Optimization in Machine Learning: Lecture 4

Convex Optimization

by Xiaolin Huang xiaolinhuang@sjtu.edu.cn SEIEE 2-429

Institute of Image Processing and Pattern Recognition

http://www.pami.sjtu.edu.cn/

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

目录 Contents

Convex Optimization

- **convex optimization** is to minimize a convex function over a convex set
  - convex combination
  - convex sets
  - operations that preserve convexity
  - separating hyperplane
  - supporting hyperplane

目录 Contents

Standard form

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

- $x \in \mathbf{R}^n$ is the optimization variable
- $f_0: \mathbf{R}^n \to \mathbf{R}$ is the object or cost or loss function
- $f_i: \mathbf{R}^n \to \mathbf{R}$ are the inequality constraint functions
- $h_i: \mathbf{R}^n \to \mathbf{R}$ are the equality constraint functions
- implicit constraint: the domain of the problem (i.e., of all the above functions)

Standard form

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{array}$$

feasibility problem

$$\begin{array}{ll} \text{find} & x \\ \text{s.t.} & f_i(x) \le 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{array}$$

$$\begin{array}{ll} \min & 0 \\ \text{s.t.} & f_i(x) \le 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{array}$$

- $x \in \mathbf{R}^n$ is the optimization variable
- $f_0: \mathbf{R}^n \to \mathbf{R}$ is the object or cost or loss function
- $f_i: \mathbf{R}^n \to \mathbf{R}$ are the inequality constraint functions
- $h_i: \mathbf{R}^n \to \mathbf{R}$ are the equality constraint functions
- implicit constraint: the domain of the problem (i.e., of all the above functions)
- **feasible** solution, **feasible** set, problem is **feasible**

Convex Optimization

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

"**convex optimization** is to minimize a convex function over a convex set"

- f_0: $\mathbf{R}^n \to \mathbf{R}$ is convex
- f_i: $\mathbf{R}^n \to \mathbf{R}$ are convex
- h_i: $\mathbf{R}^n \to \mathbf{R}$ are affine

Standard Form Convex Optimization Problem

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \dots, m \\ & a_i^\top x = b_i, i = 1, \dots, p \end{aligned}$$

"**convex optimization** is to minimize a convex function over a convex set"

- $f_0 : \mathbf{R}^n \to \mathbf{R}$ is convex
- $f_i : \mathbf{R}^n \to \mathbf{R}$ are convex
- $h_i : \mathbf{R}^n \to \mathbf{R}$ are affine

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & x_1/(1 + x_2^2) \leq 0 \\ & (x_1 + x_2)^2 = 0 \end{aligned}$$

it not a convex optimization problem
but is equivalent to a convex problem

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & x_1 \leq 0 \\ & x_1^2 + x_2^2 = 0 \end{aligned}$$

Local and Global Optima

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{aligned}$$

- a feasible solution is (globally) optimal if there is no better feasible solution

$$f_0(x^*) \le f_0(x), \forall x: f_i(x) \le 0, h_i(x) = 0$$

- a feasible problem does not necessarily have a optimal solution
  - infinite set: $\min \ a^\top x + b$
  - infinite value: $\min 1/x, \ \text{s.t.} \ 0 < x < 2$

Local and Global Optima

$$\begin{array}{ll} \min & f_0(x) \\ \mathrm{s.t.} & f_i(x) \le 0, i = 1, \dots, m \\ & h_i(x) = 0, i = 1, \dots, p \end{array}$$

- a feasible solution is (globally) optimal if there is no better feasible solution

$$f_0(x^*) \le f_0(x), \forall x: f_i(x) \le 0, h_i(x) = 0$$

- a feasible solution is locally optimal if there is no better feasible solution

  around it

$$\exists \varepsilon > 0, \hat{x} \le f_0(x), \forall x: f_i(x) \le 0, h_i(x) = 0, \|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \le \boldsymbol{\varepsilon}$$

Convex Optimization Problem

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, i = 1, \dots, m \\ & a_i^\top x = b_i, i = 1, \dots, p \end{aligned}$$

"**convex optimization** is to minimize a convex function over a convex set"

- any locally optimal point of a convex problem is (globally) optimal

Proof. $\hat{x}$ is locally optimal but not globally optimal:

