1. Introduction

- mathematical optimization
- least-squares and linear programming
- convex optimization
- example
- course goals and topics
- nonlinear optimization
- brief history of convex optimization

Mathematical optimization

(mathematical) optimization problem $\iff \max_x \frac{1}{f_0(x)}$

$$\begin{array}{ll} \text{minimize} & f_0(x) \iff \max_x -f_0(x) \\ \text{subject to} & f_i(x) \le b_i, \quad i = 1, \ldots, m \end{array}$$

$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

- $x = (x_1, \ldots, x_n)$: optimization variables
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: objective function
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$: constraint functions

solution or **optimal point** $x^\star$ has smallest value of f_0 among all vectors that satisfy the constraints

Examples

portfolio optimization

- variables: amounts invested in different assets
- constraints: budget, max./min. investment per asset, minimum return
- objective: overall risk or return variance

device sizing in electronic circuits

- variables: device widths and lengths
- constraints: manufacturing limits, timing requirements, maximum area
- objective: power consumption

data fitting

- variables: model parameters
- constraints: prior information, parameter limits
- objective: measure of misfit or prediction error, plus regularization term

Solving optimization problems

general optimization problem

- very difficult to solve

global

- methods involve some compromise, $e.g.$, very long computation time, or not always finding the solution (which may not matter in practice)

local

exceptions: certain problem classes can be solved efficiently and reliably

- least-squares problems
- linear programming problems
- convex optimization problems

Least-squares

skinny : $k >> n$

$\mathrm{rank}(A) = n$

$R^{k\times n}$ $\quad R^{k}$

$$\text{minimize} \quad \|Ax - b\|_2^2$$

x

$\boxed{(A^T A)x = A^T b}$ $\quad \frac{O(n^3)}{3}$

$O(n^2 k)$ ① $\quad O(nk)$ ②

solving least-squares problems

- analytical solution: $x^\star = (A^T A)^{-1} A^T b$
- reliable and efficient algorithms and software
- computation time proportional to $n^2 k$ ($A \in \mathbf{R}^{k\times n}$); less if structured
- a mature technology

A sparse

$x^T\beta$

$y = f(x) + \varepsilon \sim N(0, \sigma^2)$

$x^T\beta$

$\sim N(f(x), \sigma^2)$

using least-squares

- least-squares problems are easy to recognize
- a few standard techniques increase flexibility ($e.g.$, including weights, adding regularization terms)

Linear programming

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m \end{array}$$

$Ax \preceq b$

$\in R^{m \times n}$

solving linear programs

- no analytical formula for solution
- reliable and efficient algorithms and software
- computation time proportional to $n^2 m$ if $m \geq n$; less with structure
- a mature technology

$\min_{x} \boxed{\max_{i=1,\ldots,k} |a_i^T x - b_i|} = \|Ax - b\|_\infty$

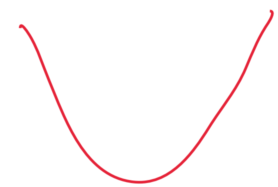
$\Rightarrow \min_{x,t} t, \quad \text{s.t. } |a_i^T x - b_i| \leq t, \quad i = 1, \ldots, m$

using linear programming

- not as easy to recognize as least-squares problems
- a few standard tricks used to convert problems into linear programs ($e.g.$, problems involving ℓ_1- or ℓ_∞-norms, piecewise-linear functions)

Convex optimization problem

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m
\end{array}
$$



- objective and constraint functions are convex:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y)$$

  if $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$

  LS LP

- includes least-squares problems and linear programs as special cases

solving convex optimization problems

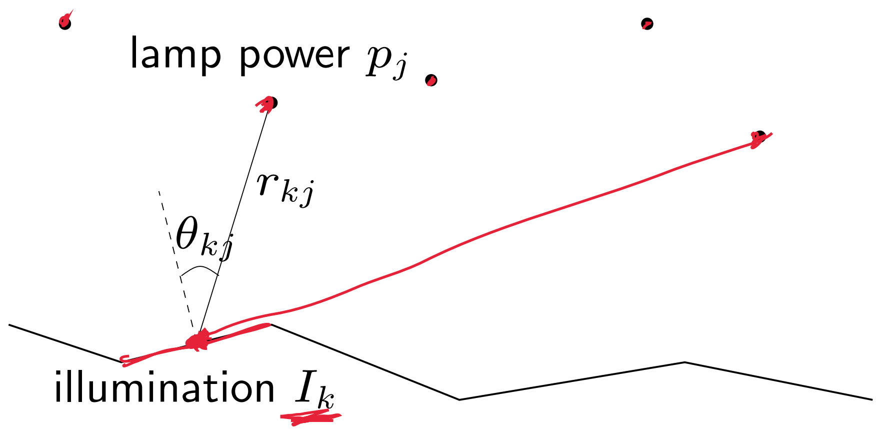
- no analytical solution
- reliable and efficient algorithms interior-point method.
- computation time (roughly) proportional to $\max\{n^3, n^2m, F\}$, where F is cost of evaluating f_i's and their first and second derivatives
- almost a technology

using convex optimization

- often difficult to recognize
- many tricks for transforming problems into convex form
- surprisingly many problems can be solved via convex optimization

