CSE203B Convex Optimization
Lecture 2 Convex Set

CK Cheng
Dept. of Computer Science and Engineering
University of California, San Diego

Chapter 2 Convex Set

1. Set Convexity and Specification
   i. Convexity
   ii. Set Specification: Qualification vs. Enumeration Oriented Description
2. Convex Set Terms and Definitions
3. Separating Hyperplanes
4. Dual Cones

Convex Optimization Problem:

$$\min_x f_0(x),\ x \in R^n$$

Subject to

$$f_i(x) \le b_i\ , i = 1, \cdots, m$$

1. $f_0(x)$ is a convex function
2. For $f_i(x) \le b_i, i = 1, \ldots, m$

$\{x | f_i(x) \le b_i, i = 1, \cdots, m\}$ is a convex set

Convex Optimization Problem:

A. Convex Function Definition:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y), \forall \alpha + \beta = 1, \alpha, \beta \geq 0$$

Convex Optimization Problem:

A. Convex Function Definition:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y), \forall \alpha + \beta = 1, \alpha, \beta \geq 0$$

B. Convex Set Definition: $\forall x, y \in C$

We have $\alpha x + \beta y \in C, \forall \alpha + \beta = 1, \alpha, \beta \geq 0$

1. Set Convexity and Specification: Convexity

A set S is convex if we have

$$\alpha x + \beta y \in S, \forall \alpha + \beta = 1, \alpha, \beta \geq 0, \forall x, y \in S$$

Examples:

1. Set Convexity and Specification: Convexity

A set S is convex if we have

$$\alpha x + \beta y \in S, \forall \alpha + \beta = 1, \alpha, \beta \geq 0, \forall x, y \in S$$

Remark:

1. Most used sets in the class
   1. Scalar set: $S \subset R$
   2. Vector set: $S \subset R^n$
   3. Matrix set: $S \subset R^{n \times m}$
2. Set S is convex if every two points in S has the connected straight segment in the set.
3. For convex sets S_1 and S_2: $S_1 \cap S_2$ is also convex

1. Set Convexity and Specification:
Set Specification via Qualification or Enumeration

<u>Qualification Oriented Expression</u> $S_Q = \{x | Ax \leq b, x \in R^n\}$

<u>Enumeration Oriented Expression</u> $S_E = \{Ax \mid x \in R^n_+\}$

Qualification Oriented Expression:

Constraints

Min $f_o(x)$

Subject to

$$Ax \leq b, x \in R^n$$

Enumeration Oriented Expression:

Obj function

Min $f_o(Ax), x \in R^n_+$

1. Qualification vs Enumeration Oriented Description

Qualification Oriented Expression

Example: $\{x | Ax \leq b\}$

Remark: Simultaneous linear constraints imply AND, Intersection of the constraints

$$
\begin{aligned}
x_1 \quad +2x_2 \quad +3x_3 &\leq 4 \\
2x_1 \quad -x_2 \qquad\qquad &\leq 3 \\
x_2 \quad +x_3 &\leq 5 \\
x_3 &\leq 10
\end{aligned}
$$

$$
A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \; x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \; b = \begin{bmatrix} 4 \\ 3 \\ 5 \\ 10 \end{bmatrix}
$$

1. Qualification vs. Enumeration Oriented Description

$S_1 = \{x | Ax \leq b, x \in R^n\}$ is a convex set

Proof: Given two vectors $u, v \in S_1$, $i.e.\ Au \leq b, Av \leq b$

For $w = \theta_1 u + \theta_2 v, \forall \theta_1 + \theta_2 = 1, \theta_1, \theta_2 \geq 0$

we have $Aw \leq b$.

$(Aw = \theta_1 Au + \theta_2 Av \leq \theta_1 b + \theta_2 b = b)$

The inequality implies $w \in S_1$

By definition, set S_1 is convex.

Remark:

1. Simultaneous linear constraints imply AND, Intersection of the constraints
2. Linear constraints constitute a convex set.

