01. Introduction to Convex Optimization –最优化导论

By Yang Lin[1] (2023 秋季, @NJU)

Contents

0 About this course

1 Elements for mathematical optimization

2 Examples of optimization problems

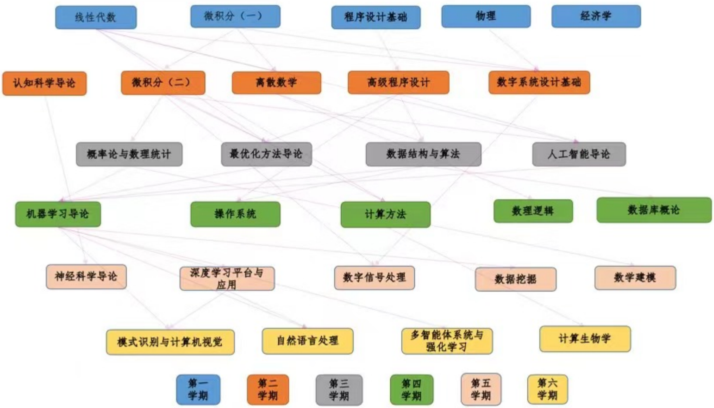
2.1 Least Squares
2.2 Linear Programming
2.3 Convex Optimization

3 Examples of applications

4 Brief history of convex optimization

[1] Institute: Nanjing University. Email: linyang@nju.edu.cn.

0 About this course



0 About this course

1 助教

李紫微、张浩然、陈浩旭、（TBD）

2 课程时间

周一 3-4 节（1-17 周）；周四 3-4 节（双数周）

3 课程考核

- 期末考试（闭卷）60%
- 作业（两周一次）、课堂或期中考核 30%
- 课程项目（TBD，文献阅读等）10%

4 参考书

"Convex Optimization" Stephen Boyd and Lieven Vandenberghe

1 Mathematical Optimization

"Optimization" as a tool in engineering:

"在给定的决策集中（通过某种方式）寻找最优解"

$$\begin{array}{ll} \texttt{minimize} & f_0(x) \\ \texttt{subject to} & f_i(x) \le b_i, \quad i = 1, 2, \ldots, m \end{array}$$

- □ $x = (x_1, x_2, \ldots, x_n)$: optimization variables
- □ $f_0 : \boldsymbol{R}^n \to \boldsymbol{R}$: Objective function (objective)
- □ $f_i : \boldsymbol{R}^n \to \boldsymbol{R}, \ i = 1, \ldots, m$: Constraint functions (constraints)

Optimal solution x^* has smallest value of f_0 among all vectors that satisfy the constraints

Preliminaries: norms

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

$$\begin{array}{ll} (1) & \|x\| \geq 0;\ \|x\| = 0 \text{ if and only if } x = 0 \\ (2) & \|tx\| = t\|x\|, t \in \boldsymbol{R} \\ (3) & \|x+y\| \leq \|x\| + \|y\| \text{ (triangle inequality)} \end{array}$$

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

Distance: between vectors x and y as the length of their difference, *i.e.*,

$$\mathtt{dist}(x, y) = \|x - y\|$$

Preliminaries: norms

Some common norms on $\boldsymbol{R}^n$:

1. Sum-absolute-value, or l_1-norm:

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|, \;\; x \in \boldsymbol{R}^n$$

2. l_2-norm

$$\|x\|_2 = (x_1^2 + x_2^2 + \cdots + x_n^2)^{\frac{1}{2}}, \;\; x \in \boldsymbol{R}^n$$

3. l_p-norm:

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{\frac{1}{p}}, \;\; x \in \boldsymbol{R}^n$$

4. l_∞-norm (Chebyshev norm):

$$\|x\|_\infty = \lim_{p \to \infty} \|x\|_p = \max\{|x_1|, \ldots, |x_n|\}$$

5. For $P \in \boldsymbol{S}_{++}^n$, P-quadratic norm is

$$\|x\|_P = (x^T P x)^{\frac{1}{2}} = \|P^{\frac{1}{2}} x\|_2$$

Preliminaries: norms

proof.

$$\begin{aligned} \|x\|_p =& (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{\frac{1}{p}} \geq |x_{\max}| \\ \|x\|_p \leq& (n \cdot |x_{\max}|^p)^{\frac{1}{p}} = n^{\frac{1}{p}} |x_{\max}| \end{aligned}$$

$$n^{\frac{1}{p}} |x_{\max}| \rightarrow |x_{\max}| \quad \text{when} \quad p \rightarrow \infty$$

$$\ldots$$

Preliminaries: norms

Equivalence between norms:

1. Suppose that $\|\cdot\|_a$ and $\|\cdot\|_b$ are norms on $\boldsymbol{R}^n$, there exist positive constants α and β, for all $x \in \boldsymbol{R}^n$:

$$\alpha\|x\|_a \leq \|x\|_b \leq \beta\|x\|_a$$

2. If $\|\cdot\|$ is any norm on $\boldsymbol{R}^n$, then there exists a quadratic norm $\|\cdot\|_P$ for which

$$\|x\|_P \leq \|x\| \leq \sqrt{n}\|x\|_P$$

holds for all x

Preliminaries: norms

proof. of first part:

It is sufficient to prove $\alpha\|x\|_1 \leq \|x\|_b \leq \beta\|x\|_1$

Or $\alpha \leq \|u\|_b \leq \beta$, where $u = x/\|x\|_1$ has norm $\|u\|_1 = 1$

Let

$$\alpha = \min_{\|u\|_1=1} \|u\|_b$$

$$\beta = \max_{\|u\|_1=1} \|u\|_b$$

α, β exist for given b

Proof completes

2 Examples: least squares

Least squares formulation:

$$\texttt{minimize} \quad f_0(x) = \|Ax - b\|_2^2 = \sum_{i=1}^{k} (a_i^T x - b_i)^2$$