Example

m lamps illuminating n (small, flat) patches



intensity I_k at patch k depends linearly on lamp powers p_j:

$$I_k = \sum_{j=1}^{m} a_{kj} p_j, \qquad a_{kj} = r_{kj}^{-2} \max\{\cos\theta_{kj}, 0\}$$

problem: achieve desired illumination I_{des} with bounded lamp powers

$$\begin{array}{ll} \text{minimize} & \max_{k=1,\ldots,n} |\log I_k - \log I_{\text{des}}| \\ \text{subject to} & 0 \leq p_j \leq p_{\max}, \quad j = 1, \ldots, m \end{array}$$

how to solve?

1. use uniform power: $p_j = p$, vary p
2. use least-squares:

$$\text{minimize} \quad \sum_{k=1}^{n} (I_k - I_{\text{des}})^2$$

   round p_j if $p_j > p_{\text{max}}$ or $p_j < 0$

3. use weighted least-squares:

$$\text{minimize} \quad \sum_{k=1}^{n} (I_k - I_{\text{des}})^2 + \sum_{j=1}^{m} w_j (p_j - p_{\text{max}}/2)^2$$

   iteratively adjust weights w_j until $0 \le p_j \le p_{\text{max}}$

4. use linear programming:

$$\begin{array}{ll} \text{minimize} & \max_{k=1,\ldots,n} |I_k - I_{\text{des}}| \\ \text{subject to} & 0 \le p_j \le p_{\text{max}}, \quad j = 1, \ldots, m \end{array}$$

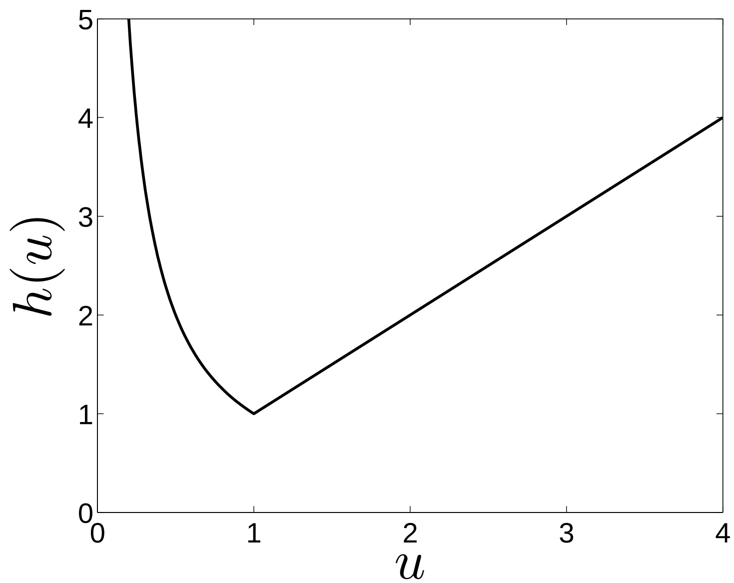
   which can be solved via linear programming

of course these are approximate (suboptimal) 'solutions'

5. use convex optimization: problem is equivalent to

$$\begin{array}{ll} \text{minimize} & f_0(p) = \max_{k=1,\ldots,n} h(I_k/I_{\mathsf{des}}) \\ \text{subject to} & 0 \le p_j \le p_{\mathsf{max}}, \quad j = 1, \ldots, m \end{array}$$

with $h(u) = \max\{u, 1/u\}$



f_0 is convex because maximum of convex functions is convex

exact solution obtained with effort $\approx$ modest factor $\times$ least-squares effort

additional constraints: does adding 1 or 2 below complicate the problem?

1. no more than half of total power is in any 10 lamps
2. no more than half of the lamps are on ($p_j > 0$)

- answer: with (1), still easy to solve; with (2), extremely difficult
- moral: (untrained) intuition doesn't always work; without the proper background very easy problems can appear quite similar to very difficult problems

Course goals and topics

goals

1. recognize/formulate problems (such as the illumination problem) as convex optimization problems
2. develop code for problems of moderate size (1000 lamps, 5000 patches)
3. characterize optimal solution (optimal power distribution), give limits of performance, etc.

topics

1. convex sets, functions, optimization problems
2. examples and applications
3. algorithms

Nonlinear optimization

traditional techniques for general nonconvex problems involve compromises

local optimization methods (nonlinear programming)

- find a point that minimizes f_0 among feasible points near it
- fast, can handle large problems
- require initial guess
- provide no information about distance to (global) optimum

global optimization methods

- find the (global) solution
- worst-case complexity grows exponentially with problem size

these algorithms are often based on solving convex subproblems

Brief history of convex optimization

theory (convex analysis): 1900–1970

algorithms

- 1947: simplex algorithm for linear programming (Dantzig)
- 1970s: ellipsoid method and other subgradient methods
- 1980s & 90s: polynomial-time interior-point methods for convex optimization (Karmarkar 1984, Nesterov & Nemirovski 1994)
- since 2000s: many methods for large-scale convex optimization

applications

- before 1990: mostly in operations research, a few in engineering
- since 1990: many applications in engineering (control, signal processing, communications, circuit design, . . .)
- since 2000s: machine learning and statistics

