THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
MATH4230 2024-25
Lecture 18
March 25, 2025 (Tuesday)

1 Recall: Convex Optimization

$$P := \inf_{x \in K} f(x) \quad \text{subject to} \quad \begin{array}{ll} g_i(x) \le 0, & i = 1, \dots, m \\ h_j(x) = 0, & j = 1, \dots, \ell \end{array} \qquad (P)$$

where $f, g_i : \mathbb{R}^n \to \mathbb{R}$ are convex differentiable functions while $h_j : \mathbb{R}^n \to \mathbb{R}$ are affine functions, i.e. $h_j(x) = A_j^T x + b_j$.

and the **dual problem**

$$D = \sup_{\substack{\lambda \in \mathbb{R}^m_+ \\ \mu \in \mathbb{R}^\ell}} d(\lambda, \mu), \qquad \text{where } d(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu)$$

where $L(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^\ell \mu_j h_j(x)$ and the feasible set

$$K = \{x \in \mathbb{R}^n : g_i(x) \le 0,\ h_j(x) = 0\}$$

KKT Theorem

Proposition 1. *Assume that there exists $\bar{x}$ such that Slater condition holds, i.e.*

$$\begin{cases} g_i(\bar{x}) \le 0, & i = 1, \dots, m \\ h_j(\bar{x}) = 0, & j = 1, \dots, \ell \end{cases} \qquad \textit{and} \quad \{A_j : j = 1, \dots, \ell\} \textit{ is linearly independent}$$

*Then **Mangasarian-Fromovitz Qualification condition** holds for all*

$$x \in K = \{x \in \mathbb{R}^n : g_i(x) \le 0,\ h_j(x) = 0\}.$$

Corollary 2. *Let $x^* \in K$ be an optimal solution to P and it satisfies the qualification condition (Mangasarian-Fromovitz or Abadie). Then there exists $\lambda^* \in \mathbb{R}^m_+,\ \mu^* \in \mathbb{R}^\ell$ such that*

$$\begin{cases} \sum \lambda_i^* g_i(x^*) = 0 \\ \nabla_x L(x^*, \lambda^*, \mu^*) = 0 \end{cases}$$

Theorem 3. *Assume that the constraint K satisfies the qualification condition, and $x^* \in K$. Then x^* is an optimal solution to P iff (*) there exists $(\lambda^*, \mu^*) \in \mathbb{R}^m_+ \times \mathbb{R}^\ell$ such that*

$$\begin{cases} \sum \lambda_i^* g_i(x^*) = 0 \\ \nabla_x L(x^*, \lambda^*, \mu^*) = 0 \end{cases}$$

Remarks. It is the **necessary and sufficient condition** for **convex optimization**. In general, this is a necessary condition only.

Proof.

1. We already know that $(*)$ is necessary condition for the optimality of x^* in the previous lectures.
2. Assume that $(*)$ holds. Since $x \mapsto L(x, \lambda, \mu)$ is convex, then

$$L(x^*, \lambda^*, \mu^*) = \min_{x \in \mathbb{R}^n} L(x, \lambda^*, \mu^*) = d(\lambda^*, \mu^*).$$

This is equivalent to for all $y \in \mathbb{R}^n$, we have

$$L(y, \lambda^*, \mu^*) \geq L(x^*, \lambda^*, \mu^*) = f(x^*) + \underbrace{\sum \lambda_i^* g_i(x^*)}_{=0 \text{ (By (*))}} + \underbrace{\sum \mu_j h_j(x^*)}_{=0} = f(x^*)$$

Moreover, for all $y \in K$,

$$L(y, \lambda^*, \mu^*) = f(y) + \sum \lambda_i^* g_i(y) \leq f(y)$$

Together with the above inequalities, we have

$$f(x^*) \leq L(y, \lambda^*, \mu^*) \leq f(y), \quad \forall y \in K$$

thus this implies that x^* is an optimal solution to (P).

