Operations Research II

Lecture 2: Elementary Convex Analysis

Shuming Wang

School of Economics & Management
University of Chinese Academy of Sciences
Email: wangshuming@ucas.ac.cn

Outline

Outline

1. Convex Sets and Cones
2. Polyhedra and Linear Programs
3. The Modeling Power of LP
4. Convex Functions

Convex Sets and Cones

Definition (Convex Set)

A set C (multi-dimensional) is convex if the line segment between any two points in C lies in C. That is, for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in C$ and any $\lambda \in [0, 1]$, we have

$$\lambda \boldsymbol{x}_1 + (1 - \lambda)\boldsymbol{x}_2 \in C.$$

- A, B convex, then $A \cap B$ is also convex
- A, B convex, then $A + B$ is also convex
- $A \cup B$ is not!

A, B convex, then $A \cap B$ is also convex.

Proof.

$\forall\ \boldsymbol{x}_1, \boldsymbol{x}_2 \in A \cap B$, we have

$$\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in A, \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in B.$$

That is

$$\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in A \cap B.$$

□

Convex Sets and Cones

A, B convex, then $A + B$ is also convex.

Proof.

Recall that
$$A + B := \{\mathbf{z} = \boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in A, \boldsymbol{y} \in B\}.$$

$\forall\ \mathbf{z}_1, \mathbf{z}_2 \in A + B$, note that

$$\mathbf{z}_i = \boldsymbol{x}_i + \boldsymbol{y}_i, i = 1, 2,$$

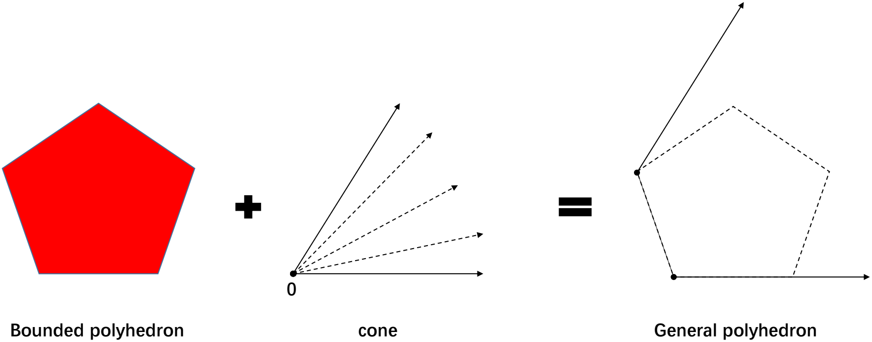
where $\boldsymbol{x}_i \in A, \boldsymbol{y}_i \in B$. We have

$$\begin{aligned} \lambda\mathbf{z}_1 + (1-\lambda)\mathbf{z}_2 &= \lambda(\boldsymbol{x}_1 + \boldsymbol{y}_1) + (1-\lambda)(\boldsymbol{x}_2 + \boldsymbol{y}_2) \\ &= \underbrace{\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2}_{\in A} + \underbrace{\lambda\boldsymbol{y}_1 + (1-\lambda)\boldsymbol{y}_2}_{\in B} \in A + B. \end{aligned}$$

□

Convex Sets and Cones

$$\mathbf{A} + \mathbf{B} = \mathbf{C}$$



Example (Subspace)

Subspace: a nonempty subset $\boldsymbol{S}$ of $\mathbb{R}^n$ is called a *subspace* of $\mathbb{R}^n$ if $a\boldsymbol{x}_1 + b\boldsymbol{x}_2 \in \boldsymbol{S}$ for every $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{S}$ and every $a, b \in \mathbb{R}$.

- $\boldsymbol{S}_1 = \{(x, 0)' \mid x \in \mathbb{R}\} \subseteq \mathbb{R}^2$: subspace, convex set.
- $\boldsymbol{S}_2 = \boldsymbol{S}_1 + (0, 1)' = \{(x, 1) \mid x \in \mathbb{R}\} \subseteq \mathbb{R}^2$: convex set, not subspace.

Example (Subspace)

Subspace: a nonempty subset $\boldsymbol{S}$ of $\mathbb{R}^n$ is called a *subspace* of $\mathbb{R}^n$ if $a\boldsymbol{x}_1 + b\boldsymbol{x}_2 \in \boldsymbol{S}$ for every $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{S}$ and every $a, b \in \mathbb{R}$.

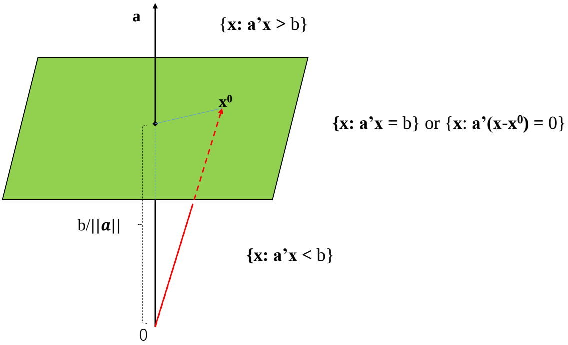
- $\boldsymbol{S}_1 = \{(x, 0)' \mid x \in \mathbb{R}\} \subseteq \mathbb{R}^2$: subspace, convex set.
- $\boldsymbol{S}_2 = \boldsymbol{S}_1 + (0, 1)' = \{(x, 1) \mid x \in \mathbb{R}\} \subseteq \mathbb{R}^2$: convex set, not subspace.
- Given $\mathbf{A} \in \mathbb{R}^{m \times n}$,

$$\mathcal{N}(\mathbf{A}) := \{\boldsymbol{x} \in \mathbb{R}^n \mid \mathbf{A}\boldsymbol{x} = \mathbf{0}\},\ \mathcal{R}(\mathbf{A}') := \{\mathbf{A}'\mathbf{z} \in \mathbb{R}^n \mid \mathbf{z} \in \mathbb{R}^m\},$$

both are subspaces of $\mathbb{R}^n$.

Example (Hyperplane, Halfspace)

Hyperplane: a set $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}'\boldsymbol{x} = b\}$ is called a *hyperplane*.
Halfspace: a set $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}'\boldsymbol{x} \geq b\}$ is called a *halfspace*, where $\boldsymbol{a} \in \mathbb{R}^n$ is a nonzero vector and b is a scalar ($b \in \mathbb{R}$).



Example (Support Vector Machine)

- Distance (signed) of a point $\boldsymbol{x}$ to a hyperplane $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}'\boldsymbol{x} = b\}$:

$$(\boldsymbol{x} - \boldsymbol{x}_0)' \left(\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|} \right) = \frac{\boldsymbol{a}'\boldsymbol{x} - b}{\|\boldsymbol{a}\|}$$

  where $\boldsymbol{x}_0 : \boldsymbol{a}'\boldsymbol{x}_0 = b$.

- Given training data points $\{(\hat{\boldsymbol{x}}_i, \hat{y}_i), i \in [N]\}$ assumed separable, where $\hat{\boldsymbol{x}}_i \in \mathbb{R}^n, \hat{y}_i \in \{-1, +1\}$, we can then always have

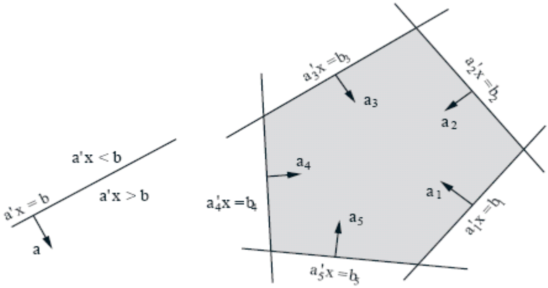
$$\hat{y}_i \left(\frac{\boldsymbol{a}'\hat{\boldsymbol{x}}_i - b}{\|\boldsymbol{a}\|} \right) > 0, \forall i \in [N].$$

- SVM model:

$$\operatorname*{Max}_{\boldsymbol{a}, b} \left\{ M : \hat{y}_i \left(\frac{\boldsymbol{a}'\hat{\boldsymbol{x}}_i - b}{\|\boldsymbol{a}\|} \right) \geq M, \forall i \in [N] \right\}.$$

Example (Polyhedron)

Polyhedron: a set $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}\}$ (or the intersection of many halfspaces) is called a *polyhedron*, where A is an $m \times n$ matrix and $\boldsymbol{b}$ is a vector in $\mathbb{R}^m$.