2. Convex sets

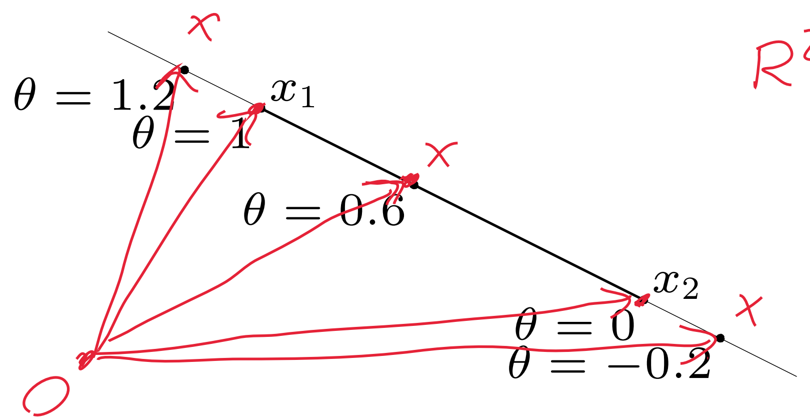
- affine and convex sets
- some important examples
- operations that preserve convexity
- generalized inequalities
- separating and supporting hyperplanes
- dual cones and generalized inequalities

Affine set

$\in R^n$

line through x_1, x_2: all points

$$\boxed{x = \theta x_1 + (1-\theta)x_2} \qquad (\theta \in \mathbf{R})$$



R^2

affine set:

$\{x \mid x_1, x_2 \in C, \theta x_1 + (1-\theta)x_2 \in C, \theta \in R\}$

affine set: contains the line through any two distinct points in the set

example: solution set of linear equations $\{x \mid Ax = b\}$

(conversely, every affine set can be expressed as solution set of system of linear equations)

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

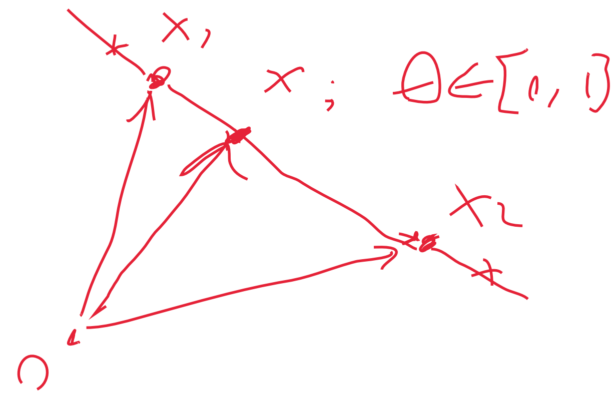
$Ax_1 = b,\ Ax_2 = b$ $A(\theta x_1 + (1-\theta)x_2) = b$

$Ax_1 - b = 0,\ Ax_2 - b = 0$ $= \theta(Ax_1 - b) + (1-\theta)(Ax_2 - b) = 0$

Convex set

line segment between x_1 and x_2: all points

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



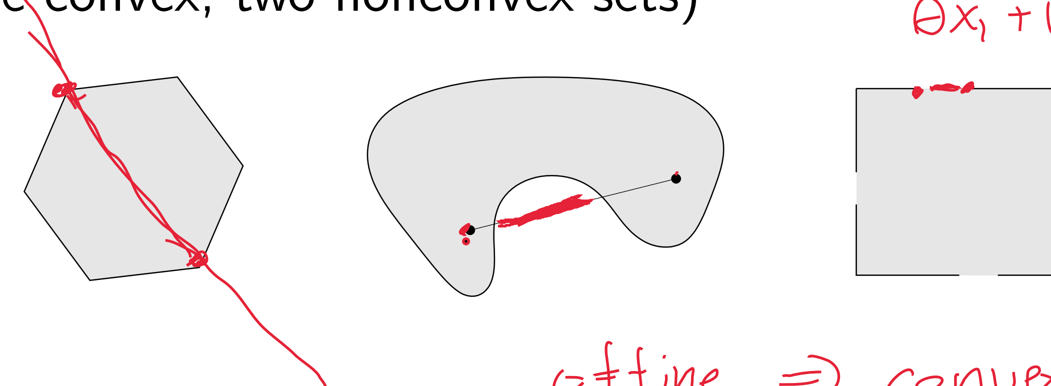
with $0 \le \theta \le 1$

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

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

$\{x \mid x_1, x_2 \in C,\ \theta \in [0,1],\ \theta x_1 + (1-\theta)x_2 \in C\}$

