Optimization

Qingfu Zhang

Dept of CS, CityU

2025

Overview

Table of Contents

Topics in This Course

- Widely used computational problem solving methodoloy.
- Fundamental techniques in AI and Computer Science.

- Traditional Opt (8 weeks)
- Opt Algorithms for Machine Learning (1-2 weeks)
- Heuristics and Multiobjective Opt (1-2 weeks)

Pre-requisites: basic calculus and linear algebra.

Ref Books and Resources

- Convex Optimization, Stephen Boyd and Lieven Vandenberghe, Cambridge University Press.
- Convex Optimization: Algorithms and Complexity, Sébastien Bubeck.
- Convex Analysis and Nonlinear Optimization, Theory and Examples, Jonathan Borwein , Adrian Lewis.
- Handbook of Metaheuristics (International Series in Operations Research & Management Science) 2nd ed. 2010, Michel Gendreau (Editor), Jean-Yves Potvin (Editor).
- Papers et al.

Acknowledgment: Thanks to Prof. S Boyd for course materials in this course. Most materials from week 1 to 8 are from his slides for convex opt.

Assessment

- Course work: 40%
  - Tutorial exercises and take-home assignments 20%
  - Midterm 20%
- Examination: 60%

How to study

- Read notes before lecture and take notes during lecture.
- Be active. Ask questions if you have.
- Do some mini-research on optimization.

An optimization problem can be written as

$$\begin{aligned} &\min \ \ f(x) \\ &s.t. \ \ x \in \Omega \end{aligned} \tag{1}$$

where $f : \Omega \to R$ is the objective function.

- globally optimal solution x^*: $f(x^*) \leq f(x)$ for any $x \in \Omega$
- local optimal solution x^*: there exists $N(x^*)$, a neighborhood of x^*, such that $f(x^*) \leq f(x)$ for any $x \in \Omega \cap N(x^*)$.
- Discrete opt, continuous opt. multi-objective opt

An example:

Table of Contents

Vector Space

A vector space is a nonempty set V of objects (called vectors) with two operations: addition $(+)$ and multiplication by real numbers $(*)$, subject to the following properties for any $u, v, w \in V$ and $a, b \in R$:

1. $u + v \in V$, $a * u \in V$.
2. $u + v = v + u$
3. $u + (v + w) = (u + v) + w$
4. There exists a zero vector $\mathbf{0}$ such that $u + \mathbf{0} = u$.
5. For each $u \in V$, there exists $-u \in V$ such that $u + (-u) = \mathbf{0}$
6. $a(u + v) = au + av$, $(a + b)u = au + bu$
7. $1u = u$,
8. $(ab)u = a(bu)$

Q: Is zero vector unique? Why

- Finite dimensional vector space
- Examples: R^n (n-D real vector space), S^n

Finite-Dimensional Euclidean Space

Let V be a vector space, an inner product $\langle\cdot,\cdot\rangle$ is a mapping: $V \times V \to R$ such that, for any $u, v, w \in V$ and $a \in R$

- $\langle v, v\rangle \geq 0$; "=" iff $v = 0$,
- $\langle v, u\rangle = \langle u, v\rangle$.
- $\langle av + bu, w\rangle = a\langle v, w\rangle + b\langle u, w\rangle$.

Norm induced by the inner product:

$$\|v\| = \sqrt{\langle v, v\rangle}$$

We can prove:

$$|\langle v, u\rangle| \leq \|v\| * \|u\|$$

Finite-dimensional Euclidean space E: a finite-dimensional vector space with an inner product $\langle\cdot,\cdot\rangle$.

Two Examples:

- R^n=? $\langle x, y\rangle = x^T y = \sum x_i y_i$.
- S^n=the set of all the $n \times n$ symmetric matrices. $\langle A, B\rangle = tr(AB)$.

Table of Contents

Convex sets

Line segment between x_1 and x_2 : the set of all points of form

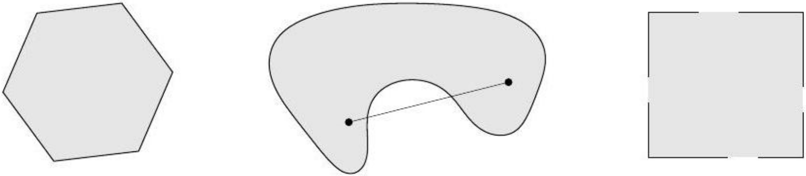
$$x = \theta x_1 + (1-\theta)x_2$$

with $0 \leq \theta \leq 1$.