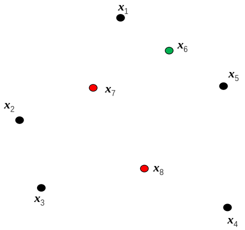


(a) (b)

Definition (Convex Hull)

The convex hull of n points (vectors) $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$, denoted by $\mathrm{conv}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$, is the set of all the convex combinations of these points:

$$\mathrm{conv}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} := \left\{ \sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i \mid \lambda_i \geq 0, i = 1, 2, \cdots, n, \sum_{i=1}^{n} \lambda_i = 1 \right\}.$$



Convex Sets and Cones

Definition (Convex Hull)

The convex hull of n points (vectors) $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$, denoted by $\text{conv}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$, is the set of all the convex combinations of these points:

$$\text{conv}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} := \left\{ \sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i \mid \lambda_i \geq 0, i = 1, 2, \cdots, n, \sum_{i=1}^{n} \lambda_i = 1 \right\}.$$

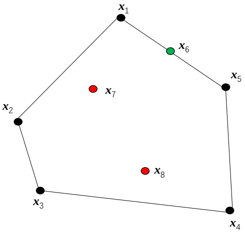
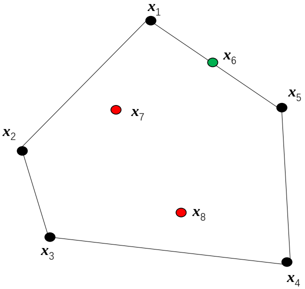


Figure 1: Convex hull of $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_8\}$

- $\text{conv}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ is the smallest convex set that contains set $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$. (H.W.)

Theorem (Representability of Bounded Polyhedron)

Any bounded polyhedron $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}\}$ is the convex hull of a finite set of vectors.



- Can we do it better? What
- Can we do it better? How Many?

Convex Sets and Cones

Definition (Cone & Convex Cone)

A set C is called a cone, if for every $\boldsymbol{x} \in C$ and $\lambda \geq 0$, we have $\lambda \boldsymbol{x} \in C$. Furthermore, a set C is a convex cone if it is a cone and is convex.

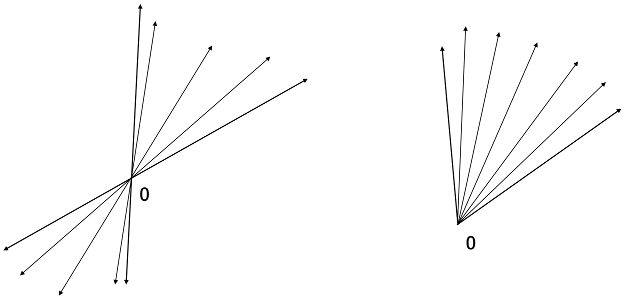


Figure 2: nonconvex cone and convex cone

Definition (Conic hull)

The conic hull of points $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ is the set of all conic combinations of these points:

$$\text{cone}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} := \left\{ \sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i \mid \lambda_i \geq 0, i = 1, 2, \cdots, n \right\}.$$

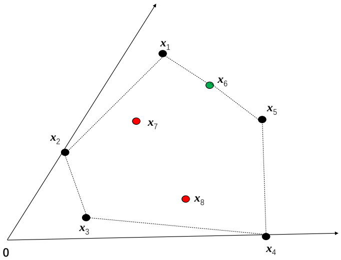


Figure 3: Conic hull $\text{cone}\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_8\}$ and convex hull $\text{conv}\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_8\}$

Convex Sets and Cones

Proposition

Conic hull $\text{cone}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ *is the smallest convex cone that contains* $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$.

Proof.

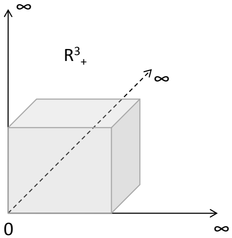
1. Convex cone $\boldsymbol{B}$: $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{B} \Rightarrow \lambda_1\boldsymbol{x}_1 + \lambda_2\boldsymbol{x}_2 \in \boldsymbol{B}$, for $\lambda_1, \lambda_2 \geq 0$.
2. For any convex cone $\boldsymbol{B}$ that contains $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$, $\text{cone}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\} \subseteq \boldsymbol{B}$.

□

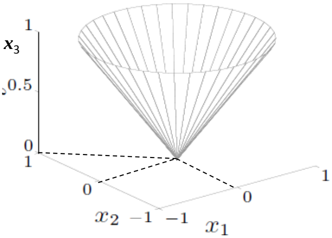
Example

1. The nonnegative orthant $\mathbb{R}^n_+$
2. The Lorantz cone (Second order cone, or ice-cream cone):

$$\mathbf{L}^m := \left\{ \boldsymbol{x} = (x_1, x_2, \cdots, x_m)' \in \mathbb{R}^m \mid \|(x_1, x_2, \cdots, x_{m-1})'\|_2 \le x_m \right\}$$



(a) $\mathbb{R}^3_+$



(b) Second order cone (3D)

Example

The semidefinite cone $\mathbf{S}_+^m$: The cone lives in the space of $m \times m$ symmetric matrices $\mathbf{S}^m$ and consists of all $m \times m$ positive semidefinite matrices $\boldsymbol{A}$:

1. $\boldsymbol{A} = \boldsymbol{A}'$;
2. $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0, \ \forall \boldsymbol{x} \in \mathbb{R}^m$.

Example

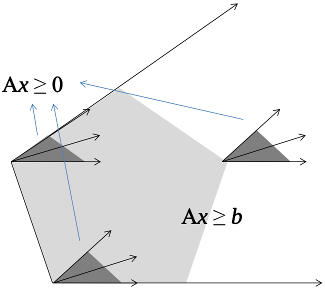
The polyhedron $\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \mathbf{0}\}$ is also a (convex) cone, called **polyhedral cone**.

Example (polyhedral cone as the recession cone of a polyhedron)

Given a polyhedron $\boldsymbol{P} = \{\boldsymbol{y} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{y} \geq \boldsymbol{b}\}$, then

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}(\boldsymbol{y} + \lambda \boldsymbol{x}) \geq \boldsymbol{b}, \forall \lambda \geq 0\} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{0}\}$$

is call the **recession cone** (of $\boldsymbol{P}$).



Theorem (polyhedral cone as conic hull of finite vectors)

Any polyhedral cone $\boldsymbol{C} := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \mathbf{0}\}$ *can be generated by a finite number of vectors: there exists a set of finite vectors* $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_m\} \subseteq \mathbb{R}^n$ *such that* $\boldsymbol{C} = \operatorname{cone}\{\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_m\}$ *i.e.*

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \mathbf{0}\} = \left\{\sum_{i=1}^{m} \lambda_i \boldsymbol{b}_i \mid \lambda_i \geq 0\right\}$$

Recall that

- Bounded polyhedron vs. convex hull
- Polyhedral cone vs. conic hull

Theorem (Separating Hyperplane Theorem)

Let $\boldsymbol{S}$ be a nonempty closed convex subset of $\mathbb{R}^n$, and let $\boldsymbol{x}^* \in \mathbb{R}^n$ be a vector that does not belong to $\boldsymbol{S}$. Then there exists some vector $\mathbf{d} \in \mathbb{R}^n$ such that $\mathbf{d}'\boldsymbol{x}^* < \mathbf{d}'\boldsymbol{x}$ for any $\boldsymbol{x} \in \boldsymbol{S}$.