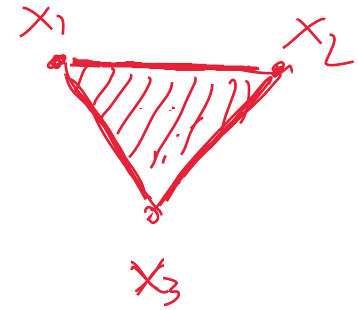
examples (one convex, two nonconvex sets)



affine $\Rightarrow$ convex

Convex combination and convex hull

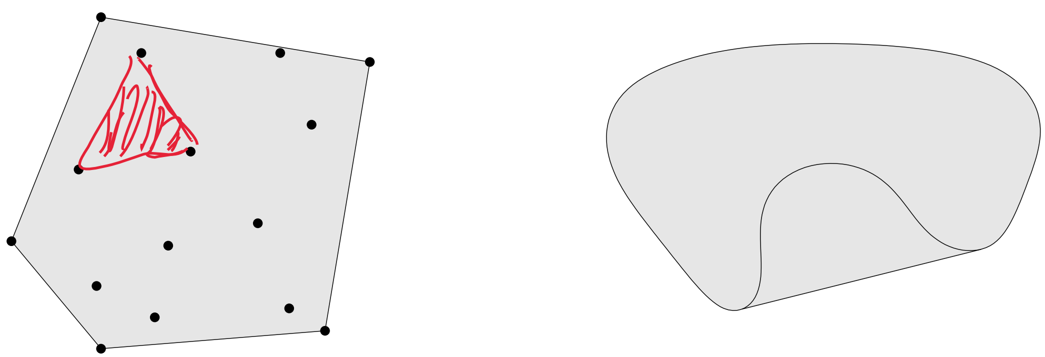
convex combination of x_1,. . . , x_k: any point x of the form



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

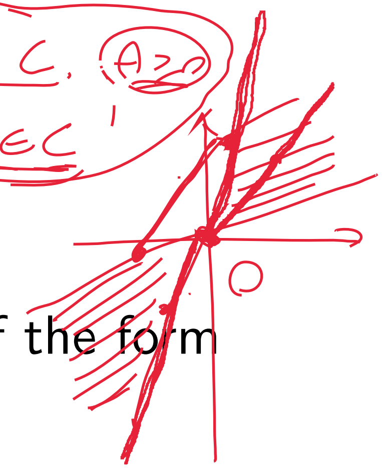
with $\theta_1 + \cdots + \theta_k = 1$, $\theta_i \geq 0$

convex hull $\mathbf{conv}\, S$: set of all convex combinations of points in S



Convex cone

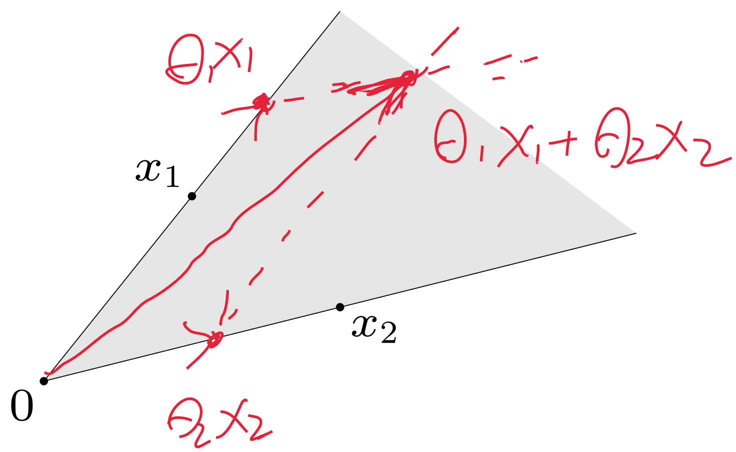
$\boxed{\text{cone}}$: $x \in C$, $\theta \geq 0$, $\theta x \in C$



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

must pass through the origin 0.

Affine Set, Convex Set, Convex Conic Set

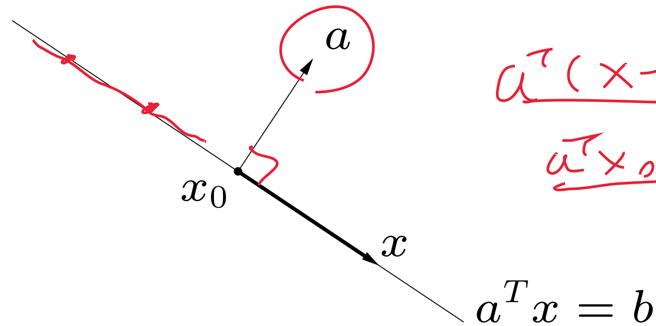
$x_1, x_2 \in C$, $\theta_1 x_1 + \theta_2 x_2 \in C$

Affine	Convex	Convex cone
$\theta_1 + \theta_2 = 1$	$\theta_1 + \theta_2 = 1$ $\theta_1, \theta_2 \geq 0$.	$\theta_1, \theta_2 \geq 0$

1. affine $\Rightarrow$ convex
2. convex cone $\Rightarrow$ convex
3. convex $\not\Rightarrow$ affine
4. convex $\not\Rightarrow$ convex cone

