Tutorial notes

2024-2025 Optimization theory

1 Norms

Definition 1. *Let $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}^+$ be a non-negative function. If it satisfies the following conditions:*

- ***Positivity***: *For all $\mathbf{v} \in \mathbb{R}^n$, we have $\|\mathbf{v}\| \geq 0$, and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}_{n\times 1}$.*
- ***Homogeneity***: *For all $\mathbf{v} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, we have $\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\|$.*
- ***Triangle inequality***: *For all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have*

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|. \tag{1}$$

*Then, $\|\cdot\|$ is called a **vector norm** in the vector space $\mathbb{R}^n$.*

The most commonly used vector norms are the ℓ_p norms (where $p \geq 1$):

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}}, \quad \|\mathbf{v}\|_\infty = \max_{1\leq j\leq n} |v_j|. \tag{2}$$

Cauchy-Schwarz inequality: For any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, we have

$$|\mathbf{a}^T\mathbf{b}| \leq \|\mathbf{a}\|_2\|\mathbf{b}\|_2, \tag{3}$$

with equality if and only if $\mathbf{a}$ and $\mathbf{b}$ are linearly dependent.

Definition 2. *Matrix norms can be extended from vector norms. Common matrix norms include:*

$$\|A\|_1 = \sum_{i,j} |A_{ij}| \tag{4}$$

$$\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2} = \sqrt{\mathrm{Tr}(AA^\top)} \tag{5}$$

Operator norms are a special class of matrix norms, which are induced by vector norms:

$$\|A\|_{(m,n)} = \max_{\mathbf{x}\in\mathbb{R}^n, \|\mathbf{x}\|_{(n)}=1} \|A\mathbf{x}\|_{(m)}. \tag{6}$$

For specific values of p, we have:

- When $p = 1$:

$$\|A\|_{p=1} = \max_{\|\mathbf{x}\|_1=1} \|A\mathbf{x}\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|. \tag{7}$$

- When $p = 2$:

$$\|A\|_{p=2} = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 = \sqrt{\lambda_{\max}(A^\top A)}, \tag{8}$$

  also known as the **spectral norm** of A.

- When $p = \infty$:

$$\|A\|_{p=\infty} = \max_{\|\mathbf{x}\|_\infty=1} \|A\mathbf{x}\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|. \tag{9}$$

Definition 3. *The **nuclear norm** (or trace norm) of a matrix is defined as:*

$$\|A\|_* = \sum_{i=1}^{r} \sigma_i, \tag{10}$$

where $\sigma_i (i = 1, \ldots, r)$ are the nonzero singular values of A, and $r = \operatorname{rank}(A)$.

Definition 4. *The **inner product** of two matrices A and B is defined as:*

$$\langle A, B\rangle = \operatorname{Tr}(AB^\top) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij}. \tag{11}$$

Cauchy-Schwarz inequality: Given matrices $A, B \in \mathbb{R}^{m\times n}$, we have

$$|\langle A, B\rangle| \le \|A\|_F\|B\|_F. \tag{12}$$

Equality holds if and only if A and B are linearly dependent.

2 Convex Sets

In $\mathbb{R}^n$, a line passing through two distinct points x_1 and x_2 is given by:

$$y = \theta x_1 + (1-\theta)x_2, \quad \theta \in \mathbb{R}. \tag{13}$$

In particular, when $0 \le \theta \le 1$, the line reduces to the line segment with endpoints x_1 and x_2.

Definition 5. *A set C is called a **affine set** if for any two points in C, the entire line passing through them is also contained in C. That is,*

$$x_1, x_2 \in C \Rightarrow \theta x_1 + (1-\theta)x_2 \in C, \quad \forall\theta \in \mathbb{R}. \tag{14}$$

Example: The solution set $\mathcal{X}$ of a linear system $Ax = b$ is an affine set, because for any $x_1, x_2 \in \mathcal{X}$ $(x_1 \ne x_2)$, we have

$$\theta Ax_1 + (1-\theta)Ax_2 = b. \tag{15}$$

Conversely, any affine set can be represented as the solution set of some linear system.

