Introduction to Convex Optimization

Lecture 3: Convex Set

Silin DU

Department of Management Science and Engineering
Tsinghua University
dsl21@mails.tsinghua.edu.cn

July 29, 2022

Contents

Lecture Overview

In this lecture, we focus on some definitions and properties in convex sets, which are the foundation of convex optimization.

1. Basic definitions: line and line segment, affine set, convex set, cone, hyperplane, polyhedron, etc.
2. Operations that preserve convexity.
3. Generalized inequalities.
4. Dual cones, minimum and minimal.
5. Projection onto convex sets.
6. Optimization over convex set.
7. Separating hyperplane and supporting hyperplane.

We put some proofs in appendix.

Contents

Line and Line Segment

Definition 1 (Line)

Suppose x_1 and x_2 are two points in $\mathbb{R}^n$, and $x_1 \neq x_2$. Then $y = \theta x_1 + (1 - \theta)x_2, \theta \in \mathbb{R}$ forms a line passing through x_1 and x_2.

Expressing y in the form

$$y = x_2 + \theta(x_1 - x_2)$$

gives another interpretation: y is the sum of the *base point* x_2 and the *direction* $x_1 - x_2$ (which points from x_2 to x_1) scaled by the parameter θ.

Definition 2 (Line Segment)

Suppose x_1 and x_2 are two points in $\mathbb{R}^n$, and $x_1 \neq x_2$. Points of the form $y = \theta x_1 + (1 - \theta)x_2, 0 \leq \theta \leq 1$ are the line segment connecting x_1 and x_2.

Affine Set I

Definition 3 (Affine Set)

A set $\mathcal{C} \subset \mathbb{R}^n$ is affine if the line through any two distinct points in $\mathcal{C}$ lies within $\mathcal{C}$, i.e., if $x_1, x_2 \in \mathcal{C}, x_1 \neq x_2$ and $\theta \in \mathbb{R}$, then $\theta x_1 + (1-\theta)x_2 \in \mathcal{C}$.

- We refer to a point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$, where $\theta_1 + \cdots + \theta_k = 1$, as an *affine combination* of the points $x_1, ..., x_k$.
- An affine set contains every affine combination of its points.
- The solution set of a system of linear equations, $C = \{x \mid Ax = b\}$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, is an affine set. Suppose $Ax_1 = b, Ax_2 = b$. Then for any θ, we have

$$A(\theta x_1 + (1-\theta)x_2) = \theta A x_1 + (1-\theta)Ax_2 = b$$

Conversely, every affine set can be expressed as the solution set of a system of linear equations.

Affine Set II

Definition 4 (Affine Hull)

The set of all affine combinations of points in some set $\mathcal{C} \subset \mathbb{R}^n$ is called the affine hull of $\mathcal{C}$, and denoted $\mathbf{aff}(\mathcal{C})$:

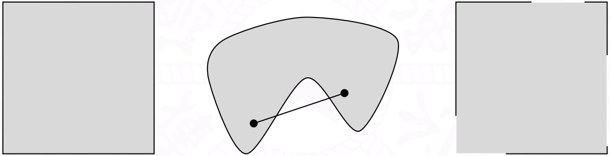
$$\mathbf{aff}(\mathcal{C}) = \{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_1, ..., x_k \in \mathcal{C}, \theta_1 + \cdots + \theta_k = 1\}$$

The affine hull is the smallest affine set that contains $\mathcal{C}$.

Definition 5 (Convex Set)

A set $\mathcal{C}$ is convex if the line segment between any two points in $\mathcal{C}$ lies in $\mathcal{C}$, i.e., if for any $x_1, x_2 \in \mathcal{C}$ and any θ with $0 \le \theta \le 1$, we have $\theta x_1 + (1-\theta)x_2 \in \mathcal{C}$.

Examples:



($\surd$) ($\times$) ($\times$)

- Roughly speaking, a set is convex if every point in the set can be seen by every other point, along an unobstructed straight path between them, where unobstructed means lying in the set.
- Every affine set is convex.

Convex Set II

Definition 6 (Convex Combination)

We refer to a point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$, where $\sum_{i=1}^{k} \theta_i = 1, \theta_i \geq 0$, as a convex combination of points $x_1, ..., x_k$.

- A set is convex if and only if it contains every convex combination of its points.

Definition 7 (Conic Combination)

Conic combination of x_1 and x_2 is $\theta_1 x_1 + \theta_2 x_2, \theta_1, \theta_2 \geq 0$.

Convex Set III

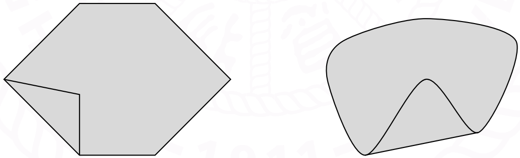
Definition 8 (Convex Hull)

The convex hull of a set $\mathcal{S}$, $\mathbf{conv}(\mathcal{S})$, is the set of all convex combinations of points in $\mathcal{S}$:

$$\mathbf{conv}(\mathcal{S}) = \left\{ \sum_{i=1}^{k} \theta_i x_i \mid x_i \in \mathcal{S}, \theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1 \right\}$$

- The convex hull is the smallest convex set that contains $\mathcal{S}$.

Examples:

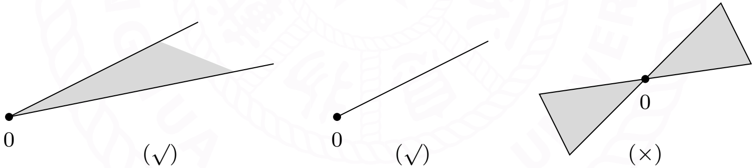


Cone I

Definition 9 (Cone)

A set $\mathcal{C}$ is called a cone, or nonnegative homogenous, if for every $x \in \mathcal{C}$ and $\theta \geq 0$ we have $\theta x \in \mathcal{C}$. A set $\mathcal{C}$ is called a convex cone if it is convex and a cone, which means for any $x_1, x_2 \in \mathcal{C}$ and θ_1, θ_2, we have $\theta x_1 + \theta_2 x_2 \in \mathcal{C}$. In other words, a convex cone is a set that contains all conic combinations of points in the set.

Examples:



Non-convex cone?

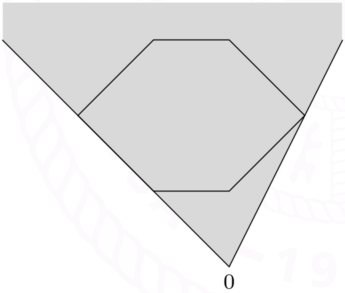
Cone II

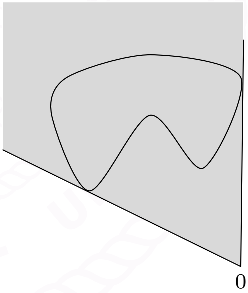
Definition 10 (Conic Hull)

The conic hull of a set $\mathcal{C}$ is the set of all conic combinations of points in $\mathcal{C}$, i.e.,

$$\{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in \mathcal{C}, \theta_i \geq 0, i = 1, ..., k\}$$

which is also the smallest convex cone that contains $\mathcal{C}$.





Some Important Examples

Some important examples of convex sets:

- The empty set $\emptyset$, any single point $\{x_0\}$, and the whole space $\mathcal{R}^n$ are affine (hence, convex) subsets of $\mathbb{R}^n$.
- Any line is affine. If it passes through zero, it is a subspace, hence also a convex cone.
- A line segment is convex, but not affine (unless it reduces to a point).
- A ray, which has the form $\{x_0 + \theta x \mid \theta > 0\}$, where $v \neq 0$, is convex, but not affine. It is a convex cone if its base x_0 is 0.
- Any subspace is affine, and a convex cone (hence convex).