Convex set: contains line segment between any two points in the set

$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \Longrightarrow \quad \theta x_1 + (1-\theta)x_2 \in C$$

examples: (one convex, two nonconvex sets)



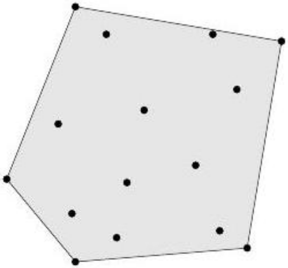
Convex combination and convex hull

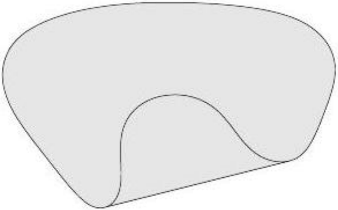
Convex combination of $x_1, \ldots, x_k$: any point x of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

with $\theta_1 + \cdots + \theta_k = 1, \theta_i \geq 0$.

Convex hull: conv(S**)**=set of all convex combinations of points in S



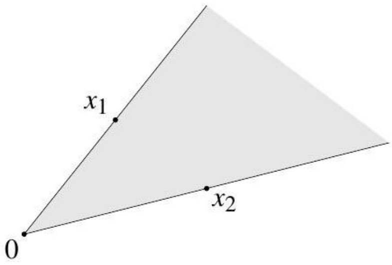


Convex cone

conic (nonnegative) combination of x_1 **and** x_2 : any point of the form

$$x = \theta_1 x_1 + \theta_2 x_2$$

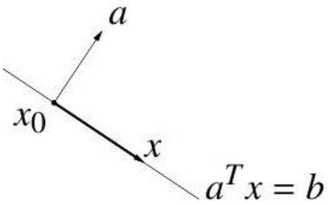
with $\theta_1 \geq 0, \theta_2 \geq 0$



Convex cone: set that contains all conic combinations of points in the set

Hyperplanes and halfspaces

hyperplane: set of the form $\{x \mid a^T x = b\}$, with $a \neq 0$.
halfspace: set of the form $\{x | a^T x \leq b\}$, with $a \neq 0$



- a is the normal vector
- hyperplanes are convex; halfspaces are convex

Euclidean balls and ellipsoids

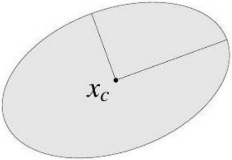
(Euclidean) ball with center x_c and radius r :

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

ellipsoid: set of the form

$$\left\{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\right\}$$

with $P \in \mathbf{S}^n_{++}$ (i.e., P symmetric positive definite)



another representation: $\{x_c + Au \mid \|u\|_2 \leq 1\}$ with A square and nonsingular

Norm balls and norm cones

norm: a function $\|\cdot\|$ that satisfies:

- $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$
- $\|tx\| = |t|\|x\|$ for $t \in \mathbf{R}$
- $\|x + y\| \leq \|x\| + \|y\|$

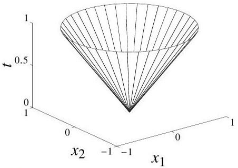
notation: $\|\cdot\|$ is general (unspecified) norm; $\|\cdot\|_{\text{symb}}$ is particular norm

- norm ball with center x_c and radius r : $\{x \mid \|x - x_c\| \leq r\}$
- norm cone: $\{(x, t) \mid \|x\| \leq t\}$
- norm balls and cones are convex

Euclidean norm cone

$$\{(x, t) \mid \|x\|_2 \leq t\} \subset \mathbf{R}^{n+1}$$

is called second-order cone



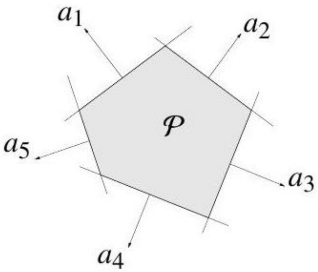
Polyhedra

- polyhedron is solution set of finitely many linear inequalities and equalities

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

($A \in \mathbf{R}^{m \times n}$, $C \in \mathbf{R}^{p \times n}$, $\preceq$ is componentwise inequality)

