F}^n, \mathbb{F}, +, \cdot)$ is a vector space over $\mathbb{F}$.

Example $((\mathbb{F}^{m \times n}, \mathbb{F}, +, \cdot))$

Let $m, n \in \mathbb{N}$. Define

$$\mathbb{F}^{m \times n} := \{A = (a_{ij}) : a_{ij} \in \mathbb{F}\}.$$

Addition and scalar multiplication are defined entrywise:

$$(A+B)_{ij} = a_{ij} + b_{ij}, \qquad (aA)_{ij} = a\, a_{ij}.$$

Then $(\mathbb{F}^{m \times n}, \mathbb{F}, +, \cdot)$ is a vector space over $\mathbb{F}$.

Example $((\mathbb{P}_n, \mathbb{F}, +, \cdot))$
Let $n \in \mathbb{N}$. Define

$$\mathbb{P}_n := \left\{ a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \;\middle|\; a_0, \ldots, a_n \in \mathbb{F} \right\}.$$

Addition and scalar multiplication are defined by

$$(p+q)(x) := p(x) + q(x), \qquad (ap)(x) := a\, p(x).$$

Then $(\mathbb{P}_n, \mathbb{F}, +, \cdot)$ is a vector space over $\mathbb{F}$ of dimension $n+1$.

Definition 1.5 (Linear Combination)
Let $(V, \mathbb{F}, +, \cdot)$ be a vector space and let

$$\mathbf{v}_1, \ldots, \mathbf{v}_k \in V, \qquad a_1, \ldots, a_k \in \mathbb{F}.$$

A *linear combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is a vector of the form

$$a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \cdots + a_k \mathbf{v}_k.$$

Definition 1.6 (Subspace)
Let $(V, \mathbb{F}, +, \cdot)$ be a vector space. A subset $W \subseteq V$ is called a *subspace* of V if $(W, \mathbb{F}, +, \cdot)$ is itself a vector space.

Theorem 1.7 (Subspace Criterion)
Let $(V, \mathbb{F}, +, \cdot)$ be a vector space and let $S \subseteq V$ be nonempty. Then S is a subspace of V if and only if

$$\alpha u + v \in S \quad \text{for all } \alpha \in \mathbb{F} \text{ and all } u, v \in S.$$

Proof. $(\Rightarrow)$ Assume S is a subspace of V. Then S is closed under vector addition and scalar multiplication. Hence, for any $\alpha \in \mathbb{F}$ and any $u, v \in S$,

$$\alpha u \in S \quad \text{and} \quad \alpha u + v \in S.$$

$(\Leftarrow)$ Conversely, assume $S \subseteq V$ is nonempty and satisfies

$$\alpha u + v \in S \quad \text{for all } \alpha \in \mathbb{F},\ u, v \in S.$$

Closure under scalar multiplication. Fix $u \in S$ and $\alpha \in \mathbb{F}$. Since S is nonempty, choose $v \in S$. Taking $v = 0u = 0 \cdot u$, we obtain

$$\alpha u = \alpha u + 0 \in S.$$

Closure under addition. Let $u, v \in S$. Taking $\alpha = 1$, we have

$$u + v = 1 \cdot u + v \in S.$$

Existence of additive identity. Let $u \in S$. Taking $\alpha = 0$, we obtain

$$0 = 0 \cdot u + u \in S.$$

Existence of additive inverse. Let $u \in S$. Since $0 \in S$, taking $\alpha = -1$ gives

$$-u = (-1)u + 0 \in S.$$

Thus S contains 0, is closed under addition and scalar multiplication, and contains additive inverses. Therefore $(S, +, \cdot)$ is a vector space, and hence S is a subspace of V. $\square$

Theorem 1.8 (Intersection of Subspaces)
Let $(V, \mathbb{F}, +, \cdot)$ be a vector space and let $\mathcal{S}$ be a nonempty collection of subspaces of V. Define

$$W := \bigcap_{S \in \mathcal{S}} S.$$

Then W is a subspace of V.

Proof. Since each $S \in \mathcal{S}$ is a subspace, we have $\mathbf{0} \in S$ for all $S \in \mathcal{S}$. Hence $\mathbf{0} \in W$, and thus W is nonempty.