Hyperplanes and halfspaces

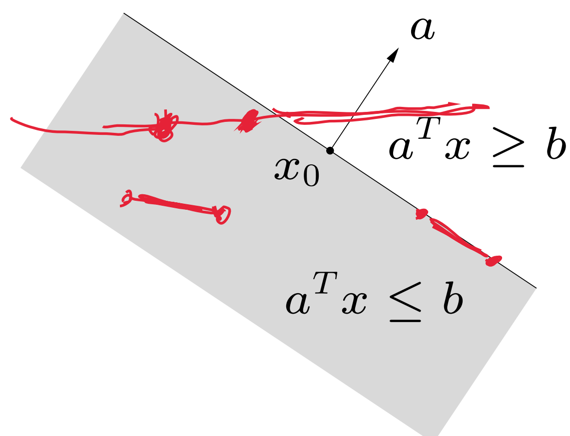
hyperplane: set of the form $\{x \mid a^T x = b\}$ $(a \neq 0)$



$a^T(x - x_0) = 0$

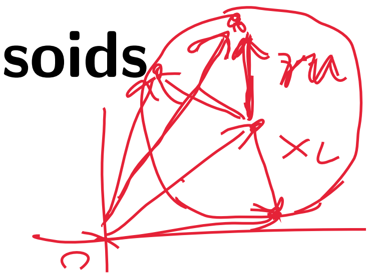
$a^T x_0 = b$

halfspace: set of the form $\{x \mid a^T x \leq b\}$ $(a \neq 0)$



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

Euclidean balls and ellipsoids



$\mathbb{R}^n$ $\quad\quad$ $\mathbb{R}$

(Euclidean) ball with center x_c and radius r:

generalization.

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

$(x-x_c)^T (rI)^{-1} (x-x_c) \le 1$ $\quad\quad$ $x_c + (rI)\cdot u$

ellipsoid: set of the form

$$\{x \mid (x - x_c)^T P^{-1} (x - x_c) \le 1\}$$

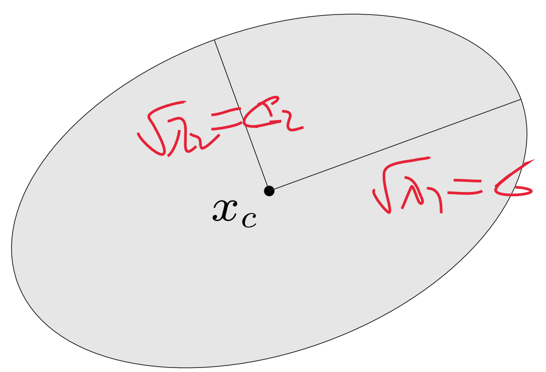
with $P \in \mathbf{S}^n_{++}$ (*i.e.*, P symmetric positive definite)

$x, y : \|x\| \le 1, \ \|y\| \le 1$

$z = \theta x + (1-\theta) y, \ (\theta \ge 0)$

$\|x + y\| \le \|x\| + \|y\|$

$\|tx\| = |t| \, \|x\|$



$\|z\| = \|\theta x + (1-\theta) y\|$

$\le \|\theta x\| + \|(1-\theta) y\|$

$= \theta \|x\| + (1-\theta) \|y\|$

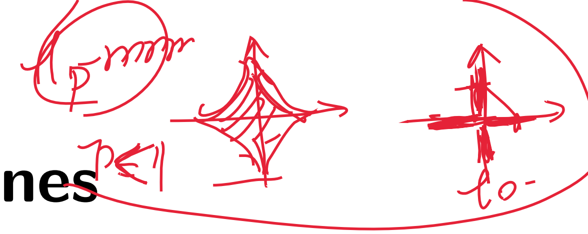
other representation: $\{x_c + Au \mid \|u\|_2 \le 1\}$ with A square and nonsingular