Hyperplanes and Halfspaces

Definition 11 (Hyperplanes and Halfspaces)

For $a \in \mathbb{R}^n, a \neq 0$ and $b \in \mathbb{R}$:

1. $\mathcal{H}_{a,b} = \left\{x \in \mathbb{R}^n \mid a^\top x \leq b\right\}$ defines a halfspace;
2. $\mathcal{H}_{a,b} = \left\{x \in \mathbb{R}^n \mid a^\top x = b\right\}$ defines a hyperplane;

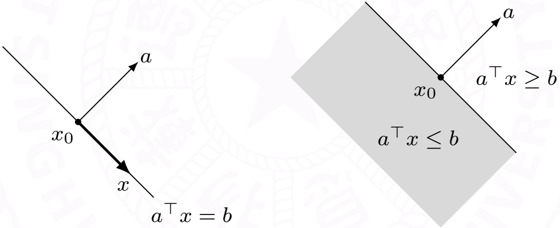


Fig. 1: Hyperplanes and Halfspaces

- The set $\left\{a^\top x < b\right\}$, which is the interior of the halfspace $\left\{x \mid a^\top x \leq b\right\}$, is called an open halfspace.

Euclidean Balls and Ellipsoids I

Definition 12 (Euclidean Ball)

A Euclidean ball in $\mathbb{R}^n$ has the form

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \left\{x \mid (x - x_c)^\top (x - x_c) \leq r^2\right\}$$

where $r > 0$, and $\|\cdot\|_2$ denotes the Euclidean norm. The vector x_c is the center of the ball and the scalar r is its radius. Another common representation for the Euclidean ball is

$$B(x_c, r) = \{x_c + ru \mid \|u\|_2 \leq 1\}$$

- A Euclidean ball is a convex set.

Definition 13 (Ellipsoids)

An ellipsoid has the form

$$\mathcal{E} = \left\{x \mid (x - x_c)^\top P^{-1}(x - x_c) \le 1\right\} \tag{1}$$

where P is symmetric and positive definite.

- The matrix P determines how far the ellipsoid extends in every direction from x_c; the lengths of the semi-axes of $\mathcal{E}$ are given by $\sqrt{\lambda_i}$, where λ_i are the eigenvalues of P.
- A ball is an ellipsoid with $P = r^2 I$.
- Another common representation of an ellipsoid is

$$\mathcal{E} = \{x_c + Au \mid \|u\|_2 \le 1\} \tag{2}$$

where A is square and nonsingular. By taking $A = P^{1/2}$, this representation gives the ellipsoid defined in Eq. 1. When A is symmetric positive semidefinite but singular, the set in Eq. 2 is called a degenerate ellipsoid; its affine dimension is equal to the rank of A. Degenerate ellipsoid are also convex.

Norm Balls and Norm Cones I

Definition 14 (Norm Ball)

Suppose $\|\cdot\|$ is any norm on $\mathbb{R}^n$. A norm ball (with center x_c and radius r) defined based on $\|\cdot\|$ is:

$$\{x \mid \|x - x_c\| \leq r\}$$

Definition 15 (Norm Cone)

Suppose $\|\cdot\|$ is any norm on $\mathbb{R}^n$. A norm cone defined based on $\|\cdot\|$ is:

$$\{(x, t) \mid \|x\| \leq t\} \subseteq \mathbb{R}^{n+1}$$

- Norm cones are convex.
- The second-order cone is the norm cone for Euclidean norm.
- The second-order cone is also called ice-cream cone or Lorentz cone.

Norm Balls and Norm Cones II

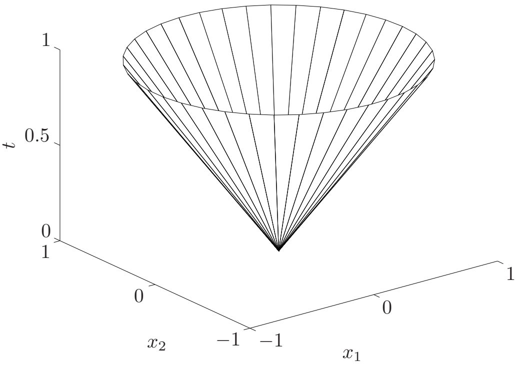


Fig. 2: Boundary of second-order connecting in $\mathbb{R}^3$, $\left\{(x_1, x_2, t) \mid \left(x_1^2 + x_2^2\right)^{1/2} \leq t\right\}$

Polyhedron I

Definition 16 (Polyhedron)

A polyhedron is defined as the solution set of a finite number of linear equalities and inequalities:

$$P = \left\{ x \mid a_j^\top x \leq b_j, j = 1, ..., m; c_i^\top x = d_i, i = 1, ..., p \right\}$$

or

$$P = \{ x \mid Ax \preceq b, Cx = d \}$$

where

$$A = \begin{bmatrix} a_1^\top \\ \vdots \\ a_m^\top \end{bmatrix}, \quad C = \begin{bmatrix} c_1^\top \\ \vdots \\ c_p^\top \end{bmatrix}$$

Polyhedron II

- A polyhedron is the intersection of a finite number of halfspaces and hyperplanes.
- Affine sets (e.g., subspaces, hyperplanes, lines), rays, line segments, and halfspaces are all polyhedra.
- Polyhedra are convex sets.
- A bounded polyhedron is sometimes called a polytope.
- The nonnegative orthant is the set of points with nonnegative components, i.e.,

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_i \geq 0, i = 1, ..., n\} = \{x \in \mathbb{R}^n \mid x \succeq 0\}$$

The nonnegative orthant is a polyhedron and a cone.

Polyhedron III

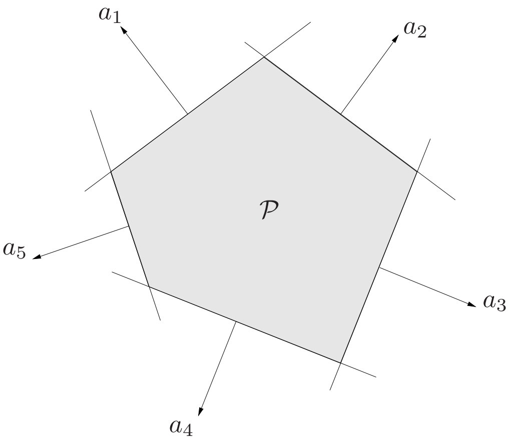


Fig. 3: The polyhedron $\mathcal{P}$ is the intersection of five halfspaces, with normal vectors $a_1, \ldots, a_5$.

Convex Hull Description of Polyhedra

A generalization of the convex hull description is

$$\{\theta_1 v_1 + \cdots + \theta_k v_k \mid \theta_1 + \cdots + \theta_m = 1, \theta_i \geq 0, i = 1, \cdots k\} \tag{3}$$

where $m \leq k$. This is the convex hull of the points $v_1, ..., v_m$ plus the conic hull of the points $v_{m+1}, ..., v_l$. The set in Eq. 3 defines a polyhedron, and conversely, every polyhedron can be represented in this form.

Simplexes

Definition 17 (Simplex)

Suppose the $k+1$ points $v_0, \ldots, v_k \in \mathbb{R}^n$ are affinely independent, which means $v_1 - v_0, \ldots, v_k - v_0$ are linearly independent. The simplex determined by them is given by

$$C = \mathbf{conv}\{v_0, \ldots, v_k\} = \left\{\theta_0 v_0 + \cdots + \theta_k v_k \mid \theta \succeq 0, \mathbf{1}^\top \theta = 1\right\}$$

where $\mathbf{1}$ denotes the vector with all entries one. The affine dimension of this simplex is k.