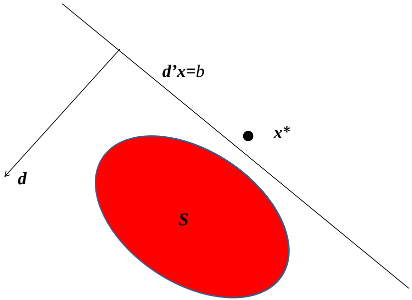


Figure 4: Hyperplane $\{\boldsymbol{x} \mid \boldsymbol{d}'\boldsymbol{x} = b\}$ separates set $\boldsymbol{S}$ and point $\boldsymbol{x}^*$

Theorem (Farkas' Lemma)

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{b} \in \mathbb{R}^m$. *Then, exactly one of the following two alternatives holds:*

1. *There exists some* $\boldsymbol{x} \geq 0$ *such that* $\boldsymbol{Ax} = \boldsymbol{b}$.
2. *There exists some vector* $\boldsymbol{p}$ *such that* $\boldsymbol{p}'\boldsymbol{A} \geq \boldsymbol{0}'$ *and* $\boldsymbol{p}'\boldsymbol{b} < 0$.

- Certificate of infeasibility for a given (standard) system of linear inequalities.
- The set $\{\boldsymbol{Ax} \mid \boldsymbol{x} \geq 0\}$ is closed.

Theorem (Separating Hyperplane Theorem)

Let $\boldsymbol{C}$ and $\boldsymbol{D}$ be two nonempty convex subset, with $\boldsymbol{C} \cap \boldsymbol{D} = \emptyset$. Then there exist $\boldsymbol{a} \neq \mathbf{0}$ and b, such that

$$\boldsymbol{a}'\boldsymbol{x} \leq b, \forall \boldsymbol{x} \in \boldsymbol{C}, \ \boldsymbol{a}'\boldsymbol{x} \geq b, \forall \boldsymbol{x} \in \boldsymbol{D}.$$

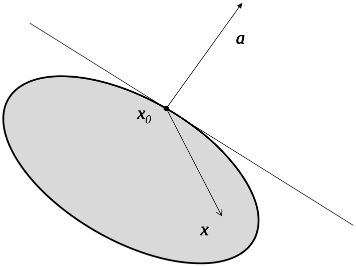
Convex Sets and Cones

Theorem (Supporting Hyperplane Theorem)

For any nonempty convex set $\boldsymbol{C}$, and $\boldsymbol{x}_0$ on the boundary of $\boldsymbol{C}$, there exists a **supporting hyperplane** *at $\boldsymbol{x}_0$, i.e.*
$\boldsymbol{S} := \{\boldsymbol{x} \mid \boldsymbol{a}'(\boldsymbol{x} - \boldsymbol{x}_0) = 0\}$*: $\boldsymbol{S}$ is tangent to $\boldsymbol{C}$ at $\boldsymbol{x}_0$, and*

$$\boldsymbol{a}'(\boldsymbol{x} - \boldsymbol{x}_0) \leq 0, \forall \boldsymbol{x} \in \boldsymbol{C}.$$

- Can be proved by Separating Hyperplane Theorem.



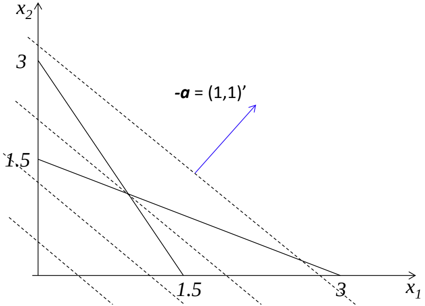
Outline

Consider the following LP problem:

$$
\begin{aligned}
\text{Min} \quad & -x_1 - x_2 \\
\text{s.t.} \quad & x_1 + 2x_2 \leq 3 \\
& 2x_1 + x_2 \leq 3 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

Consider the following LP problem:

$$\begin{aligned} \text{Min} \quad & -x_1 - x_2 \\ \text{s.t.} \quad & x_1 + 2x_2 \le 3 \\ & 2x_1 + x_2 \le 3 \\ & x_1, x_2 \ge 0 \end{aligned}$$



LP and Polyhedra

The linear programming problem in general has the form as follows

Linear Programming

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{a}_i'\boldsymbol{x} = b_i, i \in M_1 \\
& \boldsymbol{a}_i'\boldsymbol{x} \leq b_i, i \in M_2 \\
& \boldsymbol{a}_i'\boldsymbol{x} \geq b_i, i \in M_3 \\
& x_j \geq 0, j \in N_1 \\
& x_j \leq 0, j \in N_2 \\
& x_j \in \mathbb{R}, j \in N_3
\end{aligned}
$$

The linear programming problem in general has the form as follows

Linear Programming

$$
\begin{aligned}
\underset{\boldsymbol{x}}{\text{Min}} \quad & \boldsymbol{c}'\boldsymbol{x} \Leftrightarrow \boldsymbol{c}'\boldsymbol{x} + \boldsymbol{0}'\boldsymbol{s} \\
\text{s.t.} \quad & \boldsymbol{a}_i'\boldsymbol{x} = b_i, i \in M_1 \\
& \boldsymbol{a}_i'\boldsymbol{x} \leq b_i, i \in M_2 \Leftrightarrow \boldsymbol{a}_i'\boldsymbol{x} + s_i = b_i, s_i \geq 0, i \in M_2 \\
& \boldsymbol{a}_i'\boldsymbol{x} \geq b_i, i \in M_3 \Leftrightarrow \boldsymbol{a}_i'\boldsymbol{x} - s_i = b_i, s_i \geq 0, i \in M_2 \\
& x_j \geq 0, j \in N_1 \\
& x_j \leq 0, j \in N_2 \Leftrightarrow x_j' \geq 0 (\text{replacing } x_j \text{ with } -x_j'), j \in N_2 \\
& x_j \in \mathbb{R}, j \in N_3 \Leftrightarrow x_j^+, x_j^- \geq 0 (\text{replacing } x_j \text{ with } x_j^+ - x_j^-), j \in N_3
\end{aligned}
$$

LP and Polyhedra

Example

$$\begin{aligned} \text{Min} \quad & 2x_1 + 4x_2 \\ \text{s.t.} \quad & x_1 + x_2 \geq 3 \\ & 3x_1 + 2x_2 = 14 \\ & x_1 \geq 0 \\ & x_2 \in \mathbb{R} \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} \text{Min} \quad & 2x_1 + 4x_2^+ - 4x_2^- \\ \text{s.t.} \quad & x_1 + x_2^+ - x_2^- - x_3 = 3 \\ & 3x_1 + 2x_2^+ - 2x_2^- = 14 \\ & x_1, x_2^+, x_2^-, x_3 \geq 0 \end{aligned}$$

LP and Polyhedra

Any LP problem can be transformed into the Standard Form:

Linear Programming Standard Form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq 0 \end{aligned}$$

Characteristics:

1. Minimization objective
2. Equality constraints
3. nonnegative variables

The standard form is more computationally more convenient in developing the algorithms for solving the LPs, e.g. simplex and interior point methods.