□

2 Duality by KKT Theorem

Theorem 4. *Assume that K satisfies the qualification condition, and that (P) has at least one solution. Then*

(i) $P = D$, and

(ii) there exists $(\lambda^*, \mu^*) \in \mathbb{R}_+^m \times \mathbb{R}^\ell$ such that $D = d(\lambda^*, \mu^*)$ and there exists x^* such that $L(x^*, \lambda^*, \mu^*) = d(\lambda^*, \mu^*)$ and x^* is solution to (P).

Proof.

1. First, we have the **Weak Duality:** $P \geq D$ is always true.
2. Since (P) has at least one solution. Let x^* be a solution to (P), then by KKT theorem, there exists (λ^*, μ^*) such that

$$\begin{cases} \sum \lambda_i^* g_i(x^*) = 0 \\ \nabla_x L(x^*, \lambda^*, \mu^*) = 0 \end{cases}$$

By definition, we have $D = \sup\limits_{\substack{\lambda \in \mathbb{R}_+^m \\ \mu \in \mathbb{R}^\ell}} d(\lambda, \mu) \geq d(\lambda^*, \mu^*) = \inf\limits_{x \in \mathbb{R}^n} L(x, \lambda^*, \mu^*)$.

From the first order condition $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$, we have $\inf\limits_{x \in \mathbb{R}^n} L(x, \lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*)$.

Putting all together, we have

$$\begin{aligned} D = \sup_{\substack{\lambda \in \mathbb{R}_+^m \\ \mu \in \mathbb{R}^\ell}} d(\lambda, \mu) &\geq d(\lambda^*, \mu^*) = \inf_{x \in \mathbb{R}^n} L(x, \lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*) \\ &= f(x^*) + \underbrace{\sum \lambda_i^* g_i(x^*)}_{=0} + \underbrace{\sum \mu_j h_j(x^*)}_{=0} \\ &= \inf_{x \in K} f(x) = P \end{aligned}$$

Therefore, we have $D = d(\lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*) = P$.

□

Remarks. From $D = d(\lambda^*, \mu^*)$, this implies that (λ^*, μ^*) is the dual optimizer.
From $d(\lambda^*, \mu^*) = L(x^*, \lambda^*, \mu^*)$, this implies that x^* is optimizer in $d(\lambda^*, \mu^*)$ and x^* is optimizer for (P).

3 Exercises

Exercise 1. Solve the optimization problem

$$\min_{\substack{-\log(x)-y\le 0\\ y\ge 1}} x + \frac{1}{2}y^2.$$

Solution. Let $f(x,y) = x + \frac{1}{2}y^2$. As there are two inequality constraints, so $g_1(x,y) = -\log(x) - y$ and $g_2(x,y) = 1-y$. Clearly, f is convex function. Since $\log(x)$ is a concave function, so $-\log(x)$ is convex and hence $g_1(x,y), g_2(x,y)$ are convex functions. Consider the feasible set K as

$$K := \{(x,y) : y \ge 1,\ -\log(x) \le y\} = \left\{(x,y) : y \ge 1,\ x \ge e^{-y}\right\}$$

Since f, g_1, g_2 are convex and $f(\cdot)$ is coercive function, so there exists minimizer $x^* \in K$ as a solution to the problem. Define the **Lagrangian** function as

$$L(x,y,\lambda_1,\lambda_2) = x + \frac{1}{2}y^2 - \lambda_1\left(\log(x) + y\right) + \lambda_2\left(1-y\right)$$

and the **dual** function $d(\lambda_1,\lambda_2) = \inf_{(x,y)} L(x,y,\lambda_1,\lambda_2)$. Now, we compute

$$\nabla_{(x,y)} L(x,y,\lambda_1,\lambda_2) = \begin{pmatrix} 1 - \dfrac{\lambda_1}{x} \\ y - \lambda_1 - \lambda_2 \end{pmatrix}$$

By the Euler's first order condition, by setting $\nabla_{(x,y)} = \mathbf{0}$, we have $(x^*, y^*) = (\lambda_1, \lambda_1 + \lambda_2)$.
So, we have

$$\inf_{(x,y)} L(x,y,\lambda_1,\lambda_2) = L(\lambda_1, \lambda_1+\lambda_2, \lambda_1, \lambda_2) = \lambda_1 - \lambda_1\log(\lambda_1) - \frac{1}{2}(\lambda_1+\lambda_2)^2 + \lambda_2$$