1. Qualification vs. Enumeration Oriented Description

Example:

$$S_2 = \{x | Ax \geq b, x \in R^n\}$$

$$S_3 = \{x | Ax = b, x \in R^n\}$$

1. Qualification Oriented Expression

Example: $S = \{x \in R^m \mid |p_x(t)| \leq 1 \text{ for } |t| \leq \frac{\pi}{3}\}$

$$where\ p_x(t) = x_1 \cos t + x_2 \cos 2t + \cdots + x_m \cos mt$$

1. Enumeration Oriented Expression

Expression Conversion

Example: $\{x | Ax \leq b, x \in R^n\}$ vs $\{U\theta |\ 1^T\theta = 1,\ \theta \in R_+^m\}$

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 3 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}$$

1. Qualification vs. Enumeration Oriented Description

Remark:

Qualification Oriented Expression: Constraints of the problem

Enumeration Oriented Enumeration: The objective function

The interchange may not be trivial.

$$\begin{aligned} &\min f_0(x) \\ &s.t.\, Ax \le b \\ &x \in R^n \end{aligned} \qquad\qquad \begin{aligned} &\min f_0(U\theta) \\ &s.t.\ I^T\theta \le 1 \\ &U \in R^{nm}, \theta \in R_+^m \end{aligned}$$

Every vector u_i in matrix U is a solution of n equations in constraint $Ax \le b$

$$\begin{matrix} p\ equations \\ n\ variables \end{matrix} \quad \Rightarrow \quad \begin{matrix} comb(p,n)\ possible \\ vertex\ points. \end{matrix}$$

1. Qualification vs. Enumeration Oriented Description

Mixed Description

$$S_4 = \left\{ \frac{Ax+b}{c^T x + d} \middle| (c^T x + d) > 0, x \in C_4 \right\} \text{ (Projective Function)}$$

$$S_5 = \left\{ \frac{z}{t} \middle| z \in R^n, t > 0, (z,t) \in C_5 \right\} \text{ (Perspective Function)}$$

S_4 is convex if C_4 is convex
S_5 is convex if C_5 is convex

1. Qualification vs. Enumeration Oriented Description

Statement: S_5 $is\ convex$ if C_5 $is\ convex$.

Proof: Given $\left(\frac{z_1}{t_1}\right) \in S_5$, $\left(\frac{z_2}{t_2}\right) \in S_5$, let us set

$$z_3 = \alpha z_1 + \beta z_2, t_3 = \alpha t_1 + \beta t_2, \forall \alpha + \beta = 1, \alpha, \beta \geq 0$$

We have $$\frac{z_3}{t_3} = \frac{\alpha z_1 + \beta z_2}{\alpha t_1 + \beta t_2} = \frac{\alpha t_1}{\alpha t_1 + \beta t_2}\frac{z_1}{t_1} + \frac{\beta t_2}{\alpha t_1 + \beta t_2}\frac{z_2}{t_2}$$

Let $\alpha' = \dfrac{\alpha t_1}{\alpha t_1 + \beta t_2}, \beta' = \dfrac{\beta t_2}{\alpha t_1 + \beta t_2}$

(Note that $\alpha' + \beta' = 1, \alpha', \beta' \geq 0$),

we have $\dfrac{z_3}{t_3} = \alpha' \dfrac{z_1}{t_1} + \beta' \dfrac{z_2}{t_2} \in S_5$

Therefore, by definition S_5 is convex.

2. Convex Set: Terms and Definitions

Definitions: I. Affine Set, II. Cone, and III. Convex Hull

Given $u_1, u_2, \cdots, u_k \in R^n$,

function $f(u, \theta) = \theta_1 u_1 + \theta_2 u_2 + \cdots + \theta_k u_k$

and two conditions

1. $\theta_1 + \theta_2 + \cdots + \theta_k = 1$
2. $\theta_i \geq 0 \ \forall i$

I. $\{f(u, \theta) | \text{ condition } 1\}$: Affine set

II. $\{f(u, \theta) | \text{ condition } 2\}$: Cone

III. $\{f(u, \theta) | \text{ conditions 1 and } 2\}$: Convex hull

$Ex1: \theta_1 u_1 + \theta_2 u_2 = u_1 + \theta_2 (u_2 - u_1)$

$Ex2: \theta_1 u_1 + \theta_2 u_2 + \theta_3 u_3$

2. Sets and Definitions: VI. Hyperplanes and Half Spaces

Hyperplane $\{x|\ a^T x = b\}, a \in R^n, b \in R$

or $\{x|\ a^T(x - x_0) = 0\}$, for any $x_0 \in R^n, a \in R^n, b \in R$

Half Space $\{x|\ a^T x \leq b\}\ \ a \in R^n, b \in R$

or $\{x|\ a^T(x - x_0) \leq 0\}$

Ex: $\{x|\ x_1 + x_2 = 1\}$ or $\{x|\ [1,1]\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}\right) = 0\}$