- The unit simplex is the n-dimensional simplex that can be expressed as the set of vectors that satisfy

$$x \succeq 0, \mathbf{1}^\top x \leq 1$$

- The probability simplex is the $(n-1)$-dimensional simplex that can be expressed as the set of vectors that satisfy

$$x \succeq 0, \mathbf{1}^\top x = 1$$

Positive Semidefinite Cone I

Definition 18 (Positive Semidefinite Cone)

The positive semidefinite cone $\mathcal{S}_+^n$ is defined as:

$$\mathcal{S}_+^n = \{X \mid X \succeq 0, X \in \mathcal{S}^n\}.$$

- $\mathcal{S}^n$ is the set of $n \times n$ symmetric matrices.
- $\mathcal{S}_+^n$ is the set of $n \times n$ positive semidefinite matrix and is convex.

$$\forall x \in \mathcal{S}_+^n \Leftrightarrow z^\top X z \geq 0, \forall x \in \mathbb{R}^n$$

.

- $\mathcal{S}_{++}^n$ is the set of $n \times n$ positive definite matrix

Positive Semidefinite Cone II

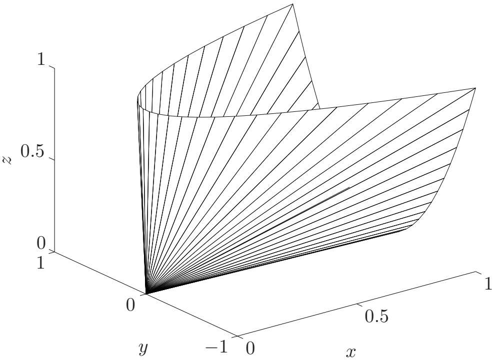


Fig. 4: Boundary of positive semidefinite cone in $\mathcal{S}^2$.

Copositive Cone

Definition 19 (Copositive Cone)

The copositive cone $\mathcal{C}_n$ is defined as:

$$\mathcal{C}_n = \left\{ X \mid X \in \mathcal{S}^n, y^\top X y \geq 0, \forall y \geq 0 \right\}$$

Definition 20 (Completely Positive Matrices Cone)

The completely positive matrices cone $\mathcal{CP}_n$ is defined as:

$$\mathcal{CP}_n = \left\{ X \mid X = \sum_{i=1}^{k} y_i y_i^\top, y_i \in \mathbb{R}^n, y_i \geq 0, i = 1, 2, ..., k \right\}$$

Contents

Intersection I

Theorem 1 (Intersection)

The intersection of (any number of) convex sets is convex. If $\mathcal{S}_\alpha$ is convex for every $\alpha \in \mathcal{A}$, then $\bigcap_{\alpha \in \mathcal{A}} \mathcal{S}_\alpha$ is convex.

- A polyhedron is the intersection of halfspaces and hyperplanes (which are convex), and therefore is convex.
- The positive semidefinite cone $\mathcal{S}_+^n$ can be expressed as

$$\bigcap_{z \neq 0} \left\{ X \in \mathcal{S}^n \mid z^\top X z \geq 0 \right\}$$

For each $z \neq 0$, $z^\top X z$ is a (identically zero) linear function of X, so the sets

$$\left\{ X \in \mathcal{S}^n \mid z^\top X z \geq 0 \right\}$$

are, in fact, halfspaces in $\mathcal{S}^n$. Thus the positive semidefinite cone is the intersection of an infinite number of halfspaces, and so is convex.

Intersection II

- Consider the set

$$\mathcal{S} = \{x \in \mathbb{R}^m \mid |p(t)| \leq 1 \text{ for } |t| \leq \pi/3\}$$

where $p(t) = \sum_{k=1}^{m} x_k \cos kt$. The set $\mathcal{S}$ can be expressed as the intersection of an infinite number of the following sets

$$\mathcal{S}_t = \left\{x \mid -1 \leq [\cos t, \cdots, \cos mt]^\top x \leq 1\right\}$$

and so is convex.

Affine Function I

Theorem 2 (Affine Function)

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine if it is a sum of a linear function and a constant, i.e., if it has the form $f(x) = Ax + b$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Suppose $\mathcal{S} \subseteq \mathbb{R}^n$ is convex and $f : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function. Then the image of $\mathcal{S}$ under f,

$$f(\mathcal{S}) = \{f(x) \mid x \in \mathcal{S}\},$$

is convex. Similarly, if $f : \mathbb{R}^k \to \mathbb{R}^n$ is an affine function, the inverse image of $\mathcal{S}$ under f,

$$f^{-1}(\mathcal{S}) = \{x \mid f(x) \in \mathcal{S}\}$$

is convex.

- If $\mathcal{S} \subseteq \mathbb{R}^n$ is convex, $\alpha \in \mathbb{R}$, and $a \in \mathbb{R}^n$, then the sets $\alpha\mathcal{S}$ and $\mathcal{S} + a$ are convex, where

$$\alpha\mathcal{S} = \{\alpha x \mid \mathcal{S}\}, \quad \mathcal{S} + a = \{x + a \mid x \in \mathcal{S}\}$$

Affine Function II

- The sum of two sets is defined as

$$\mathcal{S}_1 + \mathcal{S}_2 = \{x + y \mid x \in \mathcal{S}_1, y \in \mathcal{S}_2\}$$

  If $\mathcal{S}_1$ and $\mathcal{S}_2$ are convex, then $\mathcal{S}_1 + \mathcal{S}_2$ is convex. To see this, if $\mathcal{S}_1$ and $\mathcal{S}_2$ are convex, then so is the Cartesian product

$$\mathcal{S}_1 \times \mathcal{S}_2 = \{(x_1, x_2) \mid x_1 \in \mathcal{S}_1, x_2 \in \mathcal{S}_2\}$$

  The image of this set under the linear function $f(x_1, x_2) = x_1 + x_2$ is the sum $\mathcal{S}_1 + \mathcal{S}_2$.

- Polyhedron. The polyhedron $\{x \mid Ax \preceq b, Cx = d\}$ can be expressed as the inverse image of the Cartesian product of the nonnegative orthant and the origin under the affine function $f(x) = (b - Ax, d - Cx)$

$$\{x \mid Ax \preceq b, Cx = q\} = \left\{x \mid f(x) \in \mathbb{R}_+^m \times \{0\}\right\}$$

- The condition

$$A(x) = x_1 A_x + \cdots + x_n A_n \preceq B$$

where $B, A_i \in \mathcal{S}^m$, is called a linear matrix inequality (LMI) in x. The solution set of a LMI,$\{x \mid A(x) \preceq B\}$, is convex. Indeed, it is the inverse image of the positive semidefinite cone under the affine function $f : \mathbb{R}^n \to \mathcal{S}^m$ given by $f(x) = B - A(x)$.

- Ellipsoid. The ellipsoid

$$\mathcal{E} = \left\{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leq 1\right\}$$

where $P \in \mathcal{S}^n{++}$, is the image of the unit Euclidean ball $\{u \mid \|u\|_2 \leq 1\}$ under the affine mapping $f(u) = P^{1/2}u + x_c$. It is also the inverse image of the unit ball under the affine mapping $g(x) = P^{-1/2}(x - x_c)$.