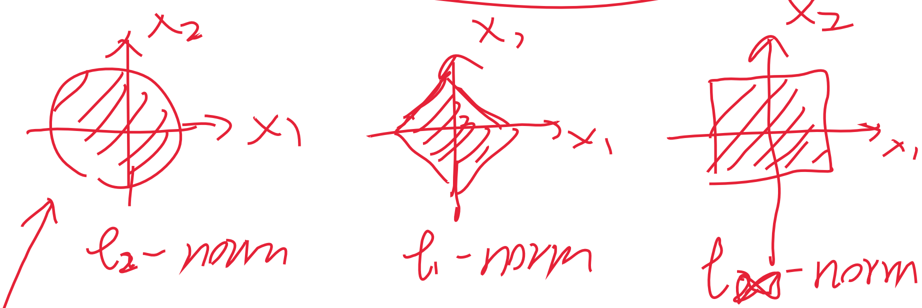
σ

$\le \theta + 1 - \theta = 1$

$|\cdot|$ is the norm in R

Norm balls and norm cones





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

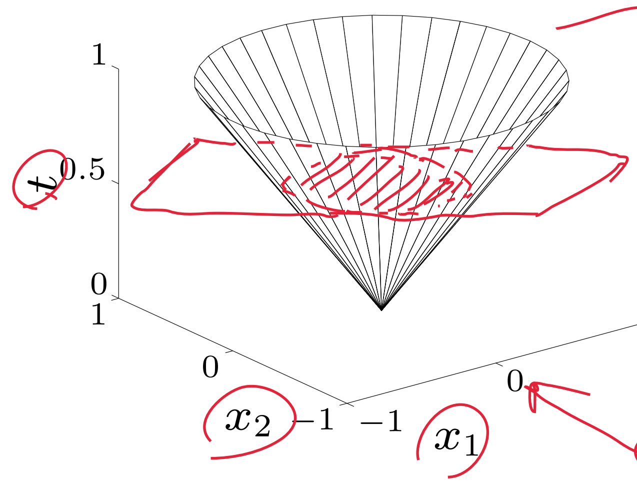
- $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$
- $\|tx\| = |t|\,\|x\|$ for $t \in \mathbf{R}$
- $\|x + y\| \leq \|x\| + \|y\|$

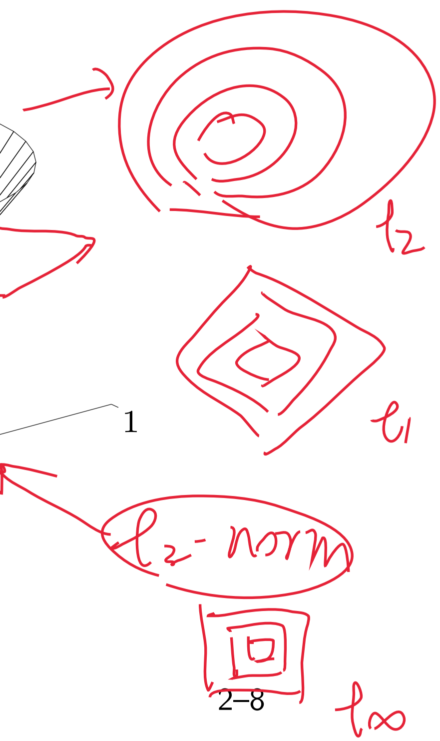
notation: $\|\cdot\|$ is general (unspecified) norm; $\|\cdot\|_{\text{symb}}$ is particular norm

norm ball with center x_c and radius r: $\{x \mid \|x - x_c\| \leq r\}$

norm cone: $\{(x,t) \mid \|x\| \leq t\}$

Euclidean norm cone is called second-order cone (ice-cream cone)





norm balls and cones are convex

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

L_p-norm:

$$\|x\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{\frac{1}{p}}$$

$p \geq 1 \Rightarrow$ norm

$0 \leq p < 1 \not\Rightarrow$ norm.

subadditive

$$\|x+y\| \leq \|x\| + \|y\|$$

Polyhedra

solution set of finitely many linear inequalities and equalities

$a_i^T x \le b_i$, $\forall i$

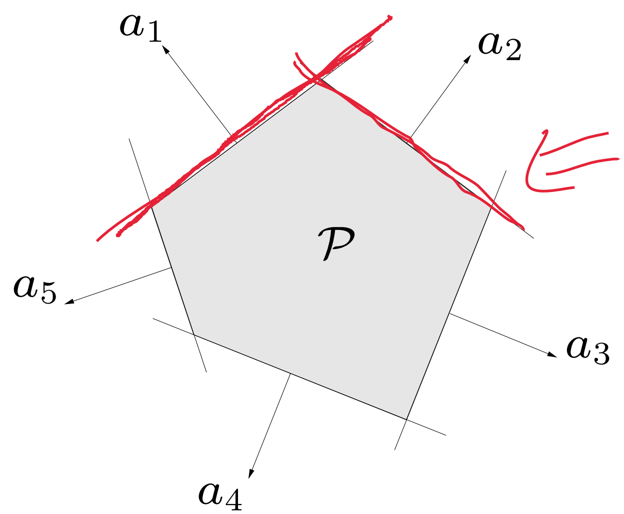
$$Ax \preceq b, \qquad Cx = d$$

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

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

$a_1^T x \le b_1$

R^3



$\Leftarrow \begin{cases} a_i^T x \le b_i, \ i = 1, 2, \dots, 5 \\ c^T x = d \end{cases}$

convex

convex

convex and affine

polyhedron is intersection of finite number of halfspaces and hyperplanes

Positive semidefinite cone

notation:

$X \in PSD$

- $\mathbf{S}^n$ is set of symmetric $n \times n$ matrices
- $\mathbf{S}^n_+ = \{X \in \mathbf{S}^n \mid X \succeq 0\}$: positive semidefinite $n \times n$ matrices

$$X \in \mathbf{S}^n_+ \iff z^T X z \geq 0 \text{ for all } z$$

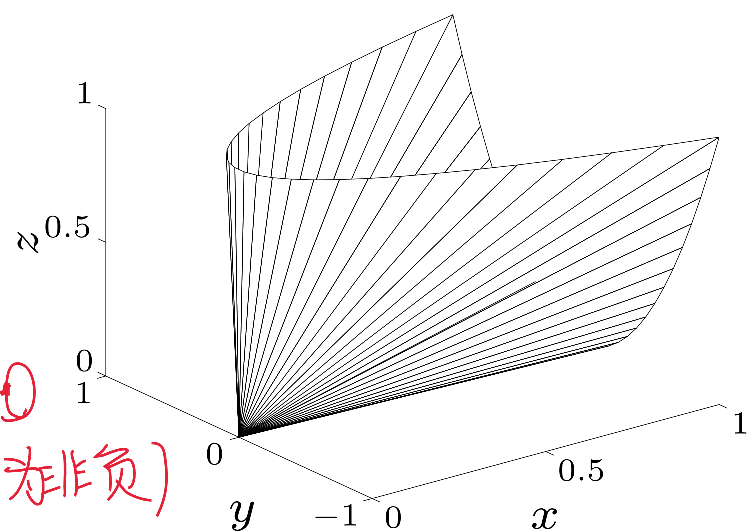
  $\mathbf{S}^n_+$ is a convex cone positive definite

- $\mathbf{S}^n_{++} = \{X \in \mathbf{S}^n \mid X \succ 0\}$: positive definite $n \times n$ matrices

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

$\Rightarrow x, z \geq 0,\ xz - y^2 \geq 0$

(PSD: 所有主子式均为非负)



Quiz:

Is S^n_{++} a convex cone? Show the reason.

Positive semidefinite cone is Convex cone.

Given: $X, Y \in S_+^n \Rightarrow z^T X z \geq 0,\ z^T Y z \geq 0,\ \forall z \neq 0$

whether: $\theta_1 X + \theta_2 Y$, $(\theta_1, \theta_2 \geq 0)$ belongs to S_+^n

---

Proof: $z^T(\theta_1 X + \theta_2 Y) z$

$$= \theta_1 \underbrace{z^T X z}_{\geq 0} + \theta_2 \underbrace{z^T Y z}_{\geq 0}$$

$$\geq 0$$

$$\Rightarrow \theta_1 X + \theta_2 Y \geq 0 \Rightarrow \theta_1 X + \theta_2 Y \in S_+^n$$

Operations that preserve convexity

practical methods for establishing convexity of a set C

1. apply definition

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

2. show that C is obtained from simple convex sets (hyperplanes, halfspaces, norm balls, . . .) by operations that preserve convexity

   - intersection
   - affine functions
   - perspective function
   - linear-fractional functions

Intersection

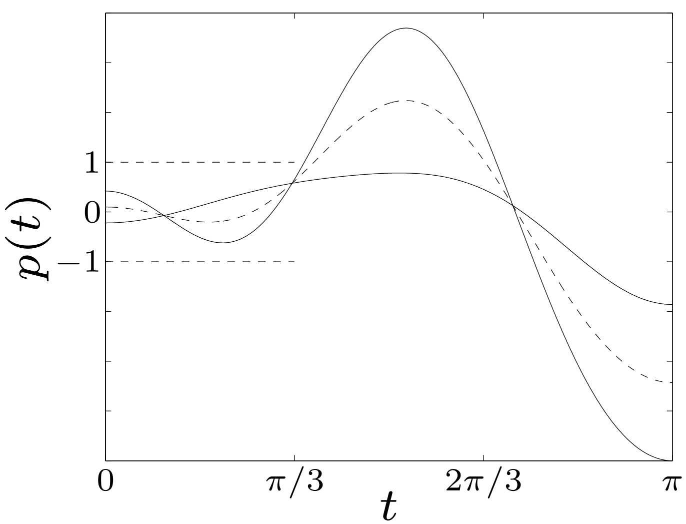
the intersection of (any number of) convex sets is convex

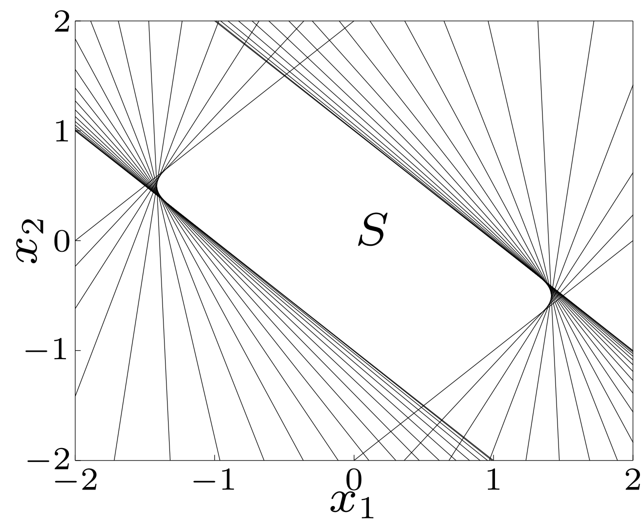
example:

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

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

for $m = 2$:





