Convex Optimization

Lecture 7: Convex Problems

Lecturer: *Dr*. Wan-Lei Zhao

Autumn Semester 2025

Contact: wlzhao@xmu.edu.cn

Outline

Optimization Problem (1)

$$
\begin{aligned}
\text{Max} \quad & f_0(x) \qquad x \in R^n \\
\text{s. t.} \quad & f_i(x) \leq 0 \qquad i = 1 \cdots m \\
& h_j(x) = 0 \qquad j = 1 \cdots p
\end{aligned}
\tag{1}
$$

- In the above equation, $f_0 : R^n \to R$ is called objective function or loss function
- $f_i(x) \leq 0, i = 1 \cdots m$ are the inequation constraints
- $h_j(x) = 0, j = 1 \cdots p$ are the equation constraints
- The domf X_f is the intersection of all domfs of these functions

$$X_f = \cap_{i=0}^{m} dom\ f_i \cap_{j=0}^{p} dom\ h_i$$

- X_f is called feasible solution set

Optimization Problem (2)

$$\begin{aligned} \text{Max} \quad & f_0(x) \qquad x \in R^n \\ \text{s. t.} \quad & f_i(x) \leq 0 \qquad i = 1 \cdots m \\ & h_j(x) = 0 \qquad j = 1 \cdots p \end{aligned} \tag{8}$$

$$X_f = \cap_{i=0}^{m} dom\ f_i \cap_{j=0}^{p} dom\ h_i$$

- Among all the feasible solution X_f, $p^* = \inf\{f_0(x)|x \in X_f\}$
  1. if $X_f = \emptyset$, there is no solution
  2. Sometimes, we can only find out a sub-optimal solution
  $$X_\epsilon = \{x|x \in X_f, f_0(x) = p^* + \epsilon\}.$$

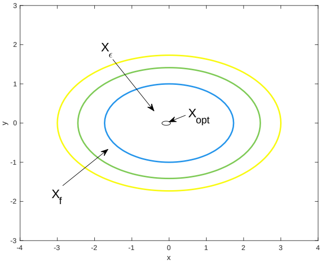
Optimization Problem (3)

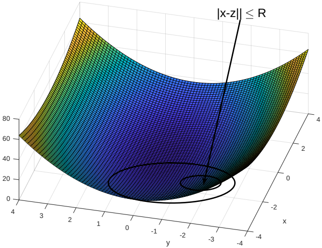
$$\begin{aligned} \text{Max} \quad & f_0(x) \qquad x \in R^n \\ \text{s. t.} \quad & f_i(x) \leq 0 \qquad i = 1 \cdots m \\ & h_j(x) = 0 \qquad j = 1 \cdots p \end{aligned} \tag{8}$$

$$X_f = \cap_{i=0}^{m} dom\ f_i \cap_{j=0}^{p} dom\ h_i$$

- If the following hold
  1. f_0 is convex
  2. If $f_i(x) \leq 0, i = 1 \cdots m$ are convex
  3. $h_j(x) = 0, j = 1 \cdots p$ are convex
  4. The domf X_f is convex
- The optimization problem is called convex problem

Optimization Problem (4)





$$X_f = \cap_{i=0}^{m} dom\ f_i \cap_{j=0}^{p} dom\ h_i$$

- Among all the feasible solution X_f, $p^* = \inf\{f_0(x)|x \in X_f\}$
  1. if $X_f = \emptyset$, there is no solution
  2. Sometimes, we can only find out a sub-optimal solution

$$X_\epsilon = \{x|x \in X_f, f_0(x) = p^* + \epsilon\}.$$

Optimization Problem: Local Optima

- In the neighboring area of x, $\|x - z\| \leq R$, if $f_0(x)$ is the minimum, $f_0(x)$ is called local optimal

$$f_0(x) = \inf_z \left\{ f_0(x) \;\middle|\; \begin{array}{ll} f_i(z) \leq 0 & i = 1 \cdots m \\ h_j(z) = 0 & j = 1 \cdots p \\ \|x - z\| \leq R & \end{array} \right\} \tag{9}$$