Perspective Functions I

Definition 21 (Perspective Function)

We define the perspective function $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$, with domain $\mathbf{dom} P = \{(x,t) \mid x \in \mathbb{R}^n, t \in \mathbb{R}_{++}\}$, as $P(x,t) = x/t$. The perspective function scales or normalizes vectors so the last component is one, and then drops the last component.

Theorem 3

If $\mathcal{C} \subseteq \mathbf{dom} P$ is convex, then its image under the perspective function

$$P(C) = \{P(x) \mid x \in \mathcal{C}\}$$

is convex.

Perspective Functions II

We show that a line segments are mapped to line segments under the perspective function. Suppose that $x = (\tilde{x}, x_{n+1}), y = (\tilde{y}, y_{n+1}) \in \mathbb{R}^{n+1}$ with $x_{n+1} > 0, y_{n+1} > 0$. Then for $0 \leq \theta \leq 1$,

$$P(\theta x + (1-\theta)y) = \frac{\theta\tilde{x} + (1-\theta)\tilde{y}}{\theta x_{n+1} + (1-\theta)y_{n+1}} = \mu P(x) + (1-\mu)P(y)$$

where

$$\mu = \frac{\theta x_{n+1}}{\theta x_{n+1} + (1-\theta)y_{n+1}} \in [0,1]$$

As θ varies between 0 and 1 ($\theta x + (1-\theta)y$ sweeps out the line segment $[x, y]$), μ varies between 0 and 1 ($P(\theta x + (1-\theta)y)$ sweeps out the line segment $[P(x), P(y)]$).
Now suppose C is convex with $C \subseteq \textbf{dom}P$ (i.e., $x_{n+1} > 0$ for all $x \in C$), and $x, y \in C$. We need to show that the line segment $[P(x), P(y)]$ is in $P(C)$.The line segment $[P(x), P(y)]$ is the line segment $[x, y]$ under P, so lies in $P(C)$.

Theorem 4

The inverse image of a convex set under the perspective function is also convex: if $C \subseteq \mathbb{R}^n$ is convex, then

$$P^{-1}(C) = \left\{(x, t) \in \mathbb{R}^{n+1} \mid x/t \in C, t > 0\right\}$$

is convex.

Suppose $(x, t) \in P^{-1}(C), (y, s) \in P^{-1}(C)$, and $0 \leq \theta \leq 1$. We need to show that

$$\theta(x, t) + (1-\theta)(y, s) \in P^{-1}(C) \Longleftrightarrow \frac{\theta x + (1-\theta)y}{\theta t + (1-\theta)s} \in C$$

This follows from

$$\frac{\theta x + (1-\theta)y}{\theta t + (1-\theta)s} = \mu(x/t) + (1-\mu)(y/s)$$

where

$$\mu = \frac{\theta t}{\theta t + (1-\theta)s} \in [0, 1]$$

Linear-fractional Functions I

Definition 22 (Linear-fractional Functions)

A linear-fractional function is formed by composing the perspective function with an affine function. Suppose $g : \mathbb{R}^n \to \mathbb{R}^{m+1}$ is affine, i.e.,

$$g(x) = \begin{bmatrix} A \\ c^\top \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix}$$

where $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$, and $d \in \mathbb{R}$. The function $f : \mathbb{R}^n \to \mathbb{R}^m$ given by $f = P \circ g$, i.e.,

$$f(x) = \frac{(Ax+b)}{(c^\top x + d)}, \quad \mathbf{dom} f = \left\{x \mid c^\top x + d > 0\right\}$$

is called a linear-fractional (or projective) function.

- If $c = 0$ and $d > 0$, the domain of f is $\mathbb{R}^n$, and f is an affine function. So we can think of affine and linear functions as special cases of linear-fractional functions.

Linear-fractional Functions II

Theorem 5

If C is convex and lies in the domain of the linear-fractional function f (i.e., $c^\top x + d > 0$ for all $x \in C$), then its image $f(C)$ is convex. Similarly, if $C \subseteq \mathbb{R}^m$ is convex, then the inverse image $f^{-1}(C)$ is convex.

- The image of C under the affine function is convex, and the image of the resulting set under the perspective function P, which yields $f(C)$, is convex.
- Conditional probabilities.
  Suppose u and v are random variables that take on values in $\{1, \dots, n\}$ and $\{1, \dots, m\}$, respectively, and let p_{ij} denote $\mathbf{prob}(u = i, v = j)$. Then the conditional probability $f_{ij} = \mathbf{prob}(u = i \mid v = j)$ is given by

$$f_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n} p_{kj}}$$

  Thus f is obtained by a linear-fractional mapping from p. It follows that if C is a convex set of joint probabilities for (u, v), then the associated set of conditional probabilities of u given v is also convex.

Contents

Partial Order

Definition 23 (Partial Order)

A (non-strict) partial order is a binary relation $\leq$ over a set P, satisfying particular axioms: reflexivity, antisymmetry, and transitivity.

- The vector inequality is defined as follows.

$$x \geq y, x, y \in \mathbb{R}^n \iff x_i \geq y_i, \forall i = 1, \cdots n$$

The coordinate-wise partial order satisfies:

- reflexivity: $a \geq a$.
- antisymmetry: if both $a \geq b$ and $b \geq a$, then $a = b$.
- transitivity: if both $a \geq b$ and $b \geq c$, then $a \geq c$.
- compatibility of linear operation:
  (a) homogeneity: if $a \geq b$ and λ is a non-negative real number, then $\lambda a \geq \lambda b$.
  (b) additivity: if $a \geq b$ and $c \geq d$, then $a + c \geq b + d$.

Proper Cone

Definition 24 (Proper Cone)

A cone $K \subseteq \mathbf{R}^n$ is called a proper cone if it satisfies the following:

- K is convex.
- K is closed.
- K is solid, which means it has nonempty interior.
- K is pointed, which means that it contains no line (or equivalently, $x \in K, -x \in K \Longrightarrow x = 0$).

Example:

- Nonnegative orthant.
- Positive semidefinite cone.
- Cone of polynomials nonnegative on $[0,1]$:

$$K = \left\{c \in \mathbb{R}^n \mid c_1 + c_2 t + \cdots c_n t^{n-1}, t \in [0,1]\right\}$$

Generalized Inequalities

Definition 25 (Generalized Inequalities)

A proper cone K can be used to define generalized inequalities:

$$x \preceq_K y \Longleftrightarrow y - x \in K$$

and

$$x \prec_K y \Longleftrightarrow y - x \in \mathrm{int} K$$

- Component-wise inequality ($K = \mathbb{R}^n_+$)

$$x \preceq_{\mathbb{R}^n_+} y \Longleftarrow x_i \leq y_i, \quad i = 1, \cdots, n$$

- Matrix inequality ($K = \mathcal{S}^n_+$)

$$X \preceq_{\mathcal{S}^n_+} Y \Longleftarrow Y - X \text{ is positive semidefinite}$$

- Many properties of $\preceq_K$ are similar to $\leq$ on $\mathbb{R}$, e.g.

$$x \preceq_K y, u \preceq_K v \Longrightarrow x + u \preceq_K y + v$$

Many properties of ordinary inequality on $\mathbb{R}$ (i.e., $\leq, <$) do hold for generalized inequalities (i.e., $\preceq_K, \prec_K$) , while some important ones do not.
$\leq$ on $\mathbb{R}$ is a linear ordering: any two points are comparable, meaning either $x \leq y$ or $y \leq x$. This property does not hold for other generalized inequalities.