- (1): $\exists x^*, f_0(x^*) < f_0(\hat{x})$
- (2): the convex combination from $\hat{x}$ to x^*: $x(\theta) = (1-\theta)\hat{x} + \theta x^*, \theta \in (0,1)$

$$f_0\big(x(\theta)\big) \le (1-\theta) f_0(\hat{x}) + \theta f(x^*) < f_0(\hat{x}), \theta \in (0,1)$$

- (3) contradicts local optimality:

$$\exists \varepsilon > 0, f_0(\hat{x}) \le f_0(x), \forall x: f_i(x) \le 0, h_i(x) = 0, \|x - \hat{x}\| \le \varepsilon$$

Equivalent Convex Problems

- two optimization problems are (informally) equivalent, if the solution of one could be **readily obtained** from the solution of the other, and vice-versa

- eliminating equality constraints:

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \dots, m \\ & a_i^\top x = b_i, i = 1, \dots, p \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \min & f_0(Fz + x_0) \\ \text{s.t.} & f_i(Fz + x_0) \le 0, i = 1, \dots, m \end{array}$$

where $Ax = b \quad \Longleftrightarrow \quad x = Fz + x_0$

- linear equality constraints in theory are easy to handle
- in practice, if F is to obtained, and x_0 could be found

could we eliminate equation constraints in this way?

- $h_i(x) \le 0, -h_i(x) \le 0$
- $a_i^\top x \le b_i, -a_i^\top x \le -b_i$

Equivalent Convex Problems

- two optimization problems are (informally) equivalent, if the solution of one could be **readily obtained** from the solution of the other, and vice-versa

- introducing equality constraints:

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0, i = 1, \dots, m \\ & a_i^\top x = b_i, i = 1, \dots, p \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \min & f_0(Fz + x_0) \\ \text{s.t.} & f_i(Fz + x_0) \le 0, i = 1, \dots, m \end{array}$$

  - it could be very useful

$$\min_x \quad \lambda\|x\|_1 + \|B - Ax\|_2^2$$

$$\Updownarrow$$

$$\begin{array}{ll} \min\limits_{x,z} & \lambda\|z\|_1 + \|B - Ax\|_2^2 \\ \text{s.t.} & x = z \end{array}$$

- g non-smooth but separable
- h smooth but non-separable

Equivalent Convex Problems

- two optimization problems are (informally) equivalent, if the solution of one could be **readily obtained** from the solution of the other, and vice-versa

- introducing slack variables for linear inequalities:

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & a_i^\top x \le b_i, i = 1, \dots, m \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & a_i^\top x + s_i = b_i, i = 1, \dots, m \\ & s_i \ge 0 \; i = 1, \dots, m \end{array}$$

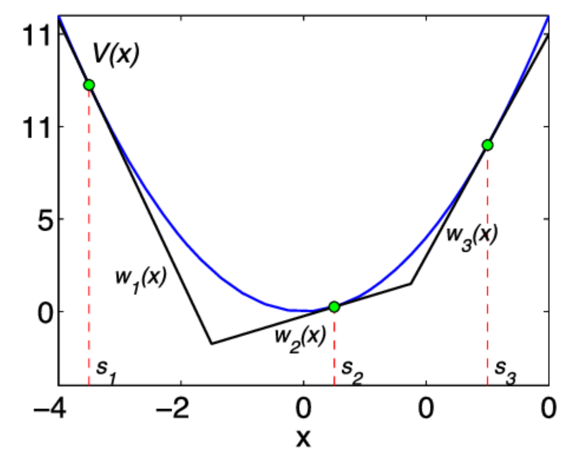
- s are separable

  - it is also very useful when meet piecewise linear functions

    flexible but non-smooth:

    l1 norm, LAD, hinge loss

    max error, (leaky) ReLu

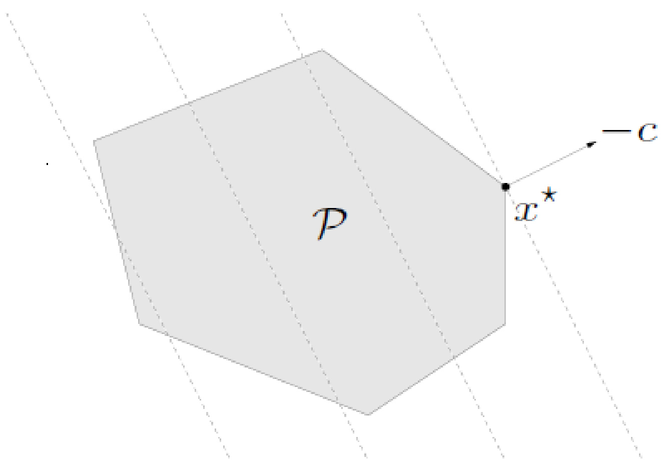


目录 Contents

Linear Programming

$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & Gx \preccurlyeq h \\ & Ax = b \end{array}$$