The linear programming problem in general has the form as follows

Linear Programming

$$
\begin{aligned}
\underset{\boldsymbol{x}}{\text{Min}} \quad & \boldsymbol{c}'\boldsymbol{x} \\
\text{s.t.} \quad & \boldsymbol{a}_i'\boldsymbol{x} = b_i, i \in M_1 \Leftrightarrow \boldsymbol{a}_i'\boldsymbol{x} \geq b_i, -\boldsymbol{a}_i'\boldsymbol{x} \geq -b_i, i \in M_1 \\
& \boldsymbol{a}_i'\boldsymbol{x} \leq b_i, i \in M_2 \Leftrightarrow -\boldsymbol{a}_i'\boldsymbol{x} \geq -b_i, i \in M_2 \\
& \boldsymbol{a}_i'\boldsymbol{x} \geq b_i, i \in M_3 \\
& x_j \geq 0, j \in N_1 \\
& x_j \leq 0, j \in N_2 \Leftrightarrow -x_j \geq 0, j \in N_2 \\
& x_j \in \mathbb{R}, j \in N_3 \Leftrightarrow x_j^+, x_j^- \geq 0 \text{ (replacing } x_j \text{ with } x_j^+ - x_j^-\text{)}, j \in N_3
\end{aligned}
$$

LP and Polyhedra

Any LP problem can be transformed into the following general form:

Linear Programming General Form

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{Min}} \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \end{aligned}$$

The general formulation of the polyhedra is convenient to develop and illustrate the geometric insights of the LP.

Constraints as a polyhedron

- $\boldsymbol{A} \in \mathbb{R}^{m \times n}$
- $$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} =: [\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_n]$$
- $$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} =: \begin{bmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \\ \vdots \\ \boldsymbol{a}_m' \end{bmatrix}$$

LP and Polyhedra

Constraints as a polyhedron

- $$\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \Leftrightarrow \boldsymbol{A}_1 x_1 + \boldsymbol{A}_2 x_2 + \cdots + \boldsymbol{A}_n x_n \geq \boldsymbol{b}$$

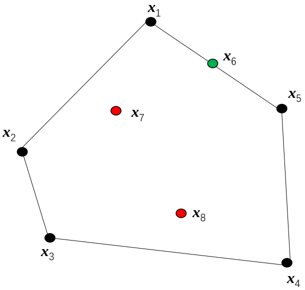
- $$\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \Leftrightarrow \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{x} \\ \boldsymbol{a}_2'\boldsymbol{x} \\ \vdots \\ \boldsymbol{a}_m'\boldsymbol{x} \end{bmatrix} \geq \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \boldsymbol{b}$$

Definition (Extreme Points)

We say $\boldsymbol{x} \in \boldsymbol{P} := \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}\}$ is an extreme point of $\boldsymbol{P}$ if

$$\nexists\ \boldsymbol{y}, \mathbf{z} \in \boldsymbol{P}, \boldsymbol{y} \neq \mathbf{z}, \lambda \in (0,1):\ \ \boldsymbol{x} = \lambda \boldsymbol{y} + (1-\lambda)\mathbf{z}.$$

- The number of extreme points of $\boldsymbol{P}$ is finite.

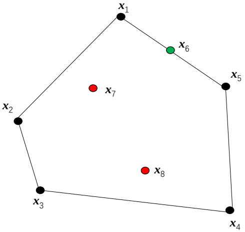


LP and Polyhedra

Theorem (Representation of Bounded Polyhedron)

Any nonempty and bounded polyhedron $\boldsymbol{P} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}\}$ *is the convex hull of its extreme points* $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_M\} \subseteq \mathbb{R}^n$*, i.e.*

$$\boldsymbol{P} = \mathrm{conv}\{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_M\} = \left\{ \sum_{i=1}^{M} \lambda_i \boldsymbol{e}_i \mid \sum_{i=1}^{M} \lambda_i = 1, \lambda_i \geq 0 \right\}.$$



- Can we do it better? What ✓
- Can we do it better? How Many?

Now we consider the following LP:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{Min}} & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{x} \in \boldsymbol{P} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \mathbf{b}\}. \end{array}$$

Theorem (Optimality of Extreme Points)

Suppose $\boldsymbol{P}$ has at least one extreme point. Then the optimal cost is either $-\infty$, or there exists an **extreme point** *that is optimal solution.*

Now we consider the following LP:

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{Min}} \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{x} \in \boldsymbol{P} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \mathbf{b}\}. \end{aligned}$$

Theorem (Optimality of Extreme Points)

Suppose $\boldsymbol{P}$ has at least one extreme point. Then the optimal cost is either $-\infty$, or there exists an **extreme point** *that is optimal solution.*

- Do we really need the condition "Suppose $\boldsymbol{P}$ has at least one extreme point"?

Existence of Extreme Points

Theorem (2.6. ★★★)

Suppose that the polyhedron $P = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}_i'\boldsymbol{x} \geq b_i, i = 1, \ldots, m\}$ *is nonempty. Then, the following are equivalent:*

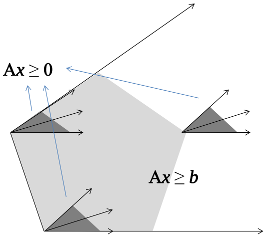
(a) *The polyhedron* P *has at least one extreme point.*

(b) *The polyhedron* P *does not contain a line.*

(c) *There exist* n *vectors out of the family* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$*, which are linearly independent.*

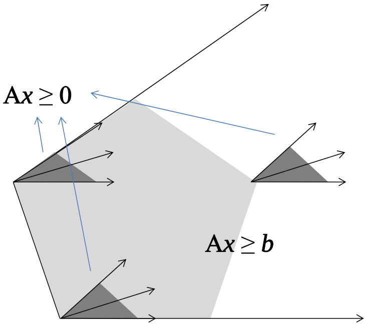
Definition (Extreme rays)

- A nonzero vector $\boldsymbol{d}$ of a polyhedral cone $\boldsymbol{C} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \mathbf{0}\}$ is called an **extreme ray** if $\boldsymbol{d}$ cannot be represented as a positive linear combination of two distant directions, i.e. if there exist $\boldsymbol{f}, \mathbf{g} \in \boldsymbol{C}$ such that $\boldsymbol{d} = \lambda_1 \boldsymbol{f} + \lambda_2 \mathbf{g}, \lambda_1, \lambda_2 > 0$, then $\boldsymbol{f} = \alpha \mathbf{g}, \alpha > 0$.
- The extreme ray of a polyhedron $\boldsymbol{P} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}\}$ is defined as the extreme ray of its recession cone $\boldsymbol{C} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \mathbf{0}\}$.