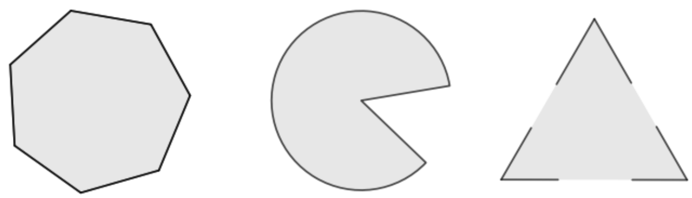
Definition 6. *A set C is called a **convex set** if, for any two points in C, the entire line segment connecting them is also contained in C. That is,*

$$x_1, x_2 \in C \Rightarrow \theta x_1 + (1-\theta)x_2 \in C, \quad \forall 0 \leq \theta \leq 1. \tag{16}$$

Affine sets are naturally convex sets.

Example: The figure below illustrates examples of convex and non-convex sets. In particular:

- (a) is a convex set.
- (b) and (c) are non-convex sets.



(a) (b) (c)

Theorem 1.
- *If S is a convex set, then $kS = \{ks \mid k \in \mathbb{R}, s \in S\}$ is also a convex set.*
- *If S and T are both convex sets, then their Minkowski sum $S + T = \{s + t \mid s \in S, t \in T\}$ is a convex set.*
- *If S and T are both convex sets, then their intersection $S \cap T$ is a convex set.*
- *The interior and closure of a convex set are also convex.*

The first two points in the theorem are obvious under the definition of convex sets. We briefly prove the third point.

Proof: Let $x, y \in S \cap T$ and $\theta \in [0, 1]$. Since both S and T are convex, we have

$$\theta x + (1-\theta)y \in S \cap T. \tag{17}$$

This proves that $S \cap T$ is convex.

In fact, the intersection of any number of convex sets is also convex. This conclusion is very useful when proving that complex sets are convex since we can consider them as the intersection of multiple convex sets.

Theorem 2. *Let S be a convex subset of a topological vector space X. Then:*

1. $\mathrm{int}(S)$ *is convex (provided* $\mathrm{int}(S) \neq \varnothing$*).*
2. $\overline{S}$ *(the closure of S) is convex.*

Proof. **(1) Convexity of** $\mathrm{int}(S)$**.**

Suppose $\mathrm{int}(S)$ is non-empty. We need to show that for any $x, y \in \mathrm{int}(S)$ and any $t \in [0, 1]$, the point

$$(1-t)x + t\,y \quad \text{also belongs to} \quad \mathrm{int}(S).$$

Since x and y lie in $\mathrm{int}(S)$, by definition there exist open neighborhoods U and V of 0 in X such that

$$x + U \subseteq S \quad \text{and} \quad y + V \subseteq S.$$

For any $t \in [0, 1]$, let

$$z_t = (1-t)x + t\,y.$$

Consider the set

$$z_t + \big((1-t)U + t\,V\big).$$

We can rewrite this as

$$(1-t)(x+U) + t\,(y+V).$$

Because S itself is convex, every convex combination of points in $x + U$ and $y + V$ remains in S. Hence,

$$(1-t)(x+U) + t\,(y+V) \;\subseteq\; S.$$

Thus,

$$z_t + \big((1-t)U + t\,V\big) \;\subseteq\; S.$$

Note that $(1-t)U + t\,V$ is an open set (the Minkowski sum of open sets in a topological vector space is open). Therefore, we have found an open neighborhood around z_t contained in S, implying

$$z_t \in \mathrm{int}(S).$$

Since x, y were arbitrary points in $\mathrm{int}(S)$, we have shown that all their convex combinations remain in $\mathrm{int}(S)$. Hence, $\mathrm{int}(S)$ is convex.