Let $\mathbf{u}, \mathbf{v} \in W$ and let $\alpha \in \mathbb{F}$. Then $\mathbf{u}, \mathbf{v} \in S$ for every $S \in \mathcal{S}$. Since each S is a subspace, it is closed under linear combinations, and therefore

$$\alpha \mathbf{u} + \mathbf{v} \in S \quad \text{for all } S \in \mathcal{S}.$$

Hence $\alpha \mathbf{u} + \mathbf{v} \in \bigcap_{S \in \mathcal{S}} S = W$.

By the subspace criterion, W is a subspace of V. $\square$

Definition 1.9 (Subspace Spanned by a Set)
Let $(V, \mathbb{F}, +, \cdot)$ be a vector space and let

$$S := \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subseteq V.$$

Let $\mathcal{S}$ denote the collection of all subspaces of V that contain S, and define

$$W := \bigcap_{K \in \mathcal{S}} K.$$

Then W is a subspace of V, called the *subspace spanned by* S, and is

denoted by
$$W = \text{span}(S).$$

Proposition 1.10

$$\text{span}(S) = \left\{ \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \;\middle|\; \alpha_1, \dots, \alpha_n \in \mathbb{F} \right\}.$$

Proof. Let
$$S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subseteq V,$$
and let
$$W := \bigcap_{K \in \mathcal{S}} K,$$
where $\mathcal{S}$ denotes the collection of all subspaces of V containing S. Let
$$\text{span}(S) := \left\{ \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \;\middle|\; \alpha_1, \dots, \alpha_n \in \mathbb{F} \right\}.$$

Step 1: span$(S) \subseteq W$.

Let $\mathbf{u} \in \text{span}(S)$. Then
$$\mathbf{u} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \quad \text{for some } \alpha_1, \dots, \alpha_n \in \mathbb{F}.$$
Let $K \in \mathcal{S}$ be arbitrary. Since K is a subspace containing S, we have $\mathbf{v}_i \in K$ for all i. By closure of K under linear combinations,
$$\mathbf{u} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \in K.$$
Since this holds for every $K \in \mathcal{S}$, it follows that
$$\mathbf{u} \in \bigcap_{K \in \mathcal{S}} K = W.$$
Hence span$(S) \subseteq W$.

Step 2: $W \subseteq$ span(S).

We first note that span(S) is a subspace of V and contains S. Therefore,
$$\text{span}(S) \in \mathcal{S}.$$
By definition of W as the intersection of all elements of $\mathcal{S}$,
$$W = \bigcap_{K \in \mathcal{S}} K \subseteq \text{span}(S).$$

Combining the two inclusions, we conclude that
$$W = \text{span}(S).$$

□

Definition 1.11 (Linear Independence)
Let $(V, \mathbb{F}, +, \cdot)$ be a vector space and let

$$S := \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq V.$$

The set S is said to be *linearly independent* if

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n = \mathbf{0}$$

implies

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0.$$

Otherwise, S is called *linearly dependent*.

Definition 1.12 (Basis)
Let $(V, \mathbb{F}, +, \cdot)$ be a vector space and let

$$S := \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq V.$$

The set S is called a *basis* of V if

1. S is linearly independent, and
2. $\text{span}(S) = V$.

Remark
A vector space $(V, \mathbb{F}, +, \cdot)$ is said to be *finite-dimensional* if there exists a finite set $B \subseteq V$ that forms a basis of V.

Remark
If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V, then the integer n is called the *dimension* of V and is denoted by $\dim V = n$.

Theorem 1.13
Let $(V, \mathbb{F}, +, \cdot)$ be a vector space and suppose

$$V = \text{span}(S), \qquad S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}.$$

Then any linearly independent set of vectors in V is finite and contains at most m vectors.

Proof. Let

$$L = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \subseteq V$$

be a linearly independent set. Since $V = \text{span}(S)$, each $\mathbf{u}_j$ can be written as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$:

$$\mathbf{u}_j = \sum_{i=1}^{m} a_{ij} \mathbf{v}_i, \qquad a_{ij} \in \mathbb{F}.$$

Suppose, for contradiction, that $k > m$. Consider a linear combination

$$\alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k = \mathbf{0}.$$

Substituting the expressions for $\mathbf{u}_j$,

$$\sum_{j=1}^{k} \alpha_j \left(\sum_{i=1}^{m} a_{ij}\mathbf{v}_i \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{k} \alpha_j a_{ij} \right) \mathbf{v}_i = \mathbf{0}.$$

This is a linear combination of the m vectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$. Since there are $k > m$ scalars $\alpha_1, \ldots, \alpha_k$, the homogeneous system

$$\sum_{j=1}^{k} \alpha_j a_{ij} = 0, \qquad i = 1, \ldots, m,$$

has a nontrivial solution. Hence there exist scalars, not all zero, such that

$$\alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k = \mathbf{0},$$

contradicting the linear independence of L.