LP and Polyhedra

- Intuitively, the extreme rays are the **directions** associated with "edges" of the polyhedron that extend to infinity.
- **The number of extreme rays is finite.** (Theorem)



LP and Polyhedra

Theorem (Representation of General Polyhedron)

Let $\boldsymbol{P} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}\}$ *be a nonempty polyhedron with at least one extreme point. Let* $\boldsymbol{x}^1, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^k$ *be all the extreme points of* $\boldsymbol{P}$*, and* $\boldsymbol{w}^1, \boldsymbol{w}^2, \cdots, \boldsymbol{w}^r$ *be all the extreme rays of* $\boldsymbol{P}$*. Let*

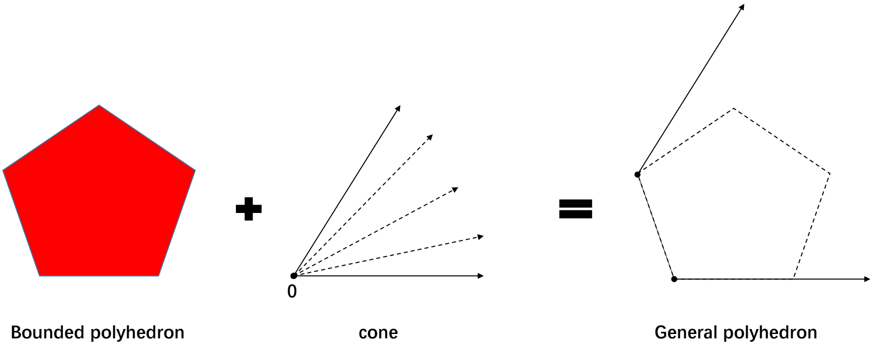
$$\boldsymbol{Q} = \boldsymbol{B} + \boldsymbol{C} = \left\{ \sum_{i=1}^{k} \lambda_i \boldsymbol{x}^i + \sum_{j=1}^{r} \theta_j \boldsymbol{w}^j \mid \lambda_i \geq 0, \sum_{i=1}^{k} \lambda = 1, \theta_j \geq 0 \right\}.$$

Then $\boldsymbol{Q} = \boldsymbol{P}$*.*

LP and Polyhedra

Theorem (Representation of General Polyhedron)

A polyhedron = A polytope (bounded polyhedron) + A generated cone



Now we consider the following LP:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{Min}} & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{x} \in \boldsymbol{P} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geq \mathbf{b}\}. \end{array}$$

Theorem (Optimality & Extreme Points)

For a general LP, the optimal cost is either $-\infty$, or there exists an **extreme point** *that is optimal solution.*

Theorem (Optimality & Extreme Ray)

For a general LP overall polyhedron $\boldsymbol{P}$, the optimal cost is $-\infty$ if and only if some **extreme ray** $\boldsymbol{d} \in \boldsymbol{P}$ *satisfies* $\mathbf{c}'\boldsymbol{d} < 0$.

Polyhedra in Standard Form and BFS

In view of Theorem 2.4, all basic solutions to a standard form polyhedron can be constructed according to the following procedure.

Procedure to Construct BSs of Standard Polyhedra

1. Choose m linearly independent columns $\boldsymbol{A}_{B(1)}, \ldots, \boldsymbol{A}_{B(m)}$.
2. Let $x_i = 0$ for all $i \neq B(1), \ldots, B(m)$.
3. Solve the system of m equations $\boldsymbol{Ax} = \boldsymbol{b}$ for the unknowns $x_{B(1)}, \ldots, x_{B(m)}$

Q: is the solution obtained a BFS?

Polyhedra in Standard Form and BFS

In view of Theorem 2.4, all basic solutions to a standard form polyhedron can be constructed according to the following procedure.

Procedure to Construct BSs of Standard Polyhedra

1. Choose m linearly independent columns $\boldsymbol{A}_{B(1)}, \ldots, \boldsymbol{A}_{B(m)}$.
2. Let $x_i = 0$ for all $i \neq B(1), \ldots, B(m)$.
3. Solve the system of m equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ for the unknowns $x_{B(1)}, \ldots, x_{B(m)}$

Q: is the solution obtained a BFS?

- If a basic solution constructed according to this procedure is nonnegative, then it is feasible, and it is a basic feasible solution.
- Conversely, since every BFS is a BS, it can be obtained from this procedure.

Theorem (Optimality Condition, 3.1)

Consider a basic feasible solution (BFS) $\boldsymbol{x}$ associated with a basis matrix $\boldsymbol{B}$, and let $\overline{\mathbf{c}}$ be the corresponding vector of reduced costs.

(a) *If $\overline{\mathbf{c}} := \boldsymbol{c}' - \boldsymbol{c}'_B \boldsymbol{B}^{-1} \boldsymbol{A} \geq \mathbf{0}'$, then $\boldsymbol{x}$ is optimal.*

(b) *If $\boldsymbol{x}$ is optimal and nondegenerate, then $\overline{\mathbf{c}} \geq \mathbf{0}'$.*

Proof.

(a) We consider any feasible solution $\boldsymbol{y}$ and the cost change rate $\boldsymbol{c}'(\boldsymbol{y} - \boldsymbol{x}) = \boldsymbol{c}'\boldsymbol{d}$. Then we claim $\boldsymbol{c}'\boldsymbol{d} \geq 0$.

(b) If $\bar{c}_j < 0$, the x_j must be a nonbasic variable. Since $\boldsymbol{x}$ is nondegenerate, the jth basic direction is always feasible and achieve a reduced cost, and $\boldsymbol{x}$ would not be optimal.

□

- Q: consider two polyhedra P_1, P_2 with

$$P_1 = \{\boldsymbol{x} \in \mathbb{R}^n : A_1\boldsymbol{x} \le \boldsymbol{b}_1\}, \quad P_2 = \{\boldsymbol{x} \in \mathbb{R}^n : A_2\boldsymbol{x} \le \boldsymbol{b}_2\},$$

how do we know $P_1 \subseteq P_2$ is correct?

Theorem (Proved by Fakars Lemma)

Let $P := \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} \le \boldsymbol{b}\}$ be a nonempty polyhedron. An inequality $\boldsymbol{c}'\boldsymbol{x} \le \delta$ is valid for P if and only if there exists $\boldsymbol{u} \ge 0$ such that $\boldsymbol{u}'A = \boldsymbol{c}'$ and $\boldsymbol{u}'\boldsymbol{b} \le \delta$.

- H.W.

Outline

Example

$$
\begin{array}{ll}
\underset{\boldsymbol{x}}{\text{Min}} & \overset{N}{\underset{i=1}{\text{Max}}}\, \boldsymbol{c}_i'\boldsymbol{x} + b_i \\
\text{s.t.} & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}
\end{array}
$$

The Modeling Power of LP

Example

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \max_{\boldsymbol{c} \in \boldsymbol{P}} \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \end{aligned}$$

where $\boldsymbol{P} := \{\boldsymbol{d} \mid \mathbf{M}\boldsymbol{d} \geq \boldsymbol{h}\}$ is a bounded polyhedron.

Example

$$\begin{array}{cl} \underset{\boldsymbol{x}}{\text{Min}} & \sum_{i=1}^{N} c_i |x_i| \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \end{array}$$

where $c_i > 0$.

- $$|x_i| = \text{Max}\{x_i, -x_i\} = \text{Min}\{\lambda_i : \lambda_i \geq x_i; \lambda_i \geq -x_i\}.$$

The Modeling Power of LP

Example

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{Min}} & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & \sum_{i=1}^{N} |x_i| \leq b \end{array}$$

Is this one still an LP?

Example

$$\begin{array}{cl} \underset{\boldsymbol{x}}{\text{Min}} & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & \sum_{i=1}^{N} |x_i| \geq b \end{array}$$

Example (Data fitting-I)

$$\underset{\boldsymbol{x}}{\text{Min}} \underset{i=1}{\overset{N}{\text{Max}}} |b_i - \boldsymbol{a}_i'\boldsymbol{x}| = \underset{\boldsymbol{x}}{\text{Min}} \|\boldsymbol{b} - A\boldsymbol{x}\|_\infty$$

Example (Data fitting-II)

$$\underset{\boldsymbol{x}}{\text{Min}} \sum_{i=1}^{N} |b_i - \boldsymbol{a}_i'\boldsymbol{x}| = \underset{\boldsymbol{x}}{\text{Min}} \|\boldsymbol{b} - A\boldsymbol{x}\|_1$$

Outline

Convex Functions

Definition (Convex Functions)

Let $\boldsymbol{S} \subseteq \mathbb{R}^n$ A real-valued function $f(\boldsymbol{x}) : \boldsymbol{S} \mapsto \mathbb{R}$ is *convex* if for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{S}$ and $\lambda \in [0, 1]$ we have

$$f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \leq \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2).$$

- **Concave**: $f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \geq \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2)$.
- $f(\boldsymbol{x})$ is convex if and only if $-f(\boldsymbol{x})$ is concave.