- The above expression searches over all feasible solution, and take the minimum

Optimization Problem: Global Optima (1)

- **Theorem:** if an optimization problem $\mathcal{P}$ is convex, the local optimal solution for the problem is also the global optima of the problem
- Namely, $\exists R > 0$, the neighboring area of x $f_0(x)$ is the local optima, then we have

$$f_0(x) = \inf \left\{ f_0(x) \middle| \begin{array}{c} x, z \in X_f \\ \|x - z\| \leq R \end{array} \right\} \tag{9}$$

- $f_0(x)$ is the global optima of the problem

Optimization Problem: Global Optima (2)

- **Theorem:** if an optimization problem $\mathcal{P}$ is convex, the local optimal solution for the problem is also the global optima of the problem
- **Proof by contradicting:**
  1. Assuming $f_0(x)$ is is not optima, and $\exists y \in X_f, f_0(y) < f_0(x)$
  2. Since x is the local optima, y is outside R neighborhood
  3. Since X_f is convex, we have $z = (1-\theta)x + \theta y$
  4. In particular, $\theta = \frac{R}{2\|y-x\|_2}$
  5. Since $f_0(x)$ is convex, we have

$$f_0(z) \leq (1-\theta)f_0(x) + \theta f_0(y)$$

- Based on $\theta = \frac{R}{2\|z-x\|_2}$, we have $\|z-x\|_2 = \frac{R}{2}$
- It is clear to see z is in the R-neighborhood of x, so we have $f_0(x) \leq f_0(z)$

Optimization Problem: Global Optima (3)

1. Assuming $f_0(x)$ is not optimal, and $\exists y \in X_f, f_0(y) < f_0(x)$
2. Since x is the local optimal, y is outside R neighborhood
3. Since X_f is convex, we have $z = (1-\theta)x + \theta y$, $\theta = \frac{R}{2\|y-x\|_2}$
4. Since $f_0(x)$ is convex, we have

$$f_0(z) \leq (1-\theta)f_0(x) + \theta f_0(y)$$

- Based on $\theta = \frac{R}{2\|z-x\|_2}$, we have $\|z-x\|_2 = \frac{R}{2}$
- It is clear to see z is in the R-neighborhood of x, namely $f_0(x) \leq f_0(z)$, so we have

$$f_0(y) < f_0(x) \leq f_0(z)$$
$$\Longrightarrow f_0(y)\theta + f_0(x)(1-\theta) < f_0(z)$$

- So we have

$$f_0(z) \leq (1-\theta)f_0(x) + \theta f_0(y) < f_0(z).$$

Optimization Problem: Linear Programming

- Given an optimization problem $\mathcal{P}$
- If both the objective function and the constraints are linear
- Problem $\mathcal{P}$ is called linear programming (LP) problem

$$Max.\ \ c^T x + d$$

$$s.t. \begin{cases} Gx + s = h \\ Ax = b \end{cases}$$

- where $s \succeq 0$ are slack variables
- Moreover, let $x = x^+ - x^-$, $x^+ \succeq 0,\ x^- \succeq 0$

$$Max.\ \ c^T x^+ - C^T x^- + d$$

$$s.t. \begin{cases} Gx^+ - Gx^- + s = h \\ Ax^+ - Ax^- = b \end{cases}$$

$$\Rightarrow$$

$$Max.\ \ c^T x + d$$

$$s.t. \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

Optimization Problem: Quadratic Programming

- Given an optimization problem $\mathcal{P}$
- If the objective function is quadratic and the constraints are linear
- Problem $\mathcal{P}$ is called quadratic programming (QP) problem

$$
\begin{aligned}
&Max. \;\; \frac{1}{2}x^T Px + g^T x + r \\
&s.t. \begin{cases} Qx \leq h, \quad P \in S_+^n \\ Ax = b \end{cases}
\end{aligned}
\tag{9}
$$