Definition 26 (Minimum)

$x \in S$ is the minimum element of S (with respect to the generalized inequality $\preceq_K$), if

$$\forall y \in S \rightarrow x \preceq_K y$$

Definition 27 (Minimal)

$x \in S$ is a minimal element of S (with respect to the generalized inequality $\preceq_K$), if

$$y \in S, y \preceq_K x \Rightarrow y = x$$

Minimum and Minimal II

- In general, minimum is unique (if exists), while minimal points are not unique.
- We can define maximum and maximal points similarly.
- A point $x \in S$ is the minimum element of S if and only if

$$S \subseteq x + K$$

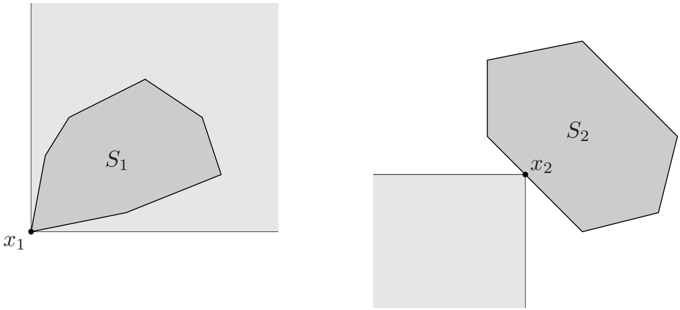
Here $x + K$ denotes all the points that are comparable to x and greater than or equal to x (according to $\preceq_K$).

- A point $x \in S$ is a minimal element if and only if

$$(x - K) \cap S = \{x\}$$

Here $x - k$ denotes all the points that are comparable to x and less than or equal to x (according to $\preceq_K$); the only point in common with S is x.

Minimum and Minimal III



- The set S_1 has a minimum element x_1 with respect to component-wise inequality in $\mathbb{R}^2$. The set $x_1 + K$ is shaded lightly; x_1 is the minimum element of S_1 since $S_1 \subseteq x_1 + K$.
- The point x_2 is a minimal point of S_2. The set $x_2 - K$ is shown lightly shaded. The point x_2 is minimal because $x_2 - K$ and S_2 intersect only at x_2.

Contents

Dual Cone I

Definition 28 (Dual Cone)

Dual cone of a cone K is defined as

$$K^* = \left\{ y \mid y^\top x \geq 0, \forall x \in K \right\}$$

- The dual cone of a subspace $V \subseteq \mathbb{R}^n$ (which is a cone) is its orthogonal complement $V^\perp = \{y \mid v^\top y = 0, \text{ for all } v \in V\}$.
- The cone $\mathbb{R}^n_+$ is its own dual:

$$x^\top y \geq 0 \text{ for all } x \succeq 0 \iff y \succeq 0$$

We call such a cone self-dual.

Dual Cone II

- The positive semidefinite cone $\mathcal{S}_+^n$ is self-dual, i.e.,

$$\left(\mathcal{S}_+^n\right)^* = \{Y \mid \operatorname{tr}(XY) \geq 0 \text{ for all } X \succeq 0\} = \{Y \mid Y \succeq 0\}.$$

Suppose $Y \notin \mathcal{S}_+^n$. Then there exists $q \in \mathbb{R}^n$ with

$$q^\top Y q = \operatorname{tr}\left(q^\top Y q\right) = \operatorname{tr}\left(qq^\top Y\right) < 0$$

Hence the positive semidefinite matrix $X = qq^\top$ satisfies $\operatorname{tr}(XY) < 0$; it follows that $Y \notin \left(\mathcal{S}_+^n\right)^*$.
Now suppose $X, Y \in \mathbf{S}_+^n$. We can express X in terms of its eigenvalue decomposition as $X = \sum_{i=1}^n \lambda_i q_i q_i^\top$, where (the eigenvalues) $\lambda_i \geq 0, i = 1, \ldots, n$. Then we have

$$\operatorname{tr}(YX) = \operatorname{tr}\left(Y \sum_{i=1}^n \lambda_i q_i q_i^\top\right) = \sum_{i=1}^n \lambda_i q_i^\top Y q_i \geq 0$$

This shows that $Y \in \left(\mathbf{S}_+^n\right)^*$.

Dual Cone III

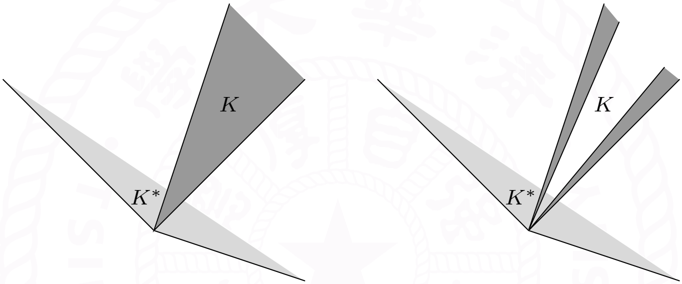


Fig. 5: Dual cone examples

Properties of dual cones:

1. K^* is closed and convex (even if K is not).
2. If $K_1 \subseteq K_2$, then $K_2^* \subseteq K_1^*$.
3. If the closure of K is pointed, then K^* has non-empty interior.
4. If K has nonempty interior, then $K^*$is pointed.

Dual Cone IV

5. K^{**} is the closure of the convex hull of K. If K is convex and closed, then $K^{**} = K$.

6. Dual cone of a proper cone is proper.

Theorem 6

The dual of the norm cone $K = \left\{(x,t) \in \mathbb{R}^{n+1} \mid \|x\| \le t\right\}$ associated with a norm $\|\cdot\|$ in $\mathbb{R}^n$ is the cone defined by the dual norm,

$$K^* = \left\{(u,s) \in \mathbb{R}^{n+1} \mid \|u\|_* \le s\right\},$$

where the dual norm is given by $\|u\|_* = \sup\left\{u^\top x \mid \|x\| \le 1\right\}$.

Dual Cone V

We need to show

$$x^\top u + ts \geq 0, \forall \|x\| \leq t \Longleftrightarrow \|u\|_* \leq s$$

- $\Leftarrow$.
  Suppose $\|u\|_* \leq s$ and $\|x\| \leq t$ for some $t > 0$ (the case $t = 0$ is trivial), by the definition of the dual norm and the fact that $\|-x/t\| \leq 1$, we know that

  $$u^\top(-x/t) \leq \|u\|_* \leq s,$$

  Hence, $u^\top x + ts \geq 0$.
- $\Rightarrow$.
  Suppose $\|u\|_* > s$ (the right hand condition dose not hold), then by the definition of dual norm, $\exists w$ with $\|w\| \leq 1$ and $w^\top u > s$. For the left side, take $x = -w$ and $t = 1$ then we have

  $$u^\top(-w) + s < 0$$

  which contradicts with the left hand condition.

Dual Generalized Inequalities I

- Suppose that the cone K is proper, so it induces a generalized inequality $\preceq_K$.
- Its dual cone K^* is also proper, and induces a generalized inequality.
- We refer to the generalized inequality $\preceq_{K^*}$ as the dual of the generalized inequality $\preceq_K$.

Some important properties relating a generalized inequality and its dual are:

- $x \preceq_K y$ if and only if $\lambda^\top x \le \lambda^\top y$ for all $\lambda \succeq_{K^*} 0$.
- $x \prec_K y$ if and only if $\lambda^\top x < \lambda^\top y$ for all $\lambda \succeq_{K^*} 0, \lambda \ne 0$.

Dual Generalized Inequalities II

- Suppose $K \subseteq \mathbf{R}^m$ is a proper cone. Consider the strict generalized inequality

$$Ax \prec_K b \tag{4}$$

where $x \in \mathbb{R}^n$.