Now, we compute

$$\nabla_\lambda d(\lambda) = \begin{pmatrix} 1 - \log(\lambda_1) - 1 - (\lambda_1+\lambda_2) \\ -(\lambda_1+\lambda_2) + 1 \end{pmatrix}$$

By setting equals to $\mathbf{0}$, we have $\begin{cases} \log(\lambda_1^*) + \lambda_1^* + \lambda_2^* = 0 \\ \lambda_1^* + \lambda_2^* = 1 \end{cases} \implies \begin{cases} \lambda_1^* = e^{-1} \\ \lambda_2^* = 1 - e^{-1} \end{cases}$.
Therefore, we have $D = \sup_\lambda d(\lambda_1,\lambda_2) = d(\lambda_1^*, \lambda_2^*)$, where $(\lambda_1^*, \lambda_2^*) = (e^{-1}, 1-e^{-1})$.
So, from the above, we have $(x^*, y^*) = (\lambda_1^*, \lambda_1^* + \lambda_2^*) = (e^{-1}, 1)$ so that

$$f(x^*, y^*) = e^{-1} + \frac{1}{2} = \min_{(x,y)\in K} f(x,y).$$

◀

Exercise 2. Let $a_i > 0$ for all $i = 1, \ldots, n$ and $K := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n \frac{x_i^2}{a_i^2} \le 1 \right\}$. Solve

$$\min_{x \in K} \|x - u\|^2$$

for a given $u \in \mathbb{R}^n$ satisfying $\sum_{i=1}^n \frac{u_i^2}{a_i^2} > 1$.

Remarks. K is an ellipsoid.

Solution. Let $f(x) = \|x - u\|^2$ and $g(x) = \sum_{i=1}^n \frac{x_i^2}{a_i^2} - 1$.
Then, the Lagrangian function is

$$L(x, \lambda) = \|x - u\|^2 + \lambda \left(\sum_{i=1}^n \frac{x_i^2}{a_i^2} - 1 \right)$$

Note that

- f and g are convex
- $d(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda)$

Now, we compute

$$\nabla_x L(x, \lambda) = \begin{pmatrix} \vdots \\ 2(x_i - u_i) + 2\lambda x_i / a_i^2 \\ \vdots \end{pmatrix}$$

By setting $\nabla_x L(x, \lambda) = \mathbf{0}$, we obtain $x_i = \frac{u_i}{1 + \lambda / a_i^2}$.
Putting back to $d(\lambda)$, we have

$$\begin{aligned} d(\lambda) &= \inf_{x \in \mathbb{R}^n} L(x, \lambda) = L\left(\frac{u_i}{1 + \lambda / a_i},\ i = 1, \ldots, n,\ \lambda \right) \\ &= \sum_{i=1}^n \left(1 - \frac{1}{1 + \lambda / a_i} - 1 \right)^2 u_i^2 + \lambda \left(\sum_{i=1}^n \left(\frac{1}{1 + \lambda / a_i^2} \right)^2 \cdot \frac{1}{a_i^2} - 1 \right) \\ &= -\lambda + \lambda \sum_{i=1}^n \frac{u_i^2}{a_i^2 + \lambda} \end{aligned}$$

To find $\sup_{\lambda \ge 0} d(\lambda)$, we set $d'(\lambda) = 0$ so that

$$\begin{aligned} -1 + \sum_{i=1}^n \frac{u_i^2}{a_i^2 + \lambda} - \lambda \cdot \sum_{i=1}^n \frac{u_i^2}{(a_i + \lambda)^2} = 0 \\ \Longrightarrow 1 = \sum_{i=1}^n \frac{u_i^2 a_i^2}{(a_i^2 + \lambda)^2} \end{aligned}$$

Theortically, we can find λ^* for solving $1 = \sum_{i=1}^n \frac{u_i^2 a_i^2}{(a_i^2 + \lambda^*)^2}$ and $x_i^* = \frac{u_i a_i^2}{a_i^2 + \lambda^*}$. It is encouraged to think on this problem.

◀

— End of Lecture 18 —