- $A \in \boldsymbol{R}^{k\times n} (k \geq n)$, a_i^T 是矩阵 A 的行向量
- Solving the above problem is equivalent to solving $(A^TA)x = A^Tb$. Hence, $x = (A^TA)^{-1}A^Tb$
- Computation time: proportional to n^2k, less if structured

Properties of Least-squares problems:

- Easy to recognize and standard techniques
- Increase flexibility through weights, regularization terms

2 Examples: linear programming

Linear programming formulation:

$$\begin{array}{ll} \texttt{minimize} & f_0(x) = c^T x \\ \texttt{subject to} & a_i^T x \le b_i, \ \ i = 1, \ldots, m. \end{array}$$

- 向量 $c, a_1, \ldots, a_m \in \boldsymbol{R}^n$, $b_1, \ldots, b_m \in \boldsymbol{R}$ 为问题参数（parameter）
- No analytical solution (closed-form equation)
- Computation time: proportional to $n^2 m$ if $m \ge n$, less if structured

Properties of Least-squares problems:

- not as easy to recognize as least-squares problems
- A few standard tricks used to convert problems into linear programs (piecewise-linear functions)

2 Examples: convex optimization

Convex optimization formulation:

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

- Object and constraint functions are **both** convex

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

  if $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$
- Includes least-squares problems and linear programming as special cases

2 Examples: solving CO problems

- No analytical solution
- Reliable and efficient algorithms
- Computation time(roughly) proportional to $\max\{n^3, n^2 m, F\}$, where F is cost of evaluating f_i' s and their first and second derivatives

When using convex optimization (why is it popular?)

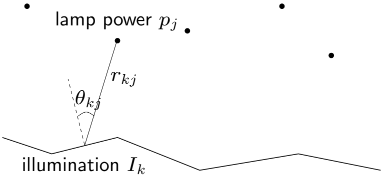
- Often difficult to recognize
- Many tricks for transforming problems into convex form
- Surprisingly many problems can be solved via convex optimization

"In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."

——*Rockafellar, 1993*

2 Examples: lamps illuminating patches

m lamps illuminating n (small, flat) patches



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

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

- Achieve desired illumination I_{des} with bounded lamp powers

$$\begin{array}{ll} \texttt{minimize} & \max_k |\log I_k - \log I_{\mathsf{des}}| \\ \texttt{subject to} & 0 \le p_j \le p_{\max}, \; j = 1, \ldots, m. \end{array}$$

2 Examples: lamps illuminating patches

How to solve?

- Use uniform power $p_j = p$, vary p
- Use least-squares

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

  round p_j if $p_j < 0$ or $p_j > p_{\max}$
- Use weighted least-squares

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

  iteratively adjust weights w_j until $0 \le p_j \le p_{\max}$
- Use linear programming

$$\begin{array}{ll} \texttt{minimize} & \max_{k=1,2,\ldots,n} |I_k - I_{\texttt{des}}| \\ \texttt{subject to} & 0 \le p_j \le p_{\max},\, j = 1, \ldots, m \end{array}$$

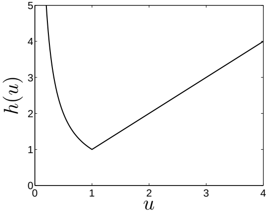
2 Examples: lamps illuminating patches

How to solve?

- Use convex optimization: problem is equivalent to

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

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



f_0 is convex because maximum of convex functions is convex

3 Examples in applications

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

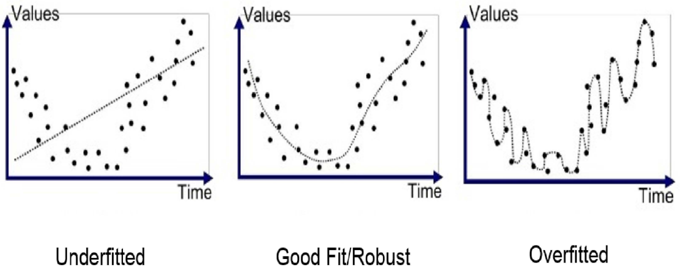
- Variables: device widths and lengths
- Constraints: manufacturing limits, timing requirements, maximum area
- Objective: power consumption

Data fitting

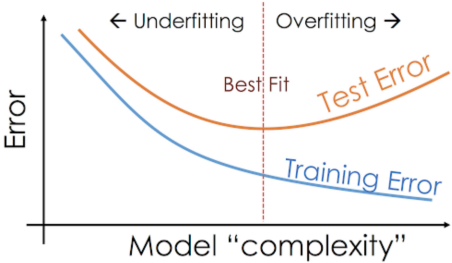
- Variables: model parameters
- Constraints: prior information, parameter limits
- Objective: measure of misfit or prediction error

3 Examples in applications

Overfitting/Underfitting



3 Examples in applications



Overfitting/Underfitting with different levels of complexity/training

4 Brief history of convex optimization

Theory(convex analysis): ca1900–1970

Algorithms

- 1947: simplex algorithm for linear programming (Dantzig)
- 1960s: early interior-point methods (Fiacco and McCormick, Dikin, ...)
- 1970s: ellipsoid method and other subgradient methods
- 1980s: polynomial-time interior-point methods for linear programming (Karmarkar1984)
- late 1980s–now: polynomial-time interior-point methods for nonlinear convex optimization (Nesterov and Nemirovski1994)

Applications

- Before 1990: mostly in operations research; few in engineering
- Since 1990: many new applications in engineering (control, signal processing, communications, circuit design, ...); new problem classes (semidefinite and second-order cone programming, robust optimization)