(2) Convexity of $\overline{S}$**.**

Let $x, y \in \overline{S}$. By the definition of closure, there exist sequences $(x_n) \subseteq S$ and $(y_n) \subseteq S$ such that

$$x_n \to x \quad \text{and} \quad y_n \to y \quad \text{as } n \to \infty.$$

For each n and each $t \in [0, 1]$, define

$$z_n(t) = (1-t)x_n + t\,y_n.$$

Since S is convex, we have $z_n(t) \in S$ for all n. Now consider the point

$$z(t) = (1-t)x + t\,y.$$

Taking the limit as $n \to \infty$, we get

$$\lim_{n\to\infty} z_n(t) = \lim_{n\to\infty} \big((1-t)x_n + t\,y_n\big) = (1-t)\lim_{n\to\infty} x_n + t\lim_{n\to\infty} y_n = (1-t)x + t\,y = z(t).$$

Since each $z_n(t)$ lies in S and $z_n(t) \to z(t)$, it follows that $z(t)$ is in the closure of S, i.e., $z(t) \in \overline{S}$. Thus, for $x, y \in \overline{S}$, every convex combination of x and y also lies in $\overline{S}$. Therefore, $\overline{S}$ is convex.

$\square$

Theorem 3. *Let X be a normed vector space (or more generally a metric vector space) and let $S \subseteq X$ be a convex set. Suppose* $\operatorname{int}(S)$ *is nonempty. Then* $\operatorname{int}(S)$ *is convex.*

Proof. We need to show that for any $x, y \in \operatorname{int}(S)$ and any $t \in [0,1]$,

$$z := (1-t)x + t\,y$$

also belongs to $\operatorname{int}(S)$.

Step 1: Existence of balls in the interior.

Because x lies in $\operatorname{int}(S)$, there exists $r > 0$ such that the open ball

$$B(x,r) = \{\, u \in X : \|u - x\| < r\}$$

is contained in S. Similarly, since $y \in \operatorname{int}(S)$, there exists $s > 0$ such that

$$B(y,s) \subseteq S.$$

Step 2: Construct a ball around the convex combination.

Define $z = (1-t)x + t\,y$. We claim z also lies in the interior of S. More precisely, we will show that there is an open ball $B(z,R)$ (for some radius $R > 0$) which is still contained in S.

Let $R = (1-t)r + t\,s$. Consider the ball

$$B(z,R) = \big\{\, w \in X : \|w - z\| < R\big\}.$$

Take any point $w \in B(z,R)$. Then

$$\|w - z\| < R.$$

We want to show $w \in S$. Notice we can rewrite w as

$$w = z + (w-z) \;=\; (1-t)x + t\,y + (w-z).$$

We would like to decompose $(w-z)$ in a way that "splits" into parts near x and y, respectively.

Since $\|w - z\| < R$, we can argue that there exist vectors u, v with $\|u\| < r$, $\|v\| < s$ such that

$$w = (1-t)(x+u) + t(y+v).$$

(An intuitive way to see this: any small enough perturbation $w-z$ can be distributed as $(1-t)u + tv$ where $\|u\| < r$ and $\|v\| < s$, provided $\|w - z\| < (1-t)r + t\,s$. Rigorously, one can scale and split $\|w - z\|$ into two parts in the ratio $(1-t) : t$.)

Step 3: Verify w remains in S by convexity.

- Note that $x + u \in B(x,r) \subseteq S$ because $\|u\| < r$. - Similarly, $y + v \in B(y,s) \subseteq S$ because $\|v\| < s$.
- Since S is convex, any convex combination of points in S remains in S. Hence

$$w \;=\; (1-t)(x+u) + t(y+v) \;\in\; S.$$

Thus for every w in $B(z, R)$, we have shown $w \in S$. Consequently,

$$B\big(z, (1-t)r + t\,s\big) \;\subseteq\; S.$$

This means z has a neighborhood entirely contained in S, i.e., $z \in \operatorname{int}(S)$.