- minimize a linear function over a polyhedron
- convex problem with a linear objective and affine constraints



Transportation Problem

- two products I and II with different profit
- three resources A, B, and C with different inventories

	Product I	Product II	inventory
Resource A	0	5	15
Resource B	6	2	24
Resource C	1	1	5
Profit	**2**	**1**	

$$
\begin{aligned}
\max \quad & 2x_1 + x_2 \\
\text{s.t.} \quad & 5x_2 \le 15 \\
& 6x_1 + 2x_2 \le 24 \\
& x_1 + x_2 \le 5
\end{aligned}
$$

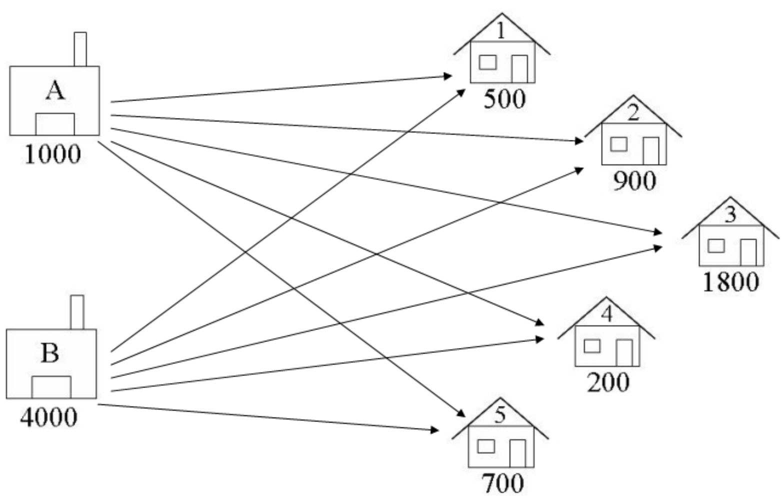
Transportation Problem

- transport goods from m suppliers with production amounts $a_1, a_2, \dots, a_m$
- n demanders/customers need production amounts $b_1, b_2, \dots, b_n$
- the unit price from the i-th supplier to the j-th customer c_{ij}
- the transportation amount from the i-th supplier to the j-th customer x_{ij}

$$\max \sum_i^m \sum_j^n c_{ij} x_{ij}$$

$$\text{s.t.} \quad \sum_{j=1}^n x_{ij} = a_i, i = 1,2, \dots m$$
$$\sum_{i=1}^m x_{ij} = b_j, j = 1,2, \dots, n$$
$$x_{ij} \geq 0, \forall i, j$$

how about non-balance?
$\sum a_i \neq \sum b_j$



Assignment Problem

- we have n workers for n jobs with cost c_{ij}
- Boolean variable $x_{ij} = \begin{cases} 1, & \text{if person } i \text{ is assinged to job } j \\ 0, & \text{otherwise} \end{cases}$

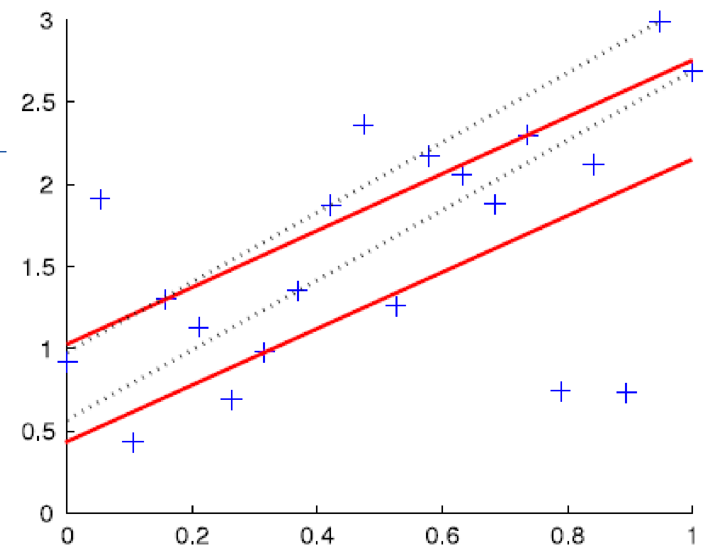
Hungarian algorithm: better than treating AP as TP