Affine function

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

- the image of a convex set under f is convex

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

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

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

examples

- scaling, translation, projection
- solution set of linear matrix inequality $\{x \mid x_1A_1 + \cdots + x_mA_m \preceq B\}$ (with $A_i, B \in \mathbf{S}^p$)
- hyperbolic cone $\{x \mid x^TPx \leq (c^Tx)^2,\ c^Tx \geq 0\}$ (with $P \in \mathbf{S}^n_+$)

Perspective and linear-fractional function

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

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

images and inverse images of convex sets under perspective are convex

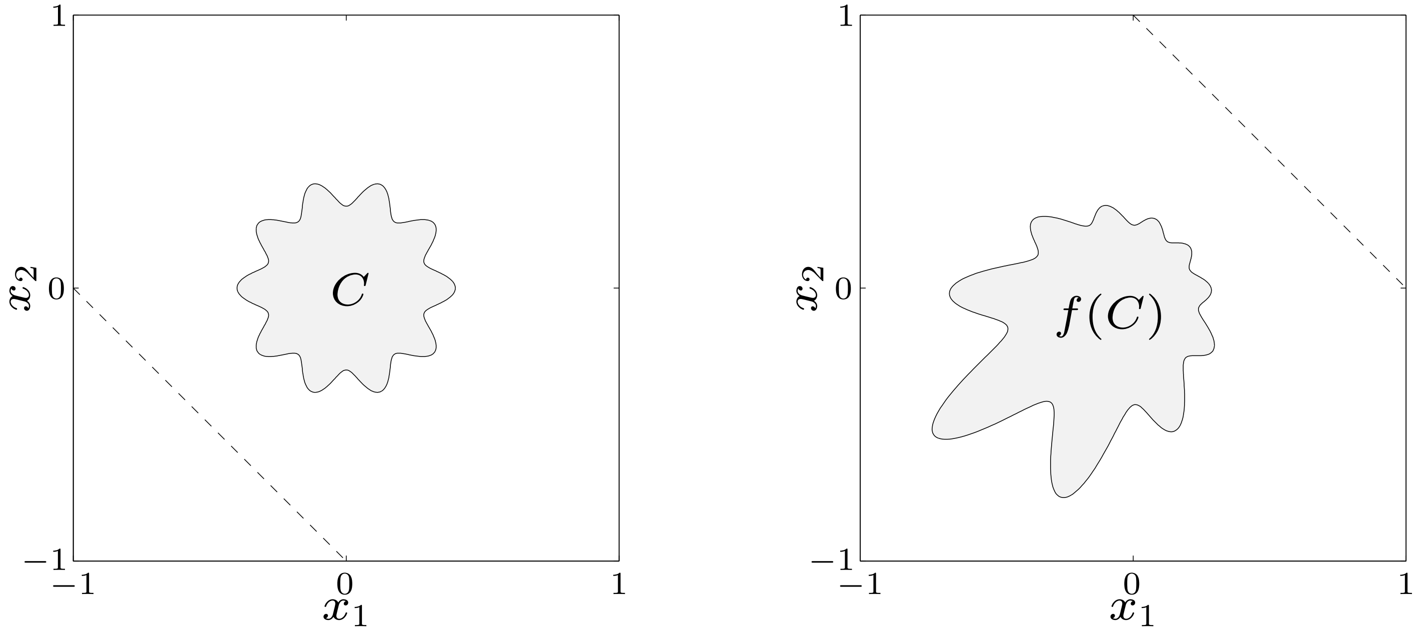
linear-fractional function $f : \mathbf{R}^n \to \mathbf{R}^m$:

$$f(x) = \frac{Ax+b}{c^T x + d}, \qquad \mathbf{dom}\, f = \{x \mid c^T x + d > 0\}$$

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

example of a linear-fractional function

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



Generalized inequalities

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 $[0, 1]$:

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

generalized inequality defined by a proper cone K:

$$x \preceq_K y \quad \Longleftrightarrow \quad y - x \in K, \qquad\qquad x \prec_K y \quad \Longleftrightarrow \quad y - x \in \mathbf{int}\, K$$

examples

- componentwise inequality ($K = \mathbf{R}^n_+$)

$$x \preceq_{\mathbf{R}^n_+} y \quad \Longleftrightarrow \quad x_i \le y_i, \quad i = 1, \ldots, n$$

- matrix inequality ($K = \mathbf{S}^n_+$)

$$X \preceq_{\mathbf{S}^n_+} Y \quad \Longleftrightarrow \quad Y - X \text{ positive semidefinite}$$

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

properties: many properties of $\preceq_K$ are similar to $\le$ on $\mathbf{R}$, *e.g.*,

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

Minimum and minimal elements

$\preceq_K$ is not in general a *linear ordering*: we can have $x \not\preceq_K y$ and $y \not\preceq_K x$

$x \in S$ is **the minimum element** of S with respect to $\preceq_K$ if

$$y \in S \quad \Longrightarrow \quad x \preceq_K y$$

$x \in S$ is **a minimal element** of S with respect to $\preceq_K$ if

$$y \in S, \quad y \preceq_K x \quad \Longrightarrow \quad y = x$$

example $(K = \mathbf{R}^2_+)$

x_1 is the minimum element of S_1
x_2 is a minimal element of S_2