Convex Functions

Some examples of one-dimensional convex functions

- $f(x) = ax + b$
- $f(x) = x^2 + bx + c$
- $f(x) = |x|$
- $f(x) = -\ln x$ for $x > 0$
- $f(x) = \exp(x)$

Convex Functions

Some examples of multiple dimensional convex functions

- $f(\boldsymbol{x}) = \boldsymbol{a}'\boldsymbol{x} + b$
- ℓ_1-norm: $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$
- ℓ_2-norm: $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}'\boldsymbol{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$
- ℓ_∞-norm: $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_\infty = \mathrm{Max}_{i=1}^{n} |x_i|$

Convex Functions

Some examples of multiple dimensional convex functions

- $f(\boldsymbol{x}) = \boldsymbol{a}'\boldsymbol{x} + b$
- ℓ_1-norm: $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$
- ℓ_2-norm: $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}'\boldsymbol{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$
- ℓ_∞-norm: $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_\infty = \mathrm{Max}_{i=1}^{n} |x_i|$

$$\|\boldsymbol{x}\|_1 \geq \|\boldsymbol{x}\|_2 \geq \|\boldsymbol{x}\|_\infty.$$

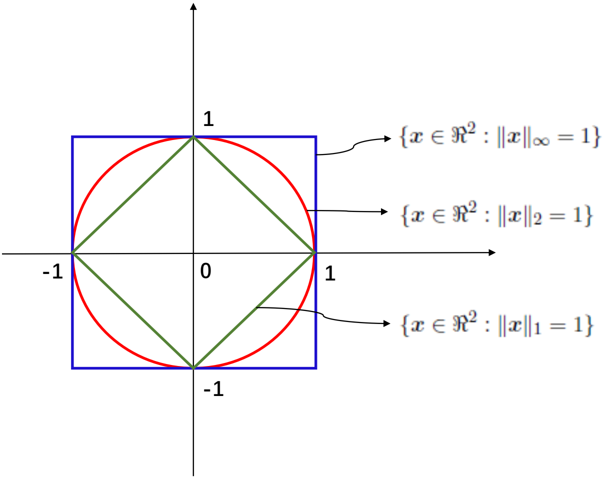


Figure 5: Norm balls $\{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_1 \leq 1\} \subseteq \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_2 \leq 1\} \subseteq \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_\infty \leq 1\}$

Definition (Dual Norm)

Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ which could be ℓ_1, ℓ_2 or ℓ_∞ norm. We denote by $\|\cdot\|^*$ the dual norm of $\|\cdot\|$, which is defined by

$$\|\boldsymbol{y}\|^* = \sup_{\|\boldsymbol{x}\|\leq 1} \boldsymbol{y}'\boldsymbol{x}.$$

Useful results

1. $\|\boldsymbol{y}\|_1^* = \sup_{\|\boldsymbol{x}\|_1\leq 1} \boldsymbol{y}'\boldsymbol{x} = \|\boldsymbol{y}\|_\infty$
2. $\|\boldsymbol{y}\|_2^* = \sup_{\|\boldsymbol{x}\|_2\leq 1} \boldsymbol{y}'\boldsymbol{x} = \|\boldsymbol{y}\|_2$
3. $\|\boldsymbol{y}\|_\infty^* = \sup_{\|\boldsymbol{x}\|_\infty\leq 1} \boldsymbol{y}'\boldsymbol{x} = \|\boldsymbol{y}\|_1$(H.W.)

Definition (Gradient Vector)

The gradient vector of function $f(\boldsymbol{x}), \boldsymbol{x} \in \mathbb{R}^n$ at point $\hat{\boldsymbol{x}}$, denoted $\nabla f(\hat{\boldsymbol{x}})$, is the vector of the partial derivative:

$$\nabla f(\hat{\boldsymbol{x}}) := \left(\frac{\partial f}{\partial x_1}(\hat{\boldsymbol{x}}), \frac{\partial f}{\partial x_2}(\hat{\boldsymbol{x}}), \cdots, \frac{\partial f}{\partial x_n}(\hat{\boldsymbol{x}}) \right)'.$$

The Hessian Matrix, denoted $\mathbf{H}_f(\hat{\boldsymbol{x}})$, is the matrix of the second partial derivatives:

$$\mathbf{H}_f(\hat{\boldsymbol{x}})_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\hat{\boldsymbol{x}}).$$

Example

$f(\boldsymbol{x}) = \boldsymbol{c}'\boldsymbol{x}, \nabla f(\boldsymbol{x}) = \boldsymbol{c}.$

Example

$f(\boldsymbol{x}) = 8x_1^2 + x_2^2 - x_1x_2 + 8x_1, \mathbf{z} := (1,0)'.$

- $f(\mathbf{z}) = 16.$
- $\nabla f(\mathbf{z}) = (16z_1 - z_2 + 8, 2z_2 - z_1)' = (24, -1)'$
- $\mathbf{H}_f(\mathbf{z}) = \begin{pmatrix} 16 & -1 \\ -1 & 2 \end{pmatrix}$

Example (H.W.)

Let $f(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{x}'\mathbf{M}\boldsymbol{x}$ with $\mathbf{M}$ symmetric, then we have $\nabla f(\boldsymbol{x}) = \mathbf{M}\boldsymbol{x}, \mathbf{H}_f(\boldsymbol{x}) = \mathbf{M}.$

Ordinary Least Squares (OLS) Estimation

Given the observation $\boldsymbol{X}$, and any parameter estimates $\boldsymbol{\beta}$

- Estimated regression equation: $\hat{\boldsymbol{y}}_i(\boldsymbol{\beta}) = \boldsymbol{x}_i'\boldsymbol{\beta}$
- Matrix Form: $\hat{\boldsymbol{y}}(\boldsymbol{\beta}) = \boldsymbol{X}\boldsymbol{\beta}$

Example (OLS Estimator $\hat{\boldsymbol{\beta}}$)

The OLS Estimator $\hat{\boldsymbol{\beta}}$ of parameter $\boldsymbol{\beta}$ solves the following SOCP:

$$\underset{\boldsymbol{\beta}}{\text{Min}} \sum_{i=1}^{n} (y_i - \hat{y}_i(\boldsymbol{\beta}))^2 = \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})^2 = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2.$$

Normal equation: $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{y}$, which implies

$$\boldsymbol{\beta}^* = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$

where the symmetric matrix $\boldsymbol{X}'\boldsymbol{X}$ is assumed to be nonsingular.

Theorem (First Order Condition)

Let $f(\boldsymbol{x}) : \mathbb{R}^n \longmapsto \mathbb{R}$ *be differentiable with a convex domain* $\mathbf{dom}\ f$*. Then* $f(\boldsymbol{x})$ *is convex if and only if*

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + [\nabla f(\boldsymbol{x})]'(\boldsymbol{y} - \boldsymbol{x})$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{dom} f$*.*

Theorem (Second Order Condition)

Let $f(\boldsymbol{x}) : \mathbb{R}^n \longmapsto \mathbb{R}$ *be twice differentiable with a convex domain* **dom** f*. Then* $f(\boldsymbol{x})$ *is convex if and only if its Hessian matrix* $\mathbf{H}_f(\boldsymbol{x})$ *is positive semi-definite (PSD):*

$$\mathbf{z}'\mathbf{H}_f(\boldsymbol{x})\mathbf{z} \geq 0, \forall \mathbf{z} \in \mathbb{R}^n.$$

for any $\boldsymbol{x}$*.*

Theorem (Second Order Condition)

Let $f(\boldsymbol{x}) : \mathbb{R}^n \longmapsto \mathbb{R}$ *be twice differentiable with a convex domain* **dom** f*. Then* $f(\boldsymbol{x})$ *is convex if and only if its Hessian matrix* $\mathbf{H}_f(\boldsymbol{x})$ *is positive semi-definite (PSD):*

$$\mathbf{z}'\mathbf{H}_f(\boldsymbol{x})\mathbf{z} \geq 0, \forall \mathbf{z} \in \mathbb{R}^n.$$

for any $\boldsymbol{x}$*.*

Example (Quadratic functions)

Consider $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}'\mathbf{M}\boldsymbol{x} + \boldsymbol{b}'\boldsymbol{x} + c$ with $\mathbf{M}$ symmetric, then f is convex iff $\mathbf{M}$ is PSD.