$$\min \sum_{i}^{m} \sum_{j}^{n} c_{ij} x_{ij}$$
$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} = 1, i = 1,2, \dots m$$
$$\sum_{i=1}^{m} x_{ij} = 1, j = 1,2, \dots, n$$
$$x_{ij} \in \{0,1\}, \forall i, j$$

$$\min \sum_{i}^{m} \sum_{j}^{n} c_{ij} x_{ij}$$
$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} = 1, i = 1,2, \dots m$$
$$\sum_{i=1}^{m} x_{ij} = 1, j = 1,2, \dots, n$$
$$x_{ij} \geq 0, \forall i, j$$

- relaxation is a typical way to deal with integer variable
- the gap is zero, for standard assignment problem

Linear Programming



- **largest error control**
  - establish a linear function $f(a) = x^\top a$ to approach observations $\{a_i, b_i\}_{i=1}^m$
  - the residual $r_i = b_i - x^\top a_i$
  - we want to obtain the least largest error $\max\limits_i |r_i|$

$$\min_x \max_i |b_i - x^\top a_i| \quad \Longrightarrow \quad \begin{array}{ll} \min\limits_{x,s} & s \\ \text{s.t.} & b_i - x^\top a_i \le s, i = 1,2,\dots,m \\ & -(b_i - x^\top a_i) \le s, i = 1,2,\dots,m \end{array}$$

- can you verify its convexity?
- can you image the equivalence to an LP?

Linear Programming

- **portfolio optimization**
  - there are m stocks could be investigated
  - the return for each stock is p_i, $i = 1,2, \dots, m$, a random with probability ρ
  - the ratio of investigation for each stock is x_i
  - the overall return is $p^\top x$, again a random
- maximize the estimated return, while control the largest loss

$$\max_x \int p^\top x \, d\rho(p) \quad \text{s.t.} \sup_{p\sim\rho}(-p^\top x) < C, \sum x_i = 1, x_i \geq 0$$

- if we use historical data p_{ij} to present ρ

$$\max_x \sum p_{\cdot j}^\top x \quad \text{s.t.} \max_j\left(-p_{\cdot j}^\top x\right) < C, \sum x_i = 1, x_i \geq 0$$

Linear-fractional Programming

$d = f = 0$

$$\begin{array}{ll} \min & \dfrac{c^\top x + d}{e^\top x + f} \\ \text{s.t.} & Gx \preccurlyeq h \\ & Ax = b \\ & e^\top x + f > 0 \end{array}$$

$$y = \frac{x}{e^\top x + f}$$
$$z = \frac{1}{e^\top x + f}$$

- **Linear-fractional Programming** is equivalent to the following LP

$$\begin{array}{ll} \min & c^\top y + dz \\ \text{s.t.} & Gy \preccurlyeq hz \\ & Ay = bz \\ & e^\top y + fz = 1 \\ & z \geq 0 \end{array}$$

$x = y/z$

$x, y \in \mathbf{R}^n$

$z \in \mathbf{R}$

Linear-fractional Programming

$d = f = 0$

$$\begin{aligned} \min \quad & \frac{c^\top x}{e^\top x} \\ \text{s.t.} \quad & Gx \preccurlyeq h \\ & Ax = b \\ & e^\top x > 0 \end{aligned}$$