- Suppose it is infeasible, i.e., the affine set $\{b - Ax \mid x \in \mathbf{R}^n\}$ does not intersect the open convex set $\text{int}K$.
- Then there is a separating hyperplane, i.e., a nonzero $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}$ such that $\lambda^\top(b - Ax) \leq \mu$ for all x, and $\lambda^\top y \geq \mu$ for all $y \in \text{int}K$.
- The first condition implies $A^\top \lambda = 0$ and $\lambda^\top b \leq \mu$. The second condition implies $\lambda^\top y \geq \mu$ for all $y \in K$, which can only happen if $\lambda \in K^*$ and $\mu \leq 0$.
- Putting it all together we find that if Eq. 4 is infeasible, then there exists λ such that

$$\lambda \neq 0, \quad \lambda \succeq_{K^*} 0, \quad A^\top \lambda = 0, \quad \lambda^\top b \leq 0. \tag{5}$$

- If Eq. 5 holds, then the inequality system in Eq. 4 cannot be feasible. Suppose that both inequality systems hold. Then we have $\lambda^\top(b - Ax) > 0$, since $\lambda \neq 0, \lambda \succeq_{K^*} 0$, and $b - Ax \succ_K 0$. But using $A^\top \lambda = 0$ we find that $\lambda^\top(b - Ax) = \lambda^\top b \leq 0$, which is a contradiction.
- For any data A, b, exactly one of them is feasible.

Minimum Elements via Dual Inequalities I

x is minimum element of S with respect to $\preceq_K$, if and only if for all $\lambda \succ_{K^*} 0, x$ is the unique minimizer of $\lambda^\top z$ over S. Note convexity of S is not required.

- $\Rightarrow$. x is minimum element of S with respect to $\preceq_K$.
  If $\lambda^\top y < \lambda^\top x$, for some $y \in S$, then $\lambda^\top(x-y) > 0$. From $\lambda \succeq_{K^*} 0$, we have $x - y \succeq_K 0 \Rightarrow y \preceq_K x$, which means x is not the minimum.
- $\Leftarrow$. For all $\lambda \succ_{K^*} 0, x$ is the unique minimizer of $\lambda^\top z$ over S.
  Suppose there exists some $y, y \preceq_K x \Rightarrow (x-y) \succeq_K 0 \Rightarrow \lambda^\top(x-y) \geq 0 \Rightarrow \lambda^\top x \geq \lambda^\top y$, thus x is not the unique minimizer.

Geometrically, for any $\lambda \succ_{K^*} 0$, the hyperplane

$$\left\{z \mid \lambda^\top (z - x) = 0\right\}$$

is a strict supporting hyperplane to S at x. By strict supporting hyperplane, we mean that the hyperplane intersects S only at the point x.

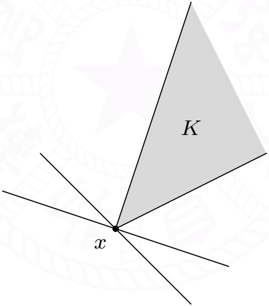


Fig. 6: The point x is the minimum element of the set S with respect to $\mathbb{R}^2_+$.

Minimal Elements via Dual Inequalities I

If x minimizes $\lambda^\top z$ over S for some $\lambda_{K^*} 0$, or $\lambda \in \operatorname{int} K^*$, then x is a minimal element of S with respect to $\preceq_K$. This is shown in Figure 7.

- Suppose that there exists $y \neq x$ such that $y \preceq_K x$, then $x - y \succeq_K 0$. Since $\lambda \succ_{K^*} 0$, $\lambda^\top (x - y) > 0$. And we know $x \neq y$, so x is not the minimizer, which contradicts with the condition.
- The converse is general false: a point x can be minimal in S, but not a minimizer of $\lambda^\top z$ over $z \in S$, for any λ, as shown in Figure 8.
- This figure suggests that convexity plays an important role in the converse, which is correct.

Minimal Elements via Dual Inequalities II

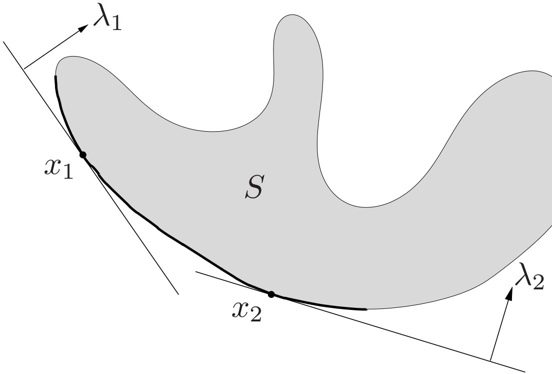


Fig. 7: A set $S \subseteq \mathbb{R}^2$. Its set of minimal points, with respect to $\mathbb{R}^2_+$, is shown as the darker section of its (lower, left) boundary. The minimizer of $\lambda_1^T z$ over S is x_1, and is minimal since $\lambda_1 \succ 0$. The minimizer of $\lambda_2^T z$ over S is x_2, which is another minimal point of S, since $\lambda_2 \succ 0$.

Minimal Elements via Dual Inequalities III

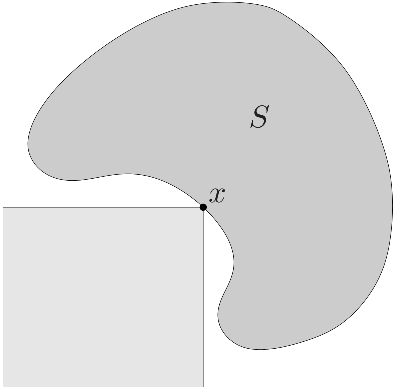


Fig. 8: The point x is a minimal element of $S \subseteq \mathbb{R}^2$ with respect to $\mathbb{R}^2_+$. However there exists no λ for which minimizes $\lambda^\top$ over $z \in S$.

Minimal Elements via Dual Inequalities IV

If x is a minimal element of a *convex set* S with respect to $\preceq_K$, then there exists a non-zero $\lambda \succeq_{K^*} 0$, such that x minimizes $\lambda^\top z$ over S.

- Suppose x is minimal,
$$((x - K)\backslash\{x\}) \cap S = \emptyset$$
- Apply the separating hyperplane theorem to the convex sets $((x - K)\backslash\{x\})$ and S, then there $\exists \lambda \neq \mathbf{0}$ and μ, such that $\lambda^\top (x - y) \leq \mu$ for $\forall y \in K$ and $\lambda^\top z \geq \mu$ for $\forall z \in S$.
- From the first inequality, we have
$$\lambda^\top y \geq \lambda^\top x - \mu \geq 0$$
since $\lambda^\top z \geq \mu, \forall z \in S$, which means $\lambda \succeq_{K^*} 0$. Since $x \in S, x \in x - K$, we have $\lambda^\top x = \mu$.
- The second inequality implies that μ is the minimum value of $\lambda^\top z$ over S. Therefore, x is a minimizer of $\lambda^\top z$ over $z \in S$.
- This cannot be strengthened to $\lambda \succ_{K^*} 0$.

Minimal Elements via Dual Inequalities V

Fig. 9: *Left.* The point $x_1 \in S_1$ is minimal, but is not a minimizer of $\lambda^\top z$ over S_1 for any $\lambda \succ 0$. (It does, however, minimize $\lambda^\top z$ over $z \in S_1$ for $\lambda = (1, 0)$.) *Right.* The point $x_2 \in S_2$ is not minimal, but it does minimize $\lambda^T z$ over $z \in S_2$ for $\lambda = (0, 1) \succeq 0$.