Example (Variance function)

Variance function $f(\boldsymbol{x}) = \mathrm{Var}(\boldsymbol{\xi}'\boldsymbol{x})$ with $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots, \xi_n)'$ being the stock returns, is convex.

Example (Variance function)

Variance function $f(\boldsymbol{x}) = \mathrm{Var}(\boldsymbol{\xi}'\boldsymbol{x})$ with $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots, \xi_n)'$ being the stock returns, is convex.

$$\mathrm{Var}(\boldsymbol{\xi}'\boldsymbol{x}) = \mathbb{E}\Big[(\boldsymbol{\xi} - \boldsymbol{\mu})'\boldsymbol{x}\Big]^2 = \boldsymbol{x}'\underbrace{\mathbb{E}\Big[(\boldsymbol{\xi} - \boldsymbol{\mu})(\boldsymbol{\xi} - \boldsymbol{\mu})'\Big]}_{\boldsymbol{\Sigma}}\boldsymbol{x}$$

where $\boldsymbol{\Sigma}$ is a covariance matrix.

- If $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ are convex functions, and $a, b \geq 0$, then $af_1(\boldsymbol{x}) + bf_2(\boldsymbol{x})$ is also convex.
- If $f(\boldsymbol{x})$ is a convex function, and let $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}$, then $g(\boldsymbol{y}) := f(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b})$ is convex in $\boldsymbol{y}$.

Example

- $f(\boldsymbol{x}) = -\sum_{i=1}^{m} \ln(b_i - \boldsymbol{a}_i'\boldsymbol{x})$ is a convex function of $\boldsymbol{x}$ $(\boldsymbol{A}\boldsymbol{x} < \boldsymbol{b})$.

- If $f_i(\boldsymbol{x})$ is convex, then so is $g(\boldsymbol{x}) := \text{Max}_i\, f_i(\boldsymbol{x})$. (H.W.)
- If $f_i(\boldsymbol{x})$ is concave, then so is $g(\boldsymbol{x}) := \text{Min}_i\, f_i(\boldsymbol{x})$. (H.W.)

The power of piece-wise convex function

$$\operatorname*{Min}_{\boldsymbol{x}} g(\boldsymbol{x}) := \operatorname*{Max}_{i \in \mathcal{I}} \{\boldsymbol{a}_i \boldsymbol{x} - b_i\} \Longleftrightarrow \begin{array}{ll} \displaystyle\operatorname*{Min}_{\boldsymbol{x},\gamma} & \gamma \\ \text{s.t.} & \gamma \geq \boldsymbol{a}_i \boldsymbol{x} - b_i, i \in \mathcal{I} \end{array}$$

Definition (Epigraph)

Given $f(\boldsymbol{x}) : \boldsymbol{S} \mapsto \mathbb{R}, \boldsymbol{S} \subseteq \mathbb{R}^n$ is a real-valued function, the *epipgraph* of f is defined as

$$\text{epi}f := \{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R} \mid f(\boldsymbol{x}) \leq t\}.$$

Theorem

Function $f(\boldsymbol{x}) : \boldsymbol{S} \mapsto \mathbb{R}$ *is a convex function if and only if* $\mathrm{epi} f$ *is a convex set.*

Example

Show that function

$$f(\boldsymbol{x}) := \ln\left(\sum_{i=1}^{n} p_i \mathrm{exp}(a_i x_i + b_i)\right), \quad p_i > 0$$

is a convex function.

First-Order Condition (FOC): $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + [\nabla f(\boldsymbol{x})]' (\boldsymbol{y} - \boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom} f$.

Example (Interpreting the FOC with epigraph)

Since epi$f$ is a convex set, using supporting hyperplane theorem, given any boundary point $(\boldsymbol{x}, f(\boldsymbol{x}))$ in epi$f$, we have

$$\begin{bmatrix} \nabla f(\boldsymbol{x}) \\ -1 \end{bmatrix}' \left(\begin{bmatrix} \boldsymbol{y} \\ t \end{bmatrix} - \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{f}(\boldsymbol{x}) \end{bmatrix} \right) \leq 0, \forall (\boldsymbol{y}, t) \in \mathrm{epi} f,$$

where $(\nabla f(\boldsymbol{x}), -1)'$ defines a supporting hyperplane to epi$f$ at $(\boldsymbol{x}, f(\boldsymbol{x}))$. This is equivalent to

$$t \geq f(\boldsymbol{x}) + [\nabla f(\boldsymbol{x})]' (\boldsymbol{y} - \boldsymbol{x}), \forall \boldsymbol{x} \in \mathrm{dom} f, \forall (\boldsymbol{y}, t) \in \mathrm{epi} f.$$

- Convex Optimization Problem:

$$\underset{\boldsymbol{x}}{\text{Min}}\{f(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{X}\},$$

where $f(\boldsymbol{x})$ is a convex function and $\boldsymbol{X}$ is a convex set.

Theorem (Local Optimality = Global Optimality)

Let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be a convex function, and $\boldsymbol{X}$ be a nonempty convex subset, then any local optimal point must be a global optimal point.

- local optimal point $\mathbf{z}$: $\exists r > 0$ such that

$$f(\mathbf{z}) = \inf\{f(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{X}, \|\boldsymbol{x} - \mathbf{z}\|_2 \leq r\}$$

Theorem (Optimality Condition for Differentiable Convex Functions)

Let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be a differentiable convex function, and $\boldsymbol{X}$ be a nonempty convex subset, then

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x} \in \boldsymbol{X}} f(\boldsymbol{x}) \iff \nabla f(\boldsymbol{x}^*)'(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \forall \boldsymbol{x} \in \boldsymbol{X}.$$

- Geometric view: If $\nabla f(\boldsymbol{x}^*) = 0$, then ?
- Geometric view: If $\nabla f(\boldsymbol{x}^*) \neq 0$, then $-\nabla f(\boldsymbol{x}^*)$ defines a supporting hyperplane to the feasible set $\boldsymbol{X}$ at $\boldsymbol{x}^*$.
- Optimization view: ?