Therefore $k \leq m$. Thus every linearly independent set in V is finite and contains no more than m vectors. □

Corollary 1.14
Let $(V, \mathbb{F}, +, \cdot)$ be a finite-dimensional vector space. Then any two bases of V contain the same number of vectors.

Proof. Let

$$B_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}, \qquad B_2 = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$$

be two bases of V. Since B_1 spans V and B_2 is linearly independent, by Theorem 1.13 we have

$$n \leq m.$$

Similarly, since B_2 spans V and B_1 is linearly independent, the same theorem gives

$$m \leq n.$$

Therefore $m = n$. □

Definition 1.15 (Linear Transformation)
Let $(V, \mathbb{F}, +, \cdot)$ and $(W, \mathbb{F}, +, \cdot)$ be vector spaces. A map

$$T : V \to W$$

is called a *linear transformation* if for all $\mathbf{u}, \mathbf{v} \in V$ and all $\alpha \in \mathbb{F}$,
(L1) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ (additivity)
(L2) $T(\alpha\mathbf{u}) = \alpha\, T(\mathbf{u})$ (homogeneity)

Remark
A map $T : V \to W$ is linear if and only if

$$T(\alpha\mathbf{u} + \mathbf{v}) = \alpha T(\mathbf{u}) + T(\mathbf{v}) \quad \forall \alpha \in \mathbb{F},\ \mathbf{u}, \mathbf{v} \in V.$$

Proposition 1.16
If $T : V \to W$ is linear, then

$$T(\mathbf{0}) = \mathbf{0}, \qquad T(-\mathbf{u}) = -T(\mathbf{u}) \quad \forall\, \mathbf{u} \in V.$$

Definition 1.17 (Vector Space Homomorphism)
Let $(V, \mathbb{F}, +, \cdot)$ and $(W, \mathbb{F}, +, \cdot)$ be vector spaces. A map

$$T : V \to W$$

is called a *(vector space) homomorphism* if for all $\mathbf{u}, \mathbf{v} \in V$ and all $\alpha \in \mathbb{F}$,

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}), \qquad T(\alpha\mathbf{u}) = \alpha\, T(\mathbf{u}).$$

Remark
A map $T : V \to W$ is a homomorphism if and only if

$$T(\alpha\mathbf{u} + \mathbf{v}) = \alpha T(\mathbf{u}) + T(\mathbf{v}) \quad \forall\, \alpha \in \mathbb{F},\ \mathbf{u}, \mathbf{v} \in V.$$

Definition 1.18 (Linear Functional)
Let $(V, \mathbb{F}, +, \cdot)$ be a vector space. A *linear functional* on V is a linear transformation

$$f : V \to \mathbb{F}.$$

That is, for all $\mathbf{u}, \mathbf{v} \in V$ and all $\alpha \in \mathbb{F}$,

$$f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v}), \qquad f(\alpha\mathbf{u}) = \alpha\, f(\mathbf{u}).$$

Example (Evaluation Functional)
Fix $a \in \mathbb{F}$. Define $f_a : \mathbb{F}[x] \to \mathbb{F}$ by

$$f_a(p) := p(a).$$

Then f_a is a linear functional.

Example (Constant-Coefficient Functional)
Define $f_0 : \mathbb{P}_n \to \mathbb{F}$ by

$$f_0(a_0 + a_1 x + \cdots + a_n x^n) := a_0.$$

Then f_0 is a linear functional.

Remark
The set of all linear functionals on V forms a vector space over $\mathbb{F}$, called the *dual space* and denoted by V^*.