□

From convex sets, we can introduce the concepts of convex combinations and convex hulls.

Definition 7. *A **convex combination** of points $x_1, \dots, x_k$ is a point of the form:*

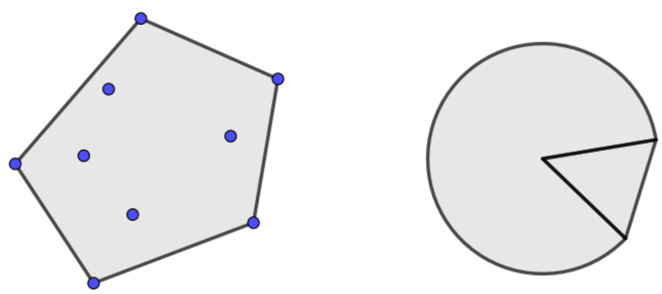
$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k, \tag{18}$$

where

$$\theta_1 + \cdots + \theta_k = 1, \quad \theta_i \geq 0, \quad i = 1, \dots, k. \tag{19}$$

Definition 8. *The **convex hull** of a set S is the set of all convex combinations of points in S. It is denoted as:*

$$\operatorname{conv} S. \tag{20}$$



Theorem 4. ***Relationship Between Convex Sets and Convex Hulls**:*

If $\operatorname{conv} S \subseteq S$*, then S is a convex set; conversely, if S is convex, then* $\operatorname{conv} S = S$*.*

The above theorem is not trivial; please try to prove it. *(Hint: Use mathematical induction.)*

Theorem 5. $\operatorname{conv} S$ *is the smallest convex set containing S.*

Proof: By the definition of the convex hull, we have:

$$S \subseteq \operatorname{conv} S, \tag{21}$$

and $\operatorname{conv} S$ is a convex set.

Now, suppose $\mathcal{X}$ is another convex set satisfying $S \subseteq \mathcal{X} \subseteq \operatorname{conv} S$. We need to show that $\mathcal{X} = \operatorname{conv} S$. To do so, we first prove an important lemma, from which this theorem follows directly.

Theorem 6. *For any vector set S, the convex hull* $\operatorname{conv} S$ *is the intersection of all convex sets containing S.*

Proof: Let $\mathcal{X}$ denote the intersection of all convex sets containing S. We previously proved that the intersection of convex sets is convex, so $\mathcal{X}$ is convex.

Since $\operatorname{conv} S$ is a convex set that contains S, we have:

$$\mathcal{X} \subseteq \operatorname{conv} S. \tag{22}$$

On the other hand, since $S \subseteq \mathcal{X}$, we also obtain:

$$\operatorname{conv} S \subseteq \mathcal{X}. \tag{23}$$

Using the relationship between convex sets and convex hulls, we conclude:

$$\operatorname{conv} S = \mathcal{X}. \tag{24}$$

Affine sets and convex sets have similar definitions, except for the range of θ. Inspired by this, we introduce the concepts of **affine combinations** and **affine hulls**.

Definition 9. *An **affine combination** of points $x_1, \ldots, x_k$ is a point of the form:*

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k, \tag{25}$$

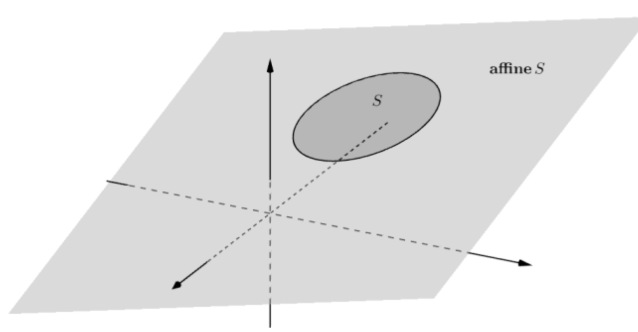
where

$$\theta_1 + \cdots + \theta_k = 1, \quad \theta_i \in \mathbb{R}, \quad i = 1, \ldots, k. \tag{26}$$

Definition 10. *The **affine hull** of a set S is the set of all affine combinations of points in S. It is denoted as:*

$$\operatorname{aff} S. \tag{27}$$

$\operatorname{aff} S$ is the smallest affine set containing S.