- intersection of finite number of halfspaces and hyperplanes
- example with no equality constraints; a_i^T are rows of A



Positive semidefinite cone

- $\mathbf{S}^n$ is set of symmetric $n \times n$ matrices
- $\mathbf{S}^n_+ = \{X \in \mathbf{S}^n \mid X \succeq 0\}$: positive semidefinite (symmetric) $n \times n$ matrices. $X \in \mathbf{S}^n_+ \iff z^T X z \geq 0$ for all z
- $\mathbf{S}^n_+$ is a convex cone, the positive semidefinite cone
- $\mathbf{S}^n_{++} = \{X \in \mathbf{S}^n \mid X \succ 0\}$: positive definite (symmetric) $n \times n$ matrices

Example:

$$\begin{bmatrix} x & y \\ y & z \end{bmatrix} \in \mathbf{S}^2_+$$

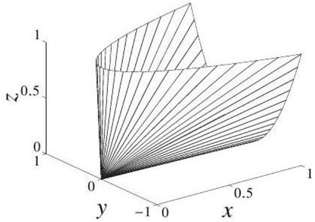


Table of Contents

Showing a set is convex

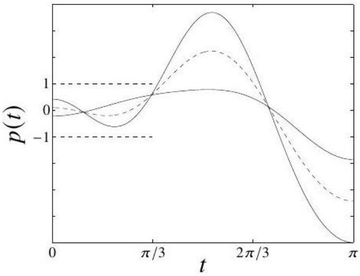
methods for establishing convexity of a set C

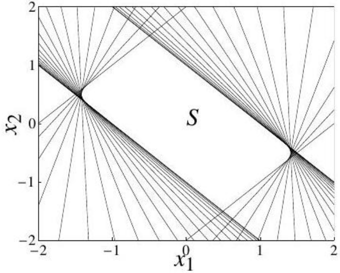
1. apply definition: show
   $x_1, x_2 \in C, 0 \leq \theta \leq 1 \implies \theta x_1 + (1-\theta)x_2 \in C$
   - recommended only for very simple sets
2. use convex functions (next lecture)
3. show that C is obtained from simple convex sets (hyperplanes, halfspaces, norm balls, ...) by operations that preserve convexity
   - intersection
   - affine mapping
   - perspective mapping
   - linear-fractional mapping

you'll mostly use methods 2 and 3

Intersection

- the intersection of (any number of) convex sets is convex
- example:
  - $S = \{x \in \mathbf{R}^m \mid |p(t)| \le 1 \text{ for } |t| \le \pi/3\}$ with $p(t) = x_1 \cos t + \cdots + x_m \cos mt$
  - write $S = \bigcap_{|t| \le \pi/3} \{x \mid |p(t)| \le 1\}$, i.e., an intersection of (convex) slabs
- picture for $m = 2$:





Affine mappings

suppose $f : \mathbf{R}^n \to \mathbf{R}^m$ is affine, i.e., $f(x) = Ax + b$ with $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$

- the image of a convex set under f is convex

$$S \subseteq \mathbf{R}^n \text{ convex } \implies f(S) = \{f(x) \mid x \in S\} \text{ convex}$$

- the inverse image $f^{-1}(C)$ of a convex set under f is convex

$$C \subseteq \mathbf{R}^m \text{ convex } \implies f^{-1}(C) = \{x \in \mathbf{R}^n \mid f(x) \in C\} \text{ convex}$$

Examples

- scaling, translation: $aS + b = \{ax + b \mid x \in S\}$, $a, b \in \mathbf{R}$
- projection onto some coordinates: $\{x \mid (x, y) \in S\}$
- if $S \subseteq \mathbf{R}^n$ is convex and $c \in \mathbf{R}^n$, $c^T S = \left\{c^T x \mid x \in S\right\}$ is an interval
- solution set of linear matrix inequality $\{x \mid x_1 A_1 + \cdots + x_m A_m \preceq B\}$ with $A_i, B \in \mathbf{S}^p$
- hyperbolic cone $\left\{x \mid x^T P x \leq \left(c^T x\right)^2, c^T x \geq 0\right\}$ with $P \in \mathbf{S}_+^n$

Perspective and linear-fractional function

perspective function $P : \mathbf{R}^{n+1} \to \mathbf{R}^n$:

$$P(x, t) = x/t, \quad \mathrm{dom}\, P = \{(x, t) \mid t > 0\}$$

(it is better to write it as ?)