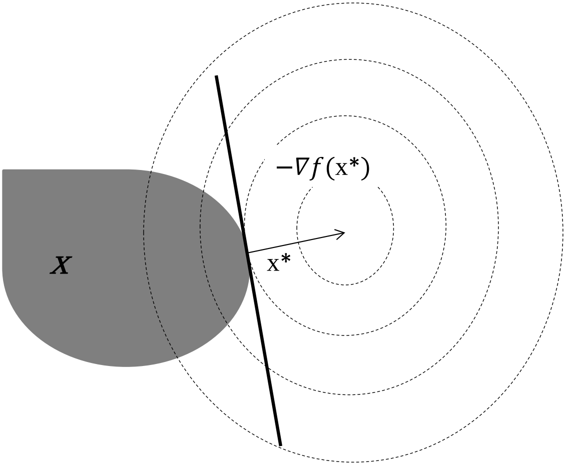


Figure 6: If $\nabla f(x) \neq 0$, then $-\nabla f(\boldsymbol{x}^*)$ defines a supporting hyperplane to the feasible set $\boldsymbol{X}$ at $\boldsymbol{x}^*$

Example (Optimality condition vs. Affine Constraints)

Consider the case where there are equality constraints but no inequality constraints, *i.e.*,

$$\begin{array}{ll} \text{minimize}_{\boldsymbol{x}} & f_0(\boldsymbol{x}) \\ \text{subject to} & A\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

where $A \in \mathbb{R}^{m \times n}$. The optimality condition reduces to

$$\left\{\nabla f_0(\boldsymbol{x}) + A^\top \boldsymbol{\nu} = 0, \boldsymbol{\nu} \in \mathbb{R}^m\right\} \neq \emptyset.$$

Optimality Condition

- The optimality condition for a feasible $\boldsymbol{x}$ now becomes

$$\nabla f_0(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}) \geq \mathbf{0}, \forall \boldsymbol{y} : A\boldsymbol{y} = \boldsymbol{b}.$$

- Since $\boldsymbol{x}$ is feasible, every feasible $\boldsymbol{y}$ has the form $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v}$ for some $\boldsymbol{v} \in \mathcal{N}(A)$. The optimality condition can therefore be expressed as

$$\nabla f_0(\boldsymbol{x})^\top \boldsymbol{v} \geq \mathbf{0}, \forall \boldsymbol{v} \in \mathcal{N}(A).$$

- <u>**An Important Fact**</u>: If a linear function is nonnegative on a subspace, then it must be zero on the subspace. So it follows that $\nabla f_0(\boldsymbol{x})^\top \boldsymbol{v} = 0$ for all $\boldsymbol{v} \in \mathcal{N}$, or

$$\nabla f_0(\boldsymbol{x}) \perp \mathcal{N}(A).$$

- Furthermore, using the fact that $\mathcal{N}^\perp = \mathcal{R}(A^\top)$, this optimality condition can be expressed as $\nabla f_0(\boldsymbol{x}) \in \mathcal{R}\left(A^\top\right)$, *i.e.*, there exists a $\boldsymbol{\nu} \in \mathbb{R}^m$ such that

$$\nabla f_0(\boldsymbol{x}) + A^\top \boldsymbol{\nu} = 0.$$

Example (Optimality condition vs. Affine Constraints)

Consider the case where there are equality constraints but no inequality constraints, *i.e.*,

$$\begin{array}{rl} \text{minimize} & f_0(\boldsymbol{x}) \\ \text{subject to} & A\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

where $A \in \mathbb{R}^{m \times n}$. The optimality condition reduces to

$$\left\{\nabla f_0(\boldsymbol{x}) + A^\top \boldsymbol{\nu} = 0, \boldsymbol{\nu} \in \mathbb{R}^m\right\} \neq \emptyset.$$

- What if $f_0(\boldsymbol{x}) = \boldsymbol{c}^\top \boldsymbol{x}$?

Theorem (Convexity recognition for general functions)

A function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex if and only if for any $\mathbf{z} \in \mathbb{R}^n$, there exists a vector $\boldsymbol{s}(\mathbf{z}) \in \mathbb{R}^n$ such that

$$f(\boldsymbol{x}) \geq f(\mathbf{z}) + \boldsymbol{s}(\mathbf{z})'(\boldsymbol{x} - \mathbf{z}),$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. The vector $\boldsymbol{s}(\mathbf{z})$ is called a **Subgradient** *of f at $\mathbf{z}$.*

- ♡ The set of all the subgradients of f at $\mathbf{z}$ is denoted by $\partial f(\mathbf{z})$ and is called the **Subdifferential** of f at $\mathbf{z}$.
- ♡ Especially, when f is differentiable, then the Subdifferential reduces to the one point set of gradient $\partial f(\mathbf{z}) = \{\nabla f(\mathbf{z})\}$.

An Optimality Condition for General Convex Functions

Let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be a proper convex function, and let $\boldsymbol{X}$ be a nonempty convex subset, and assume that $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(\boldsymbol{X}) \neq \emptyset$, then

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x} \in \boldsymbol{X}} f(\boldsymbol{x}) \Longleftrightarrow \exists \boldsymbol{s} \in \partial f(\boldsymbol{x}^*), \text{ s.t. } \boldsymbol{s}'(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \forall \boldsymbol{x} \in \boldsymbol{X}.$$

- When $\boldsymbol{X} = \mathbb{R}^n$
- When f is differentiable

Affine representation for a convex function

Any convex function can be represented as the point-wise supremum affine functions:

$$f(\boldsymbol{x}) \geq f(\mathbf{z}) + \boldsymbol{s}(\mathbf{z})'(\boldsymbol{x} - \mathbf{z}) \ \forall \mathbf{z} \in \mathbf{dom}(f) \Rightarrow$$

$$f(\boldsymbol{x}) = \operatorname*{Max}_{\mathbf{z} \in \mathbf{dom}(f)} \{f(\mathbf{z}) + \boldsymbol{s}(\mathbf{z})'(\boldsymbol{x} - \mathbf{z})\}$$

- Recall a convex set

$$\boldsymbol{S} = \bigcap_{x_0 \in \mathcal{B}(\mathbf{S})} \{\boldsymbol{x} : \boldsymbol{a}(\boldsymbol{x}_0)'\boldsymbol{x} \leq \boldsymbol{a}(\boldsymbol{x}_0)'\boldsymbol{x}_0\}$$

where $\boldsymbol{a}(\boldsymbol{x}_0)$ is the supporting vector with respect to $\boldsymbol{x}_0$.

♠ The epigraph

$$
\begin{aligned}
\text{epi} f &= \{(\boldsymbol{x}, t) : f(\boldsymbol{x}) \le t\} \\
&= \left\{ (\boldsymbol{x}, t) : \underset{\mathbf{z} \in \mathbf{dom}(f)}{\text{Max}} \{ f(\mathbf{z}) + \boldsymbol{s}(\mathbf{z})'(\boldsymbol{x} - \mathbf{z}) \le t \right\} \\
&= \bigcap_{\mathbf{z} \in \mathbf{dom}(f)} \{(\boldsymbol{x}, t) : f(\mathbf{z}) + \boldsymbol{s}(\mathbf{z})'(\boldsymbol{x} - \mathbf{z}) \le t\} \\
&= \bigcap_{\mathbf{z} \in \mathbf{dom}(f)} \left\{ (\boldsymbol{x}, t) : \begin{bmatrix} \boldsymbol{s}(\mathbf{z}) \\ -1 \end{bmatrix}' \left(\begin{bmatrix} \boldsymbol{x} \\ t \end{bmatrix} - \begin{bmatrix} \mathbf{z} \\ \boldsymbol{f}(\mathbf{z}) \end{bmatrix} \right) \le 0 \right\} \\
&= \bigcap_{(\mathbf{z}, r) \in \mathcal{B}(\text{epi} f)} \left\{ (\boldsymbol{x}, t) : \begin{bmatrix} \boldsymbol{s}(\mathbf{z}) \\ -1 \end{bmatrix}' \left(\begin{bmatrix} \boldsymbol{x} \\ t \end{bmatrix} - \begin{bmatrix} \mathbf{z} \\ r \end{bmatrix} \right) \le 0 \right\}
\end{aligned}
$$

Reference and Further Reading

1. S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2009.
2. D. Bertsimas, J.N. Tsitsiklis, Introduction to Linear Optimization, Athena Scientific, Nashua, 1997.