Compared to convex combinations and affine combinations, conic combinations do not require the coefficients to sum to 1. Therefore, in general, conic combinations are open.

Definition 11. *A **conic combination** of points $x_1, \ldots, x_k$ is a point of the form:*

$$x = \theta_1 x_1 + \cdots + \theta_k x_k, \tag{28}$$

where

$$\theta_i > 0, \quad i = 1, \ldots, k. \tag{29}$$

Definition 12. *A set S is called a **convex cone** if it contains all conic combinations of its points. That is, for any $x_1, \ldots, x_k \in S$, we have:*

$$\theta_1 x_1 + \cdots + \theta_k x_k \in S, \quad \forall \theta_i > 0. \tag{30}$$

Definition 13. *A **hyperplane** is the set of points satisfying a linear equation. Given a nonzero vector $a \in \mathbb{R}^n$, the set*

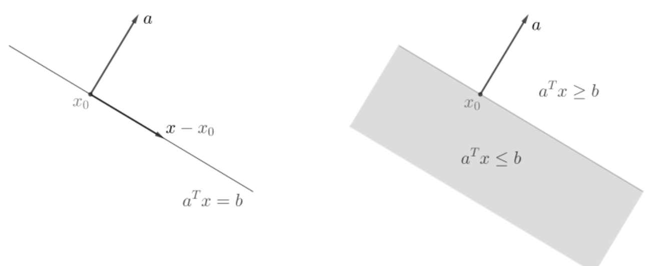
$$\{x \mid a^T x = b\} \tag{31}$$

is called a hyperplane.

Definition 14. *A **half-space** is the set of points satisfying a linear inequality. Given a nonzero vector $a \in \mathbb{R}^n$, the set*

$$\{x \mid a^T x \leq b\} \tag{32}$$

is called a half-space.



A hyperplane is both an affine set and a convex set, whereas a half-space is convex but not affine.

A set of points satisfying a system of linear equalities and inequalities is called a **polyhedron**, i.e.,

$$\{x \mid Ax \leq b, \quad Cx = d\}, \tag{33}$$

where $A \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{p \times n}$, and $x \leq y$ means that each component of vector x is less than or equal to the corresponding component of y.

A polyhedron is the intersection of a finite number of half-spaces and hyperplanes. Therefore, by the properties of convex sets, it is convex.

Affine transformations (scaling, translation, projection, etc.) preserve convexity.

Theorem 7. ***Convexity Preservation under Affine Transformations**:*

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be an affine transformation, i.e.,

$$f(x) = Ax + b, \tag{34}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then:

- *The image of a convex set under f is convex:*

$$S \subseteq \mathbb{R}^n \text{ is convex} \Rightarrow f(S) = \{f(x) \mid x \in S\} \text{ is convex.} \tag{35}$$

- *The preimage of a convex set under f is convex:*

$$C \subseteq \mathbb{R}^m \text{ is convex} \Rightarrow f^{-1}(C) = \{x \mid f(x) \in C\} \text{ is convex.} \tag{36}$$

Example 1. ***Solution Set of Linear Matrix Inequalities****:*

The set of solutions to a linear matrix inequality is convex:

$$\{x \mid x_1 A_1 + \cdots + x_m A_m \preceq B\}, \quad (A_i, i = 1, \ldots, m, B \in \mathbb{S}^p). \tag{37}$$

This follows directly from affine transformations.