- images and inverse images of convex sets under perspective are convex
- linear-fractional function $f : \mathbf{R}^n \to \mathbf{R}^m$:

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

- images and inverse images of convex sets under linear-fractional functions are convex

Linear-fractional function example

$$f(x) = \frac{1}{x_1 + x_2 + 1} x$$

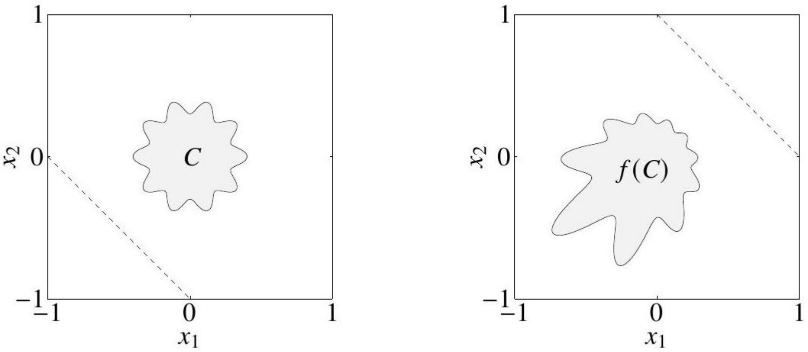


Table of Contents

Proper cones

a convex cone $K \subseteq \mathbf{R}^n$ is a proper cone if

- K is closed (contains its boundary)
- K is solid (has nonempty interior)
- K is pointed (contains no line)

examples

- nonnegative orthant
  $K = \mathbf{R}^n_+ = \{x \in \mathbf{R}^n \mid x_i \geq 0, i = 1, \ldots, n\}$
- positive semidefinite cone $K = \mathbf{S}^n_+$
- nonnegative polynomials on $[0, 1]$:

$$K = \left\{x \in \mathbf{R}^n \mid x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1} \geq 0 \text{ for } t \in [0, 1]\right\}$$

Generalized inequality

- (nonstrict and strict) generalized inequality defined by a proper cone K :

$$x \preceq_K y \iff y - x \in K, \quad x \prec_K y \iff y - x \in \operatorname{int} K$$

- examples
  - componentwise inequality $(K = \mathbf{R}_+^n)$: $x \preceq_{\mathbf{R}_+^n} y \iff x_i \leq y_i, \quad i = 1, \ldots, n$
  - matrix inequality $(K = \mathbf{S}_+^n)$: $X \preceq_{\mathbf{S}_+^n} Y \iff Y - X$ positive semidefinite

  these two types are so common that we drop the subscript in $\preceq_K$

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

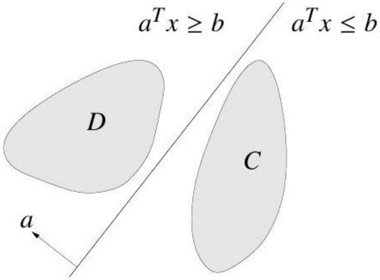
$$x \preceq_K y, \quad u \preceq_K v \quad \Longrightarrow \quad x + u \preceq_K y + v$$

Table of Contents

Separating hyperplane theorem

Theorem: if C and D are nonempty disjoint (i.e., $C \cap D = \emptyset$) convex sets, there exist $a \neq 0$, b s.t.

$$a^T x \leq b \text{ for } x \in C, \quad a^T x \geq b \text{ for } x \in D$$



- the hyperplane $\{x \mid a^T x = b\}$ separates C and D
- strict separation ("$\leq$" and "$\geq$" are replaced by "$<$" and "$>$", respectively.) requires additional assumptions (e.g., C is closed, D is a singleton)

Outline of Proof

Case 1: D has only one element d and $dist(C, d) > 0$ (i.e. $d \notin Bd(C)$

- Let $cl(C)$ be the smallest closed set including C, i.e.

$$cl(C) = C \cup bd(C)$$

- $cl(C)$ is convex, $dist(cl(C), d) = dist(C, d)$, and $C \subset cl(C)$
- without of loss of generality, we assume C is closed.