$z = 1$

$x = y$

- **Linear-fractional Programming** is equivalent to the following LP

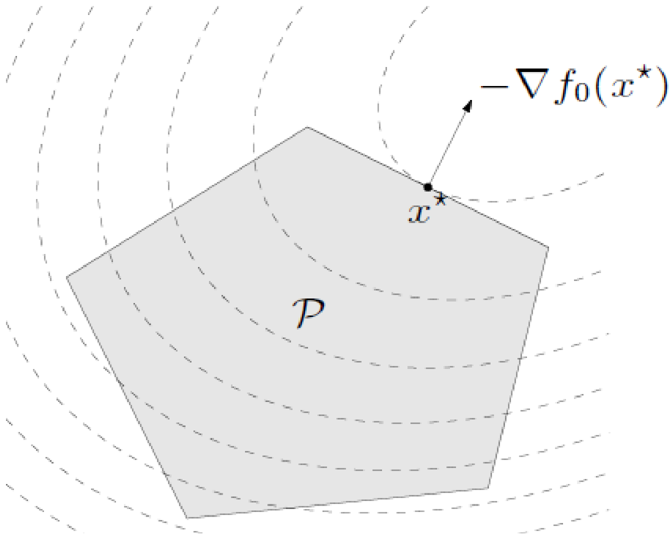
$$\begin{aligned} \min \quad & c^\top y \\ \text{s.t.} \quad & Gy \preccurlyeq h \\ & Ay = b \\ & \boldsymbol{e}^\top \boldsymbol{y} = 1 \end{aligned}$$

a useful and typical trick for **Homogeneous** object, e.g., in PCA, CCA, ICA, …, also in dual norm, steepest descent…

Quadratic Programming

$$\begin{array}{ll} \min & \dfrac{1}{2}x^\top P x + q^\top x + r \\ \text{s.t.} & Gx \preccurlyeq h \\ & Ax = b \end{array}$$

- $P \in \mathbf{S}_+^n$, the objective is convex quadratic
- minimize a convex quadratic function over a polyhedron



Quadratic Programming

- **least squares**
  - establish a linear function $f(a) = x^\top a$ to approach observations $\{a_i, b_i\}_{i=1}^m$
  - the residual $r_i = b_i - x^\top a_i$
  - we want to obtain the least sum of squared error

$$\min_x \sum (b_i - x^\top a_i)^2$$

Quadratic Programming

- **variance** is in a quadratic form
- linear programing with random cost

$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & Gx \preccurlyeq h \\ & Ax = b \end{array} \quad \xrightarrow{\text{when } c \text{ is random}} \quad \begin{array}{ll} \min & \mathbf{E}c^\top x + \gamma \mathbf{var}(c^\top x) \\ \text{s.t.} & Gx \preccurlyeq h \\ & Ax = b \end{array}$$

  - stochastic optimization
  - portfolio optimization

$$\max_x \ \bar{p}x + \gamma \sum \left(p_{\cdot,j}^\top x - \bar{p}x\right)^2 \quad \text{s.t.} \sum x_i = 1, x_i \geq 0$$

Quadratic Constraint

- we can put the variance in the constraint

$$\max_x \ \bar{p}x \ \ \text{s.t.} \ \sum\left(p_{\cdot,j}^\top x - \bar{p}x\right)^2 \le C, \sum x_i = 1, x_i \ge 0$$

- Quadratically Constrained Quadratic Programming (QCQP)

$$\begin{aligned} \min \quad & \frac{1}{2} x^\top P_0 x + q_0^\top x + r_0 \\ \text{s.t.} \quad & \frac{1}{2} x^\top P_i x + q_i^\top x + r_i \le 0, i = 1,2,\ldots,m \\ & Ax = b \end{aligned}$$

  - portfolio optimization

$$\max_x \ \bar{p}x \ \ \text{s.t.} \ \gamma\sum\left(p_{\cdot,j}^\top x - \bar{p}x\right)^2 < C, \sum x_i = 1, x_i \ge 0$$

Second-order Cone Programming

$$\begin{aligned} \min \quad & f^\top x \\ \text{s.t.} \quad & \|A_i \text{x} + b_i\|_2 \le c_i^\top x + d_i, i = 1, \dots, m \\ & Fx = g \end{aligned}$$

- **Second-order Cone Programming (SOCP)**
  - convexity: $(A_i x + b_i, c_i^\top x + d_i) \in$ second-order cone in $\mathbf{R}^{n_i+1}$
  - more general than QCQP ($c_i = 0$) and LP ($A_i = 0$)
  - when parameters are random

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & a_i^\top x \le b_i, i = 1, \dots, m \\ & a_i \in \{\overline{a_i} + P_i u, \|u\|_2 \le 1\} \end{aligned}$$

$$\sup_{a_i \in \{\overline{a_i} + P_i u, \|u\|_2 \le 1\}} a_i^\top x \le b_i$$

$$a_i^\top x + \sup_{\|u\|_2 \le 1} u^\top P_i^\top x = a_i^\top x + \left\|P_i^\top x\right\|_2$$