or $\{x|\ a^T(x - x_0) = 0\}$, $a^T = [1,1], b = 1, x_0 = [2, -1]$

For many applications, we standardize the expression:

normalize the expression: $\frac{a^T}{\|a\|_2} x = \frac{b}{\|a\|_2}$

2. Sets and Definitions: Hyperplanes

Ex : 3 variables

$\{x|a^T x = b\},\ a^T = (1,1,1),\ b = 6$

Ex : 4 variables

$\{x|a^T x = b\},\ a^T = (1,1,1,1),\ b = 6$

(1) degrees of freedom : $n - 1\ (R^n)$.

(2) all vectors $(x - y)$ are orthogonal to direction a, i.e. $a^T(x - y) = 0,\ \forall x, y$ in the hyperplane

Proof:

Exercise: Conceptually (visually) construct hyperplane.

2. Sets and Definitions: Hyperplanes

Hyperplane : as an Equal potential of cost function

$$\min f_0(x) = c^T x$$

$$e.g.\ [1, 2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\frac{\partial f_0(x)}{\partial x_1} = 1$$

$$\frac{\partial f_0(x)}{\partial x_2} = 2$$

Vector c is the sensitivity or cost of vector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

2. Sets and Definitions

Hyperplane : as a linearized constraint

$$a^T x \leq b, x \in R^n$$

$$e.g.\ [1, 2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq 10$$

Remark :

- Hyperplane is one key building block of convex optimization. (theory, algorithms, applications for machine learning, deep learning, …)
- Each hyperplane separates the space into two half spaces.
- If $n \geq p$, p hyperplanes can separate the space into 2^p disjoint regions.

2. Sets and Definitions

V. Polyhedra (plural) : Poly (many) Hedron (face)

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

$$A = \begin{bmatrix} a_1^T \\ a_2^T \\ \dots \\ a_m^T \end{bmatrix} \qquad \mathrm{C} = \begin{bmatrix} c_1^T \\ c_2^T \\ \dots \\ c_p^T \end{bmatrix}$$

2. Sets and Definitions

VI. Matrix Space : Positive Semidefinite Cone

① $S^n = \{X \in R^{n\times n} | X = X^T\}$ Symmetric Matrix

② $S_+^n = \{X \in S^n | X \succeq 0\}$ $\quad i.e.\ y^T X y \geq 0, \forall y$

$S_{++}^n = \{X \in S^n | X \succ 0\}$ $\quad i.e.\ y^T X y > 0, \forall y \neq 0$

Ex: $X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in S_+^2 \Leftrightarrow x \geq 0, z \geq 0, xz \geq y^2$

$$[a\ b]X\begin{bmatrix} a \\ b \end{bmatrix} = a^2x + b^2z + 2aby \geq 0, \forall a, b \in \mathbb{R}$$

2. Sets and Definitions

Ex: $X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in S_+^2 \Leftrightarrow x \geq 0, z \geq 0, xz \geq y^2$

$$[a\ b]X \begin{bmatrix} a \\ b \end{bmatrix} = a^2x + b^2z + 2aby \geq 0, \forall a, b \in \mathbb{R}$$

Proof : Let $R = \begin{bmatrix} 1 & -x^{-1}y \\ 0 & 1 \end{bmatrix}$

We have $[a\ b]X \begin{bmatrix} a \\ b \end{bmatrix} = [a\ b]R^{-T}R^TXRR^{-1} \begin{bmatrix} a \\ b \end{bmatrix}$

$$= [a\ b]R^{-T} \begin{bmatrix} x & 0 \\ 0 & z - x^{-1}y^2 \end{bmatrix} R^{-1} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ -x^{-1}y & 1 \end{bmatrix} \begin{bmatrix} x & y \\ y & z \end{bmatrix} \begin{bmatrix} 1 & -x^{-1}y \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} x & 0 \\ 0 & z - x^{-1}y^2 \end{bmatrix}$$

3. Separating Hyperplane

$\{x|a^T x = b\}$ (Classification, Optimization, Duality)