Definition 1.19 (Inner Product)
Let V be a vector space over $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. An *inner product*

on V is a map

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$$

satisfying the following properties for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all $\alpha \in \mathbb{F}$:

(IP1) $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$ (conjugate symmetry)

(IP2) $\langle \alpha\mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \alpha\langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ (linearity in the first argument)

(IP3) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \iff \mathbf{u} = \mathbf{0}$ (positive definiteness)

If $\mathbb{F} = \mathbb{R}$, conjugation is trivial and symmetry reduces to $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$

Definition 1.20 (Inner Product Space)
An *inner product space* is a pair $(V, \langle \cdot, \cdot \rangle)$ where V is a vector space over $\mathbb{F}$ and $\langle \cdot, \cdot \rangle$ is an inner product on V.

Example ($\mathbb{F}^n$)

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

Example ($\mathbb{P}_n$)

$$\langle p, q \rangle := \int_0^1 p(x)\overline{q(x)}\, dx.$$

Proposition 1.21
For all $\mathbf{u}, \mathbf{v} \in V$:

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0 \iff \mathbf{u} \perp \mathbf{v}, \qquad \|\mathbf{u}\| := \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Definition 1.22 (Norm)
For $\mathbf{u} \in V$, define

$$\|\mathbf{u}\| := \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Definition 1.23 (Orthogonality)
Vectors $\mathbf{u}, \mathbf{v} \in V$ are said to be *orthogonal*, written $\mathbf{u} \perp \mathbf{v}$, if

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0.$$

Proposition 1.24
Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Then for all $\mathbf{u}, \mathbf{v} \in V$:

1. $\|\mathbf{u}\| \geq 0$ and $\|\mathbf{u}\| = 0 \iff \mathbf{u} = \mathbf{0}$
2. $\|\alpha\mathbf{u}\| = |\alpha|\, \|\mathbf{u}\|$
3. $\langle \mathbf{u}, \mathbf{v} \rangle = 0 \iff \mathbf{u} \perp \mathbf{v}$

Definition 1.25 (Norm)
Let $(V, \mathbb{F}, +, \cdot)$ be a vector space, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A *norm* on V is a function

$$\|\cdot\| : V \to [0, \infty)$$

satisfying, for all $\mathbf{u}, \mathbf{v} \in V$ and all $\alpha \in \mathbb{F}$,

(N1) $\|\mathbf{u}\| \geq 0$ and $\|\mathbf{u}\| = 0$ if and only if $\mathbf{u} = \mathbf{0}$ (positive definiteness)
(N2) $\|\alpha\mathbf{u}\| = |\alpha| \, \|\mathbf{u}\|$ (absolute homogeneity)
(N3) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ (triangle inequality)

Definition 1.26 (Normed Vector Space)
A *normed vector space* is a pair $(V, \|\cdot\|)$, where V is a vector space and $\|\cdot\|$ is a norm on V.

Example (ℓ^p-type norms on $\mathbb{F}^n$)
For $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}^n$, define

$$\|\mathbf{x}\|_p := \begin{cases} \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, & 1 \leq p < \infty, \\ \max_{1 \leq i \leq n} |x_i|, & p = \infty. \end{cases}$$

Then $(\mathbb{F}^n, \|\cdot\|_p)$ is a normed vector space.

Example (Norm induced by an inner product)
Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Define

$$\|\mathbf{u}\| := \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Then $\|\cdot\|$ is a norm on V.

Example (Polynomial norm)
Let $V = \mathbb{P}_n$ and $\mathbb{F} = \mathbb{R}$. Define

$$\|p\| := \max_{x \in [0,1]} |p(x)|.$$

Then $(\mathbb{P}_n, \|\cdot\|)$ is a normed vector space.

Proposition 1.27
Let $(V, \|\cdot\|)$ be a normed vector space. Then for all $\mathbf{u}, \mathbf{v} \in V$:

1. $\|\mathbf{u} - \mathbf{v}\| = 0 \iff \mathbf{u} = \mathbf{v}$
2. $\|\mathbf{u} - \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$
3. $|\|\mathbf{u}\| - \|\mathbf{v}\|| \leq \|\mathbf{u} - \mathbf{v}\|$

Remark
Every inner product induces a norm, but not every norm arises from an inner product.