Contents

Definition 29 (Projection)

Let $z \in \mathbb{R}^n$ be some fixed vector, and set $C \subseteq \mathbb{R}^n$ be a non-empty, closed, and convex set. Then,

$$P_C(z) = \arg\min_{x \in C} \|z - x\|_2^2$$

is the projection of point z on set C.

Property 1

$\forall z \in \mathbb{R}^n, x^* = P_C(z) \Longleftrightarrow (z - x^*)^\top (x - x^*) \leq 0, \forall x \in C.$

- Let $f(x) = \frac{1}{2}\|x - z\|_2^2$, then $\nabla f(x) = x - z$. So the claim follows from

$$\begin{aligned} &\nabla f(x^*)^\top (x - x^*) \geq 0 \\ \Rightarrow &(x^* - z)^\top (x - x^*) \geq 0, \forall x \in C \\ \Rightarrow &(z - x^*)^\top (x - x^*) \geq 0, \forall x \in C \end{aligned}$$

Projection II

Property 2

For any $z \in \mathbb{R}^n$, projection is non-expansive, that is

$$\forall z, \hat{z} \in \mathbb{R}^n, \| P_C(z) - P_C(\hat{z}) \|_2 \leq \| z - \hat{z} \|_2$$

.

- First, we have
$$\begin{cases} (z - P_C(z))^\top (x - P_C(z)) \leq 0, \forall x \in C \\ (\hat{z} - P_C(\hat{z}))^\top (x - P_C(\hat{z})) \leq 0, \forall x \in C \end{cases}$$
- Since $P_C(z), P_C(\hat{z}) \in C$,
$$\begin{cases} (z - P_{\mathcal{C}}(z))^\top (P_{\mathcal{C}}(\hat{z}) - P_{\mathcal{C}}(z)) \leq 0 \\ (\hat{z} - P_{\mathcal{C}}(\hat{z}))^\top (P_{\mathcal{C}}(z) - P_{\mathcal{C}}(\hat{z})) \leq 0 \end{cases}$$
- By combining the inequalities above, we have
$$\begin{aligned} [(z - \hat{z}) - (P_{\mathcal{C}}(z) - P_{\mathcal{C}}(\hat{z}))]^\top (P_{\mathcal{C}}(\hat{z}) - P_{\mathcal{C}}(z)) \leq 0 \\ \|P_{\mathcal{C}}(z) - P_{\mathcal{C}}(\hat{z})\|_2^2 \leq (z - \hat{z})^\top (P_{\mathcal{C}}(z) - P_{\mathcal{C}}(\hat{z})) \\ \leq \|z - \hat{z}\|_2 \|P_{\mathcal{C}}(z) - P_{\mathcal{C}}(\hat{z})\|_2 \end{aligned}$$

  where the second $\leq$ in above inequality is due to Cauchy-Schwarz inequality.

Contents

Definition 30 (Local Min and Global Min)

Consider $\min_{x \in C} f(x)$ where the feasible region $C \in \mathbb{R}^n$ is convex and nonempty.

(1) Any vector $x \in C$ is called feasible.

(2) A feasible x^* is a local min if $f(x) \geq f(x^*), \forall x \in C \cap \mathcal{B}_\delta(x^*)$ for some $\delta > 0$. Here $\mathcal{B}_\delta(x^*) = \{y \mid \|y - x^*\|_2 \leq \delta\}$ is a δ-ball near x^*.

(3) A feasible x^* is a global min if $f(x) \geq f(x^*), \forall x \in C$.

Local Min and Global Min II

Let f be continuously differentiable over a closed, convex, non-empty set $C \subseteq \mathbb{R}^n$. If x^* is a local min, then $\nabla f(x^*)^\top (x - x^*) \geq 0, \forall x \in \mathcal{C} \cap \mathcal{B}_\delta(x^*)$

- Suppose that $\exists x$, s.t. $\nabla f(x^*)^\top (x - x^*) < 0$.
- Then by continuity $\nabla f(x^* + \bar{\alpha}(x - x^*))^\top (x - x^*) < 0, \forall \bar{\alpha} \in [0, \alpha]$, for some α small enough.
- By mean-value theorem (MVT), we have:

$$
\begin{aligned}
& f(x^* + \alpha(x - x^*)) - f(x^*) \\
= & \nabla f(x^* + \bar{\alpha}(x - x^*))^\top (x - x^*) < 0, \bar{\alpha} \in [0, \alpha] \\
\Rightarrow & f(x^* + \alpha(x - x^*)) < f(x^*) \text{ for some small } \alpha
\end{aligned}
$$

It contradicts to x^* being a local min.

Contents

Separating Hyperplane I

Theorem 7 (Separating Hyperplane Theorem)

Suppose C and D are nonempty disjoint convex sets, i.e., $C \cap D = \emptyset$. Then $\exists (a, b) \in \mathbb{R}^n \times \mathbb{R}, a \neq 0$, such that

$$\begin{cases} a^\top x \leq b, & \forall x \in C \\ a^\top x \geq b, & \forall x \in D \end{cases}$$

The hyperplane $\left\{x \mid a^\top x = b\right\}$ is called a separating hyperplane for the sets C and D, or is said to separate the sets C and D.

We provide a proof in Appendix 1. What's necessary to strictly separate two convex C and D?

Separating Hyperplane II

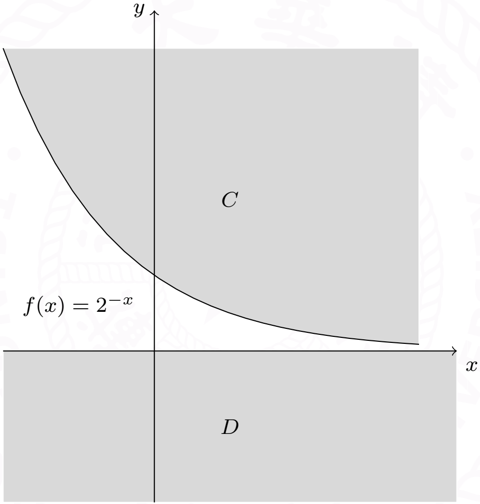


Fig. 10: Two closed convex sets without strictly separating hyperplane.

Separating Hyperplane III

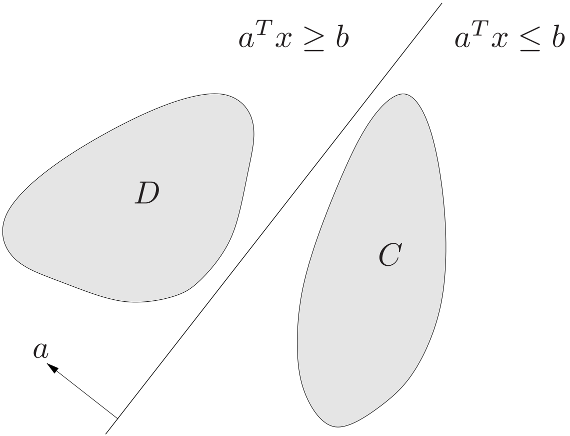


Fig. 11: The hyperplane $\left\{x \mid a^\top x = b\right\}$ separates the disjoint convex set C and D.