Theorem : Given two convex sets $C \cap D = \emptyset$ in R^n

$$\exists a \in R^n, b \in R, \quad s.t. \; a^T x \le b, \forall x \in C$$

$$a^T x \ge b, \forall x \in D$$

Actually, $a = d - c, b = \frac{\|d\|_2^2 - \|c\|_2^2}{2}$

i.e. $f(x) = a^T x - b = (d - c)^T (x - \frac{d+c}{2})$

For $dist(C, D) = \inf\{\|u - v\|_2 \mid u \in C, v \in D\}$

3. Separating Hyperplane

Proof : $\forall v \in D, a^T v \geq a^T d$ should be true

By contradiction, suppose that $a^T v < a^T d$

Then we can show that $d + t(v - d)$ is close to c for $t > 0$

Because $\frac{d}{dt} \|d + t(v - d) - c\|_2^2 = 2(d - c)^T (v - d) < 0$

We have $\|d + t(v - d) - c\|_2 < \|d - c\|_2$ for tiny $t > 0$

3. Supporting Hyperplane

Given set $C \in R^n$, and a point x_0 on the boundary of C, the hyperplan $\{x | a^T x = a^T x_0\}$ is called supporting hyperplane of C if $a^T x \leq a^T x_0, \forall x \in C$.

Supporting Hyperplane Theorem: For any nonempty convex set C, and a point x_0 on the boundary of C,

There exists a support hyperplane to C at x_0.

Proof: A separating hyperplane that separates interior C and $\{x_0\}$ is a supporting hyperplane.

4. Dual Cones

Given Cone K (i.e. $K = \{\sum_{i=1}^{k} \theta_i u_i \,|\theta_i > 0, u_i \in R^n, \forall i\}$)

$K^* = \{y|x^T y \geq 0, \forall x \in K\}$

Ex: 1. $K = R_+^n : K^* = R_+^n$

2. $K = S_+^n : K^* = S_+^n$

3. $K = \{(x,t)|\|x\|_2 \leq t\} : K^* = \{(x,t)|\|x\|_2 \leq t\}$

4. $K = \{(x,t)|\|x\|_1 \leq t\} : K^* = \{(x,t)|\|x\|_\infty \leq t\}$

4. Dual Cones

Show that cone $K = \{(x,t) \mid \|x\|_1 \leq t\}$ has its dual

$$K^* = \{(x,t) \mid \|x\|_\infty \leq t\}$$

Proof :

Statement $x^T u + tv \geq 0, \forall \|x\|_1 \leq t \;\leftrightarrow\; \|u\|_\infty \leq v$

L=>R By contradiction, suppose that $\|u\|_\infty > v$

We can find $\exists x \; s.t \; \|x\|_1 \leq 1, x^T u > v$

Setting t=1, then we have $u^T(-x) + v < 0$.

R=>L Given $\|x\|_1 \leq t, \|u\|_\infty \leq v$

$$u^T \|-x/t\|_1 \leq \|u\|_\infty \leq v$$

Thus, $u^T(-x) \leq vt$

4. Dual Cones

Definition: $x \leq_K y$ if $y - x \in K$

Theorem: $x \leq_K y$ iff $\lambda^T x \leq \lambda^T y, \forall \lambda \in K^*$

Examples

4. Dual Cones

The polyhedral cone $V = \{x|Ax \geq 0\}$ has its dual cone

$V^* = \{A^T v | v \geq 0\}$

Proof : By definition

$$V^* = \{y | x^T y \geq 0, \forall x \in V\}$$

Thus $V^* = \{y | x^T y \geq 0, \forall Ax \geq 0\}$

Let $y = A^T v$, we have $x^T y = x^T A^T v > 0$ if $v \geq 0$

Ex: $A = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}$ i.e. $x_1 + 2x_2 \geq 0, x_1 - x_2 \geq 0$

$A^T = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}$ i.e. $\{\theta_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \theta_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} | \theta_1, \theta_2 \geq 0\}$

4. Dual Cones

Remark: $\{x_0 + \Delta x | \Delta x \in K\}$

(1) K cone can be translated to x_0

(2) If $a \in K^*$, then $a^T x \geq 0, \forall x \in K$, i.e. $-ax$ is a supporting hyperplane of cone K

(3) Suppose that the current feasible search region is at corner x_0

and $\{x_0 + \Delta x | \Delta x \in K, ||\Delta x|| < r\}$ is a local feasible region of a convex set

If $\nabla f_0(x_0) \in K^*$, i.e. $\nabla f_0(x_0)^T \Delta x \geq 0, \forall \Delta x \in K$,

Then x_0 is an optimal solution