Optimization Problem: QC Quadratic Programming

- Given an optimization problem $\mathcal{P}$
- If both the objective function and the constraints are quadratic
- Problem $\mathcal{P}$ is called quadratically constrained quadratic programming (QCQP) problem

$$
\begin{aligned}
Max \ \ \frac{1}{2}x^T Px + g^T x + r, \quad P \in S^n_+ \\
s.t. \begin{cases} \frac{1}{2}xP_i x + q_i^T x + r_i \leq 0, \quad P_i \in S^n_+, i = 1 \cdots m \\ Ax = b \end{cases}
\end{aligned}
\tag{10}
$$

QC Quadratic Programming: Example-1

- Given measurement system $b = Ax + e$
- where b is the observation, A is linear transformation, and e is the noise
- Both b and A are known, we are required to recover the original signal x

$$\begin{aligned}
\hat{x} &= \underset{x}{\operatorname{argmin}} \|b - Ax\|_2 \\
&= \underset{x}{\operatorname{argmin}} \|b - Ax\|_2^2 \\
&= \underset{x}{\operatorname{argmin}}\, x^T A^T A x - 2b^T A x + b^T b \\
&= (A^T A)^{-1} A^T b
\end{aligned} \tag{11}$$

QC Quadratic Programming: Example-2

- Given $(x^{(1)}, y^{(1)}), \ldots, (x^{(i)}, y^{(i)}), \ldots, (x^{(m)}, y^{(m)})$ are independent
- $y^{(i)}$ and $x^{(i)}$ are related by following equation:

$$y^{(i)} = \theta^T \begin{bmatrix} x^{(i)} \\ 1 \end{bmatrix} + \varepsilon^{(i)}, \tag{12}$$

where $\varepsilon^{(i)}$is the error

- Given $X = \begin{bmatrix} x^{(1)}, \cdots, x^{(i)}, \cdots, x^{(m)} \\ 1 \quad , \cdots, \quad 1 \quad , \cdots, \quad 1 \end{bmatrix}$, $x^{(i)} \in R^d$
- The least square problem can be written as

$$\text{Min.} \sum_{i=1}^{m} \| y^{(i)} - \theta^T \begin{bmatrix} x^{(i)} \\ 1 \end{bmatrix} \|_2^2 . \tag{13}$$

- The analytic solution is $\theta = (X^T X)^{-1} X^T \overrightarrow{y}$

QC Quadratic Programming: ridge regression

- If we require the solution to be sparse, namely $\| \theta \|_0$ to be small
- We have

$$\text{Min.} \sum_{i=1}^{m} \| y^{(i)} - \theta^T \begin{bmatrix} x^{(i)} \\ 1 \end{bmatrix} \|_2^2 + \| \theta \|_0 . \tag{14}$$

- However, this problem is convex, so we loose it as

$$\text{Min.} \sum_{i=1}^{m} \| y^{(i)} - \theta^T \begin{bmatrix} x^{(i)} \\ 1 \end{bmatrix} \|_2^2 + \| \theta \|_2 . \tag{15}$$

$$\Downarrow$$

$$\begin{aligned} \text{Min.} \quad & \sum_{i=1}^{m} \| y^{(i)} - \theta^T \begin{bmatrix} x^{(i)} \\ 1 \end{bmatrix} \|_2^2 \\ \text{s.t.} \quad & \|\theta\|_2^2 \leq \tau \end{aligned} \tag{16}$$

QC Quadratic Programming: lasso regression

- This is called **Ridge Regression**

$$\begin{aligned} \text{Min.} \quad & \sum_{i=1}^{m} \| \, y^{(i)} - \theta^T \left[\begin{array}{c} x^{(i)} \\ 1 \end{array} \right] \|_2^2 \\ \text{s.t.} \quad & \|\theta\|_2^2 \leq \tau \end{aligned} \tag{17}$$