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & a_i^\top x + \left\|P_i^\top x\right\|_2 \le b_i, i = 1, \dots, m \end{aligned}$$

目录 Contents

Generalized inequality constraints

- convex problem with generalized inequality constraints

$$
\begin{aligned}
\min \quad & f_0(x) \\
\mathrm{s.t.} \quad & f_i(x) \preccurlyeq_{K_i} 0, i = 1, \dots, m \\
& Ax = b
\end{aligned}
$$

- f_0 is convex and f_i is convex w.r.t. proper cone K_i

$$f_i(\theta x_1 + (1-\theta)x_2) \preccurlyeq_{K_i} \theta f_i(x_1) + (1-\theta) f_i(x_2)$$

  - $K = S_+^n$: matrix convexity
    - for any z, $z^\top f(x)$z is a convex function
    - $f(X) = XX^\top$ is matrix convex

Generalized inequality constraints

- convex problem with generalized inequality constraints

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \preccurlyeq_{K_i} 0, i = 1, \dots, m \\
& Ax = b
\end{aligned}
$$

- f_0 is convex and f_i is convex w.r.t. proper cone K_i

$$
f_i(\theta x_1 + (1-\theta)x_2) \preccurlyeq_{K_i} \theta f_i(x_1) + (1-\theta) f_i(x_2)
$$

- conic form problem

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & Fx + g \preccurlyeq_K 0 \\
& Ax = b
\end{aligned}
$$

SDP and LMI

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & x_1 F_1 + x_2 F_2 + \cdots x_n F_n + G \preccurlyeq 0 \\
& Ax = b
\end{aligned}
$$

Vector optimization

- convex problem with generalized inequality defined objective

$$
\begin{array}{ll}
\min\ (\mathrm{w.r.t.}\ K) & f_0(x) \\
\mathrm{s.t.} & f_i(x) \leq 0, i = 1, \ldots, m \\
 & Ax = b
\end{array}
$$

where $f_0\colon \mathrm{R}^n \to \mathrm{R}^q$, and K is a proper cone in R^q

we could use K to measure a **partial** order for vectors

for x and y, it is not necessary that they can be compared

- $K = \mathbf{R}_+^2, x \preccurlyeq_K y$ means $(x_1 - y_1, x_2 - y_2) \in \mathbf{R}_+^2$, i.e., $x_1 \leq y_1, x_2 \leq y_2$
- $\binom{3}{2} \preccurlyeq_K \binom{4}{2}$
- $\binom{3}{2}$ $\binom{4}{1}$ neither $\preccurlyeq_K$ nor $\succcurlyeq_K$ holds

Vector optimization

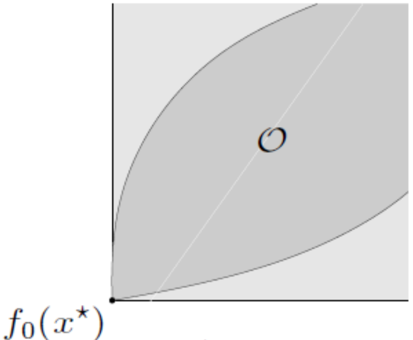
- convex problem with generalized inequality defined objective

$$
\begin{array}{ll}
\min\ (\mathrm{w.r.t.}\ K) & f_0(x) \\
\mathrm{s.t.} & f_i(x) \leq 0, i = 1, \ldots, m \\
 & Ax = b
\end{array}
$$

where $f_0 : \mathrm{R}^n \to \mathrm{R}^q$, and K is a proper cone in R^q

- minimal element:

  it is better than all the other element → optimal solution

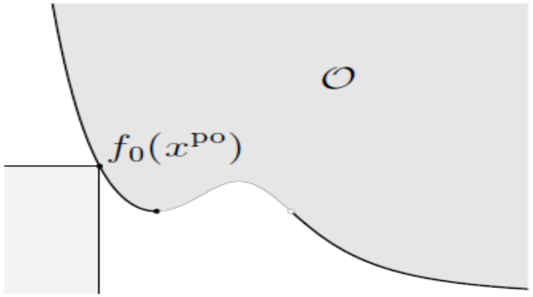


$x^\star$ is optimal

- minimum element:

  there is no element better than it

  → Pareto optimal/ efficient

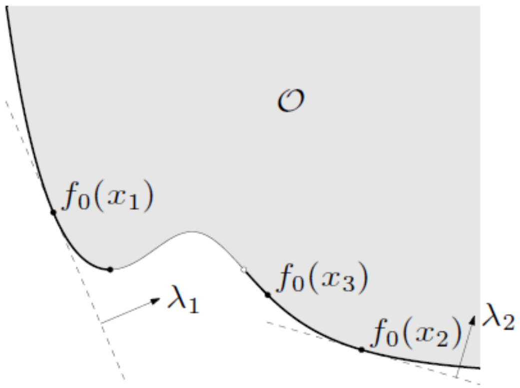


x^{po} is Pareto optimal

Scalarization

- to find the Pareto optimal,

$$\begin{array}{ll} \min\ (\mathrm{w.r.t.}\ K) & f_0(x) \\ \mathrm{s.t.} & f_i(x) \le 0, i = 1, \dots, m \\ & Ax = b \end{array}$$



- we could choose $\lambda \succ_{K^*} 0$ and solve a scalar problem

$$\begin{array}{ll} \min & \lambda^\top f_0(x) \\ \mathrm{s.t.} & f_i(x) \le 0, i = 1, \dots, m \\ & Ax = b \end{array}$$

- optimal solution for a scalar problem is Pareto-optimal
- convex problem: can find (almost) all Pareto-optimal points by scalarization

Regularized least squares

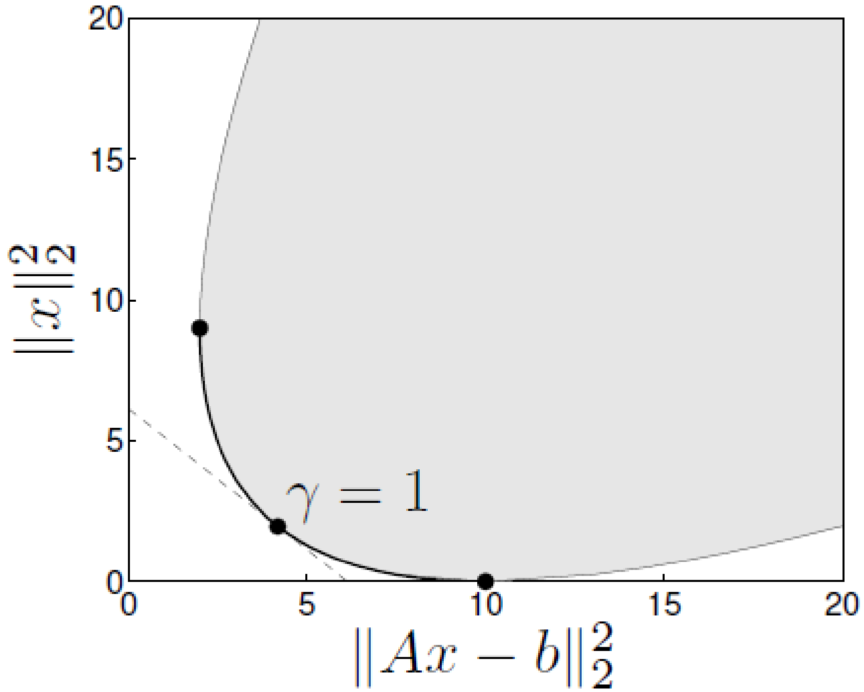
- we want to minimize both the regression loss and the model complxity

$$\min\ (\text{w.r.t.}\ \mathbf{R}_+^2)\quad (\|Ax-b\|_2^2, \|x\|_2^2)$$

- by scalarization

$$\min\ \ \|Ax-b\|_2^2 + \lambda\|x\|_2^2$$

$$\lambda > 0$$



目录 Contents

Conclusion and Home Work

- **convex optimization** is to minimize a **convex function** over a **convex set**
  - **standard formulation and equivalent transformation**
  - **local and global optimum**
  - **LP, QP, QCQP, SOCP**
  - **vector optimization and Pareto optimum**

- **Excise 4.9:** LP and its solution

  Excise 4.15: integer programming and its relaxation

  Excise 4.22: QCQP and its solution

THANKS