1. There exists a unique $c \in C$ such that $dist(c, d) = dist(C, d) > 0$ (why)
2. Let $a = d - c \neq 0$, $x_0 = \frac{d+c}{2}$. Consider $f(x) = \frac{a^T}{\|a\|}(x - x_0)$.
3. $f(d) < 0$
4. for any $x \in C$, $f(x) > 0$ (How to prove? by contradiction)

strict separation!

Outline of Proof

Case 2: D has only one element d and $dist(C, d) = 0$ (i.e. $d \in bd(C)$.)

- For any integer m, there exists $d_m \notin cl(C)$ and $dist(d_m, d) < \frac{1}{m}$.
- By Case 1, there exists affine function $f_m(x) = \frac{a_m^T}{\|a_m\|}(x - x_m)$ such that

$$f_m(d_m) < 0, \text{ and } f_m(x) > 0 \text{ for all } x \in C.$$

- There exists a sub-sequence $m_1 < m_2 < ... < m_k ....$ such that

$$\frac{a_m^T}{\|a_m\|} \to a, \text{ and } x_m \to x_0$$

- Let $f(x) = a^T(x - x_0)$, then

$$f(d) \geq 0, \text{ and } f(x) \leq 0 \text{ for all } x \in C.$$

Outline of Proof

Case 3: $D \cap C = \emptyset$ (general case).

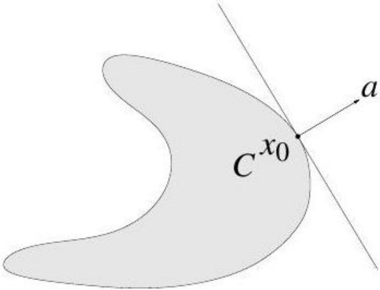
- Let $F = D - C = \{x - y | x \in D \text{ and } y \in C\}$ (F is convex)
- $0 \notin F$
- there exists $a^T x = b$ separating 0 and F. i.e.

$$f(0) = -b < 0, \text{ and } f(x - y) > b \text{ for any } x \in D, y \in C$$

- $a^T x \geq a^t y$ for any $x \in D, y \in C$
- Let $\alpha = inf\{a^T x | x \in D\}$, then $a^T x = \alpha$ is an separating plane for D and C.

Supporting hyperplane theorem

- suppose x_0 is a boundary point of set $C \subset \mathbf{R}^n$
- supporting hyperplane to C at x_0 has form $\{x \mid a^T x = a^T x_0\}$, where $a \neq 0$ and $a^T x \leq a^T x_0$ for all $x \in C$



Theorem: supporting hyperplane theorem: if C is convex, then there exists a supporting hyperplane at every boundary point of C
Pf: Noting in the proof of case 2, we only require that

$$d \in Bd(C).$$

Theorem of Alternative

Gordan Theorem: For any $a^1, \dots, a^m \in R^n$, exactly one of the following systems has a solution:

$$(1) : \sum_{i=1}^{m} \lambda_i a^i = 0, \sum_{i=1}^{m} \lambda_i = 1, 0 \leq \lambda_1, \dots, \lambda_m \in R$$

$$(2) : {a^i}^T x < 0 \text{ for } i = 1, \dots, m, x \in R^n$$

Outline of Pf: Noting that

- system (1) has a solution $\Leftrightarrow$ $0 \in conv(a^1, \dots, a^m)$.
- system (2) has a solution $\Leftrightarrow$ there is an separating plane btw 0 and $conv(a^1, \dots, a^m)$.

Farkas Lemma

Theorem: For any $a^1, \ldots, a^m$ and $c \in R^n$, exactly one of the following systems has a solution:

$$(1) : \sum_{i=1}^{m} \lambda_i a^i = c, 0 \leq \lambda_1, \ldots, \lambda_m \in R$$

$$(2) : (a^i)^T x \leq 0 \text{ for } i = 1, \ldots, m, c^T x > 0, x \in R^n$$

Outline of Pf: Noting that

- system (1) has a solution $\Leftrightarrow$ $0 \in \mathit{conichull}(a^1, \ldots, a^m)$.
- system (2) has a solution $\Leftrightarrow$ there is an separating plane btw c and $\mathit{conichull}(a^1, \ldots, a^m)$.

Table of Contents

What we learned today

- Finite-D Euclidean space: R^n and S^n.
- Convex set
- How to prove a set is convex
- Separating and supporting hyperplanes
- Sperarting plane and Farkas Lemma.