- If we change the regularization term to l_1, it is called **LASSO Regression**

$$\begin{aligned} \text{Min.} \quad & \sum_{i=1}^{m} \| \, y^{(i)} - \theta^T \left[\begin{array}{c} x^{(i)} \\ 1 \end{array} \right] \|_2^2 \\ \text{s.t.} \quad & \|\theta\|_1^1 \leq \tau \end{aligned} \tag{18}$$

The least absolute shrinkage and selection operator (Lasso).

Portifolio Problem (1)

- Given B amount of money
- We want to invest on $p_1, \cdots, p_n$ projects
- The profit from the projects are given by $p^T = [1.05, 1.05, \cdots, 1]$
- How to invest to maximize the profit??
- Intuitively, we should invest all our money to the project makes the highest profit
- However, based on our experience, it is very risky?
- Anything we missed??

Portifolio Problem (2)

- In 1952, H. Markowitz proposed a classic portifolio model
- The profit of the investment on n projects is given by $p^T = [1.05, 1.05, \cdots, 1]$
- The risk of each project is given by

$$\Sigma = \begin{bmatrix} 1.2 & 0 & 0 & \cdots \\ 0 & 2 & 0 & \cdots \\ 0 & 0 & 1.1 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

- The portifolio problem can be modeled as

$$\begin{aligned} Min \quad & x^T \Sigma x \\ s.t. \quad & p^T x \geq r_{min} \\ & \mathbf{1}^T x = B \\ & x \geq 0 \end{aligned} \tag{19}$$

Figure: Harry Markowitz (1927 - 2023)

- Nobel prize winner in 1990

Optimization on Multiple Objectives (1)

- In practice, we may want to optimize on several objectives

$$\begin{aligned} Min \quad & f(x) \\ Max \quad & g(x) \\ & \cdots\cdots \\ s.t. \quad & f_i(x) \leq 0, \quad i = 1 \cdots m \\ & h_j(x) = 0, \quad j = 1 \cdots p \\ & x \in R^n \end{aligned} \tag{20}$$

- Recall the ridge regression, we minimize
  1. The error between real signal and the recovered signal
  2. The variance between recovered signals

$$\begin{aligned} Min \quad & \|b - Ax\|_2^2 \\ Min \quad & \|x\|_2 \end{aligned}$$

Optimization on Multiple Objectives (2)

- To convert optimization problem from a multiple objectives to a single
- Weights are associated with different objectives

$$
\begin{aligned}
Min. &\quad \|b - Ax\|_2^2 \\
Min. &\quad \|x\|_2
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
Min &\quad \|b - Ax\|_2^2 + \lambda \|x\|_2^2 \\
s.t. &\quad \|x\|_2^2 \leq \varepsilon
\end{aligned}
\tag{21}
$$

Optimization on Multiple Objectives (3)

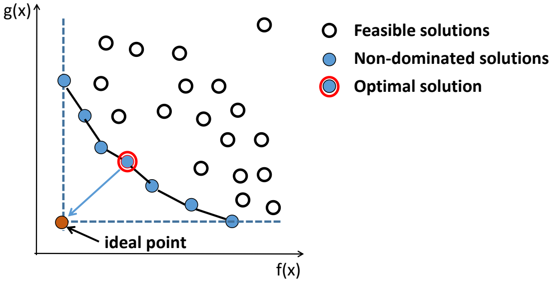


Figure: Pareto front

$$
\begin{aligned}
&Min \quad f(x) \\
&Min \quad g(x) \\
&s.t. \quad f_i(x) \leq 0 \quad i = 1 \cdots m \\
&\qquad\;\; h_j(x) = 0 \quad j = 1 \cdots p
\end{aligned} \tag{22}
$$