Definition 1.28 (Metric)
Let X be a nonempty set. A *metric* on X is a function

$$d : X \times X \to [0, \infty)$$

satisfying, for all $x, y, z \in X$:
(M1) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$ (non-negativity and identity)
(M2) $d(x, y) = d(y, x)$ (symmetry)
(M3) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality)

Definition 1.29 (Metric Space)
A *metric space* is a pair (X, d), where X is a set and d is a metric on X.

Example (Euclidean Metric)
Let $X = \mathbb{F}^n$ and define

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|_2.$$

Then $(\mathbb{F}^n, d)$ is a metric space.

Example (Discrete Metric)
Let X be any set and define

$$d(x, y) := \begin{cases} 0, & x = y, \\ 1, & x \neq y. \end{cases}$$

Then (X, d) is a metric space.

Example (Polynomial Metric)
Let $X = \mathbb{P}_n$ and define

$$d(p, q) := \|p - q\|_\infty = \max_{x \in [0,1]} |p(x) - q(x)|.$$

Then $(\mathbb{P}_n, d)$ is a metric space.

Proposition 1.30
Let $(V, \|\cdot\|)$ be a normed vector space. Define

$$d(\mathbf{u}, \mathbf{v}) := \|\mathbf{u} - \mathbf{v}\|.$$

Then d is a metric on V.

Proof. Non-negativity and symmetry follow directly from properties of the norm. Moreover,

$$d(\mathbf{u}, \mathbf{v}) = 0 \iff \|\mathbf{u} - \mathbf{v}\| = 0 \iff \mathbf{u} = \mathbf{v}.$$

Finally, by the triangle inequality for the norm,

$$d(\mathbf{u}, \mathbf{w}) = \|\mathbf{u} - \mathbf{w}\| \leq \|\mathbf{u} - \mathbf{v}\| + \|\mathbf{v} - \mathbf{w}\| = d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w}).$$

$\square$

Remark
Every normed vector space is a metric space, but a metric space need not carry any linear or vector space structure.

2 Convex Optimization

Definition 2.1 (Convex Set)
A set $C \subseteq \mathbb{R}^n$ is called *convex* if for all $\mathbf{x}, \mathbf{y} \in C$ and all $\lambda \in [0, 1]$,

$$\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} \in C.$$

Definition 2.2 (Convex Function)
Let $C \subseteq \mathbb{R}^n$ be a convex set. A function $f : C \to \mathbb{R}$ is called *convex* if for all $\mathbf{x}, \mathbf{y} \in C$ and all $\lambda \in [0, 1]$,

$$f\big(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}\big) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}).$$

Definition 2.3 (Convex Program)
A *convex program* is an optimization problem of the form

$$\min_{\mathbf{x} \in C} f(\mathbf{x}),$$

where $C \subseteq \mathbb{R}^n$ is a convex set and $f : C \to \mathbb{R}$ is a convex function.

Definition 2.4 (Convex Program)
A *convex program* is an optimization problem of the form

$$\min_{\mathbf{x} \in C} f(\mathbf{x}),$$

where $C \subseteq \mathbb{R}^n$ is a convex set and $f : C \to \mathbb{R}$ is a convex function.

Definition 2.5 (Linear Program)
A *linear program* is an optimization problem of the form

$$\min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{c}^\mathsf{T}\mathbf{x} \quad \text{subject to } A\mathbf{x} \leq \mathbf{b},$$

where $\mathbf{c} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, and $\mathbf{b} \in \mathbb{R}^m$.

Remark
Every linear program is a convex program.

Definition 2.6 (Affine Set)
A set $A \subseteq \mathbb{R}^n$ is called *affine* if for all $\mathbf{x}, \mathbf{y} \in A$ and all $\lambda \in \mathbb{R}$,

$$\lambda\mathbf{x} + (1 - \lambda)\mathbf{y} \in A.$$

Definition 2.7 (Affine Hull)
Let $S \subseteq \mathbb{R}^n$. The *affine hull* of S, denoted $\mathrm{aff}(S)$, is the smallest affine

set containing S, equivalently

$$\operatorname{aff}(S) = \left\{ \sum_{i=1}^{k} \alpha_i \mathbf{x}_i \;\middle|\; \mathbf{x}_i \in S,\ \sum_{i=1}^{k} \alpha_i = 1 \right\}.$$

Definition 2.8 (Graph)
Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$. The *graph* of f is the set

$$\operatorname{graph}(f) := \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} : t = f(\mathbf{x}),\ \mathbf{x} \in C\}.$$

Definition 2.9 (Epigraph)
Let $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$. The *epigraph* of f is the set

$$\operatorname{epi}(f) := \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} : t \geq f(\mathbf{x}),\ \mathbf{x} \in C\}.$$

Theorem 2.10
Let $C \subseteq \mathbb{R}^n$ be a convex set and let $f : C \to \mathbb{R}$ be a function. If the epigraph $\operatorname{epi}(f)$ is convex, then f is convex.