Strict Separation I

Theorem 8 (Strict Separation)

(1) If $\mathcal{C} \subseteq \mathbb{R}^n$ is a non-empty, closed, convex set and $x \notin \mathcal{C}, x \in \mathbb{R}^n$, then $\exists a \in \mathbb{R}^n$, such that

$$a^\top x > \sup_{y \in \mathcal{C}} \left(a^\top y\right).$$

(2) Separation of a disjoint point. If $\mathcal{C}, \mathcal{D}$ are non-empty, closed, convex, disjoint, and $\mathcal{D}$ is bounded, then $\exists a \in \mathbb{R}^n$, such that

$$\sup_{y \in \mathcal{C}} \left(a^\top y\right) < \inf_{x \in \mathcal{D}} \left(a^\top x\right) = \min_{x \in \mathcal{D}} \left(a^\top x\right).$$

1. Set $a = x - P_{\mathcal{C}}(x)$. Because $x \notin \mathcal{C}, a \neq 0$, therefore by Property 1 of projection,

$$\begin{aligned}
&(x - P_{\mathcal{C}}(x))^\top (y - x + x - P_{\mathcal{C}}(x)) \leq 0, \forall y \in \mathcal{C} \\
&\Leftrightarrow a^\top y - a^\top x + \|a\|_2^2 \leq 0 \\
&\Leftrightarrow a^\top x \geq a^\top y + \|a\|_2^2 && (\|a\|_2^2 > 0) \\
&\Rightarrow a^\top x > \sup_{y \in \mathcal{C}} \left(a^\top y\right)
\end{aligned}$$

Strict Separation II

2. Define

$$\mathcal{C}-\mathcal{D}=\{x-y \mid x \in \mathcal{C}, y \in \mathcal{D}\}$$

It is trivial to see that $\mathcal{C}-\mathcal{D}$ is convex and closed.

$$\mathcal{C} \cap \mathcal{D}=\emptyset \Rightarrow 0 \notin \mathcal{C}-\mathcal{D}$$

Then by (1),

$$\exists a \neq 0, \sup_{z \in \mathcal{C}-\mathcal{D}}\left(a^{\top} z\right)<a^{\top} 0=0$$

Therefore,

$$\sup_{x \in \mathcal{C}, y \in \mathcal{D}}\left\{a^{\top}(x-y)\right\}<0. \Leftrightarrow \sup_{x \in \mathcal{C}}\left(a^{\top} x\right)<\inf_{y \in \mathcal{D}}\left(a^{\top} y\right)$$

Because $\mathcal{D}$ is bounded, therefore $\inf_{y \in \mathcal{D}}\left(a^{\top} y\right)=\min_{y \in \mathcal{D}}\left(a^{\top} y\right)$.

Strict Separation III

Theorem 9 (Consequence of Separation)

Recall the definition of a halfspace

$$\mathcal{H}_{a,b} = \left\{ x \in \mathbb{R}^n \mid a^\top x \le b \right\}$$

Let $C \in \mathbb{R}^n$ be a convex, closed, and non-empty set. S_C is the intersection of all halfspaces that contain C, i.e.,

$$S_C = \bigcap_{(a,b) \in \mathbb{R}^n \times \mathbb{R}} \mathcal{H}_{a,b}, \quad \mathcal{H}_{a,b} \supseteq C$$

Then we have

(1) S_C is convex.

(2) For any closed convex C, $C = S_C$.

1. $\mathcal{H}_{a,b}$ is convex. $\Rightarrow S_c = \bigcap_{(a,b)} \mathcal{H}_{a,b}$ is convex by the preservation law.

Strict Separation IV

2. $\mathcal{H}_{a,b} \supseteq C \Rightarrow S_c \supseteq C$. We need to show $S_c \subseteq C$ (By contradiction).
   Let $y \in S_c$, but $y \notin C$, then, by strict separation,

$$\exists a \in \mathbb{R}^n, a \neq 0, \text{ s.t. } a^\top y > \sup_{x \in C} \left\{ a^\top x \right\}.$$

   Therefore $a^\top y = \sup_{x \in C} \left\{ a^\top x \right\} + \varepsilon, \varepsilon > 0$.
   Let $b = \sup_{x \in C} \left\{ a^\top x \right\}$. By the definition of supremum,

$$a^\top x - b = a^\top x - \sup_{x \in C} \left\{ a^\top x \right\} \leq 0, \forall x \in C$$

   which suggests $\mathcal{H}_{a,b}$ as defined contains C.
   Since $y \in S_c$,

$$\begin{aligned} y \in \mathcal{H}_{a,b} = & \left\{ x \in \mathbb{R}^n \mid a^\top x \leq b \right\} \\ \Rightarrow & a^\top y - \sup_{x \in C} a^\top x \leq 0 \\ \Rightarrow & a^\top y - \left(a^\top y - \varepsilon \right) \leq 0 \\ \Rightarrow & \varepsilon \leq 0 \text{ (contradiction).} \end{aligned}$$

Definition 31 (Closure)

The closure of a set is the smallest closed set containing.

Definition 32 (Boundary)

Boundary $\equiv$ closure $\setminus$ interior.

Let $\mathcal{B}_l = \{x \mid \|x\|_2 \leq l\}$, then for any set C,

$$\mathrm{cl}(C) = \bigcap \{C + \varepsilon \mathcal{B}_1 \mid \varepsilon > 0\}$$

$$\mathrm{int}(C) = \{x \mid \exists \varepsilon > 0, x + \varepsilon \mathcal{B}_l \subset C\}$$

$$\mathrm{bd}(C) = \mathrm{cl}(C) \setminus \mathrm{int}(C)$$

Supporting Hyperplane II

Definition 33 (Supporting Hyperplane)

Suppose $C \subseteq \mathbb{R}^n$, and x_0 is a point in its boundary $\text{bd}(C)$, i.e.,

$$x_0 \in \text{bd}(C) = \text{cl}(C) \setminus \text{int}(C)$$

If $a \neq 0$ satisfies $a^\top x \leq a^\top x_0$ for all $x \in C$, then the hyperplane $\{x \mid a^\top x = a^\top x_0\}$ is called a supporting hyperplane to C at the point x_0.

- The point x_0 and the set C are separated by the hyperplane $\{x \mid a^\top x = a^\top x_0\}$.
- The geometric interpretation is that the hyperplane $\{x \mid a^\top x = a^\top x_0\}$ is tangent to C at x_0, and the halfspace $\{x \mid a^\top x \leq a^\top x_0\}$ contains C. This is illustrated in Figure 12.

Supporting Hyperplane III

- If f is convex and differentiable and x^* is the minimizer of f over convex set C, i.e.,

$$\begin{aligned} & x^* = \arg\min_{x \in C} f(x) \\ \Leftrightarrow & \nabla f\left(x^*\right)^\top \left(x - x^*\right) \geq 0, \forall x \in C \\ \Leftrightarrow & \nabla f\left(x^*\right)^\top x \geq f\left(x^*\right)^\top x^* \end{aligned}$$

  which means $-\nabla f\left(x^*\right)$ is the normal of a supporting hyperplane at $x^* \in C$, if $\nabla f\left(x^*\right) \neq 0$.

Supporting Hyperplane IV

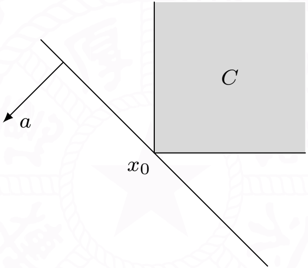


Fig. 12: The hyperplane $\left\{x \mid a^\top x = a^\top x_0\right\}$ supports C at x_0.

Theorem 10 (Supporting Hyperplane Theorem)

(1) For any nonempty convex set C, and any $x_0 \in \mathrm{bd}(C)$, there exists a supporting hyperplane to C at x_0.

(2) If a set is closed, has nonempty interior, and has a supporting hyperplane at every point in its boundary, then it is convex.

We provide a proof in Appendix 2.

Contents

Appendix 1: Separating Hyperplane Theorem I

Appendix 2: Supporting Hyperplane Theorem I