Example 2. ***Quadratic Cone****:*

The quadratic inequality defines a convex set:

$$\{x \mid x^T P x \leq (c^T x)^2, \quad c^T x \geq 0, \quad P \in \mathbb{S}^n_+\}. \tag{38}$$

Proof: *The quadratic cone can be rewritten as a* ***second-order cone****:*

$$\{x \mid \|Ax\|_2 \leq c^T x, \quad c^T x \geq 0, \quad A^T A = P\}. \tag{39}$$

Since a second-order cone can be expressed as

$$\{(x, t) \mid \|x\|_2 \leq t, \quad t > 0\}, \tag{40}$$

which is convex and preserved under affine transformation, the quadratic cone and the second-order cone are convex sets.

Example 3. ***Perspective Transformation****:*

The perspective transformation $P : \mathbb{R}^{n+1} \to \mathbb{R}$ is defined as:

$$P(x, t) = \frac{x}{t}, \quad \operatorname{dom} P = \{(x, t) \mid t > 0\}. \tag{41}$$

The image and preimage of convex sets under the perspective transformation are convex.

Example 4. ***Fractional Linear Transformation****:*

The fractional linear transformation $f : \mathbb{R}^n \to \mathbb{R}^m$ is given by:

$$f(x) = \frac{Ax + b}{c^T x + d}, \quad \operatorname{dom} f = \{x \mid c^T x + d > 0\}. \tag{42}$$

The image and preimage of convex sets under the fractional linear transformation are convex.

Hyperplanes are a special type of convex (affine) set in space. It can be proven that hyperplanes in $\mathbb{R}^n$ are of dimension $n - 1$. We can use hyperplanes to separate disjoint convex sets.

Theorem 8. ***Separating Hyperplane Theorem:***

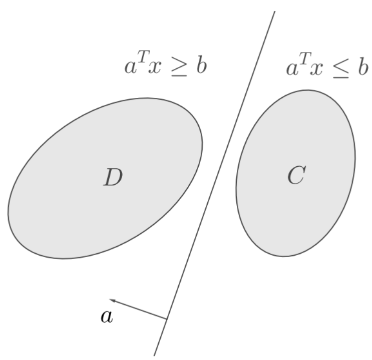
If C and D are disjoint convex sets, then there exists a nonzero vector a and a scalar b such that

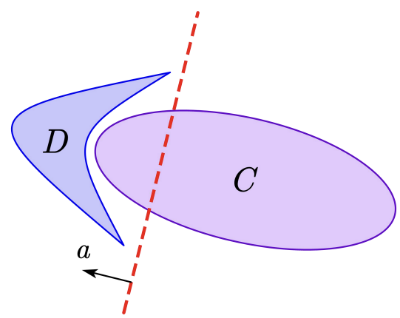
$$a^T x \leq b, \quad \forall x \in C, \tag{43}$$

and

$$a^T x \geq b, \quad \forall x \in D. \tag{44}$$

Thus, the hyperplane $\{x \mid a^T x = b\}$ separates C and D.





The separating hyperplane theorem states that if we want to **softly separate** two convex sets in $\mathbb{R}^n$, we only need to find a suitable hyperplane. In classification problems, this is relatively easy to achieve. However, if either of the sets is non-convex, the theorem generally does not hold, and more complex surfaces are required for separation.

Proof: Here, we consider a special case. Suppose there exist points $c \in C$ and $d \in D$ such that:

$$\|c - d\|_2 = \operatorname{dist}(C, D) = \inf\{\|u - v\|_2 \mid u \in C, v \in D\} > 0. \tag{45}$$

Define $a = d - c$ and

$$b = \frac{\|d\|_2^2 - \|c\|_2^2}{2}. \tag{46}$$

Then, the function

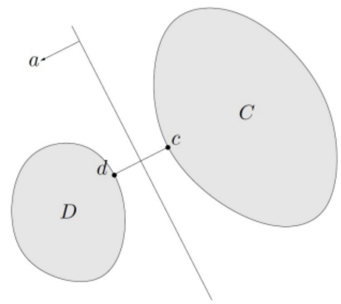
$$f(x) = a^T x - b = (d - c)^T (x - (d + c)/2) \tag{47}$$

satisfies:

- $f(x) \leq 0, \quad \forall x \in C,$
- $f(x) \geq 0, \quad \forall x \in D.$

Thus, we have constructed the separating hyperplane.