Proof. Assume that $\operatorname{epi}(f)$ is convex. Let $\mathbf{x}, \mathbf{y} \in C$ and let $\lambda \in [0, 1]$.

By definition of the epigraph,

$$(\mathbf{x}, f(\mathbf{x})) \in \operatorname{epi}(f), \qquad (\mathbf{y}, f(\mathbf{y})) \in \operatorname{epi}(f).$$

Since $\operatorname{epi}(f)$ is convex, their convex combination also belongs to $\operatorname{epi}(f)$:

$$\lambda(\mathbf{x}, f(\mathbf{x})) + (1 - \lambda)(\mathbf{y}, f(\mathbf{y})) \in \operatorname{epi}(f).$$

That is,

$$\big(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y},\ \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y})\big) \in \operatorname{epi}(f).$$

By definition of the epigraph, this implies

$$f\big(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}\big) \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}).$$

Since $\mathbf{x}, \mathbf{y} \in C$ and $\lambda \in [0, 1]$ were arbitrary, f is convex on C. □

Theorem 2.11 (Separating Hyperplane Theorem)
Let $C, D \subseteq \mathbb{R}^n$ be two nonempty, disjoint, convex sets. Assume that C is closed. Then there exist a nonzero vector $\mathbf{a} \in \mathbb{R}^n$ and a scalar $b \in \mathbb{R}$ such that

$$\mathbf{a}^{\mathsf{T}}\mathbf{x} \leq b \quad \text{for all } \mathbf{x} \in C, \qquad \mathbf{a}^{\mathsf{T}}\mathbf{y} \geq b \quad \text{for all } \mathbf{y} \in D.$$

The hyperplane

$$H := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^{\mathsf{T}}\mathbf{x} = b\}$$

is said to *separate* C and D.

Proof. Since C is nonempty, closed, and convex, and D is nonempty and disjoint from C, choose an arbitrary point $\mathbf{y} \in D$. Let

$$\mathbf{p} := \Pi_C(\mathbf{y})$$

denote the (unique) projection of $\mathbf{y}$ onto C.

By the characterization of projections onto closed convex sets,

$$(\mathbf{y} - \mathbf{p})^\mathsf{T}(\mathbf{x} - \mathbf{p}) \le 0 \quad \text{for all } \mathbf{x} \in C.$$

Define

$$\mathbf{a} := \mathbf{y} - \mathbf{p}, \qquad b := \mathbf{a}^\mathsf{T}\mathbf{p}.$$

Note that $\mathbf{a} \neq \mathbf{0}$ since $\mathbf{y} \notin C$.

For any $\mathbf{x} \in C$, we have

$$\mathbf{a}^\mathsf{T}\mathbf{x} = \mathbf{a}^\mathsf{T}\mathbf{p} + \mathbf{a}^\mathsf{T}(\mathbf{x} - \mathbf{p}) \le \mathbf{a}^\mathsf{T}\mathbf{p} = b.$$

On the other hand, for $\mathbf{y}$ itself,

$$\mathbf{a}^\mathsf{T}\mathbf{y} = \mathbf{a}^\mathsf{T}\mathbf{p} + \|\mathbf{a}\|^2 > \mathbf{a}^\mathsf{T}\mathbf{p} = b.$$

By convexity of D, the inequality

$$\mathbf{a}^\mathsf{T}\mathbf{y} \ge b$$

extends to all $\mathbf{y} \in D$. Thus the hyperplane $\{\mathbf{x} : \mathbf{a}^\mathsf{T}\mathbf{x} = b\}$ separates C and D. □

> **Remark**
> If both C and D are closed and one of them is compact, the separation can be made *strict*.

> **Remark**
> This theorem is the geometric foundation of duality, KKT conditions, and optimality certificates in convex optimization.