Step 1: Prove that $f(x) \geq 0$ for all $x \in D$.



Assume there exists $u \in D$ such that

$$f(u) = (d-c)^T(u-(d+c)/2) < 0. \tag{48}$$

We can rewrite this as:

$$f(u) = (d-c)^T(u-d) + \frac{\|d-c\|_2^2}{2}. \tag{49}$$

Since

$$(d-c)^T(u-d) < 0, \tag{50}$$

we define the convex combination:

$$z(t) = d + t(u-d), \quad t \in [0,1]. \tag{51}$$

Since $z(t)$ is a convex combination of d and u, and both belong to D, we have $z(t) \in D$.

Taking the derivative,

$$\frac{d}{dt}\|z(t)-c\|_2^2\bigg|_{t=0} = 2(d-c)^T(u-d) < 0. \tag{52}$$

Thus, there exists a sufficiently small $t_1 \in (0,1]$ such that:

$$\|z(t_1)-c\|_2 < \|d-c\|_2. \tag{53}$$

This implies that the point $z(t_1)$ is closer to c than d, contradicting our assumption of minimal separation. Therefore, the hyperplane correctly separates the two sets.

When discussing hyperplane separation, we introduced the concept of **soft separation**, which indicates that if a set is merely convex, the equality in the theorem may hold, meaning that a convex set and a hyperplane may still intersect. To ensure complete separation, stronger conditions are required.

Theorem 9. ***Strict Separation Theorem***:

If C and D are disjoint convex sets, with C being a closed set and D being a compact set, then there exists a nonzero vector a and a scalar b such that:

$$a^T x < b, \quad \forall x \in C, \tag{54}$$

and

$$a^T x > b, \quad \forall x \in D. \tag{55}$$

Thus, the hyperplane $\{x \mid a^T x = b\}$ *strictly separates* C *and* D.

A special case of this theorem occurs when D reduces to a single point set $\{x_0\}$, in which case the standard course theorem holds.

Definition 15. *Supporting Hyperplane:*

Given a set C *and a boundary point* x_0*, if* $a \neq 0$ *satisfies:*

$$a^T x \leq a^T x_0, \quad \forall x \in C, \tag{56}$$

then the set

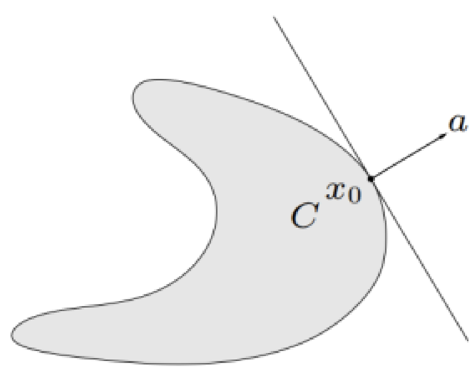
$$\{x \mid a^T x = a^T x_0\} \tag{57}$$

is called the supporting hyperplane of C *at the boundary point* x_0.

According to the definition, the point x_0 and the set C are separated by the hyperplane. Geometrically, the hyperplane $\{x \mid a^T x = a^T x_0\}$ is tangent to C at x_0, and the half-space $\{x \mid a^T x \leq a^T x_0\}$ contains C.

Theorem 10. *Supporting Hyperplane Theorem:*

If C *is a convex set, then for any boundary point of* C*, there exists a supporting hyperplane at that point.*



The supporting hyperplane theorem has a strong geometric interpretation: **Given a hyperplane, we can always find a boundary point of the convex set to serve as a supporting point, effectively placing the convex set onto the hyperplane.**

This is a unique property of convex sets. In general, non-convex sets may not even guarantee the existence of a supporting point on a given hyperplane.

References