THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
MATH4230 2024-25
Lecture 7
February 4, 2025 (Tuesday)

1 Recall

The problem (P) and the *feasible* set K as follows:

$$\inf_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} g_i(x) \le 0, & i=1,\dots,\ell \\ h_j(x)=0, & j=1,\dots,m \end{cases} \tag{P}$$

where $f, g_i, h_j \in C^1$, and

$$K = \{x\in\mathbb{R}^n : g_i(x)\le 0,\ h_j(x)=0,\ i=1,\dots,\ell,\ j=1,\dots,m\}$$

Also, together with optimal solution $\boldsymbol{x}^*$ and the **qualification**, we have the following:

<u>KKT Theorem:</u>
Let $x^*\in K$ be a solution to (P) and assume that K is **qualified** at x^*. Then there exists $\lambda_1,\cdots,\lambda_\ell \ge 0$ and $\mu_1,\cdots,\mu_m\in\mathbb{R}$ such that

$$\begin{cases} \displaystyle\sum_{i=1}^{\ell} \lambda_i g_i(x^*) = 0 \\ \displaystyle\nabla f(x^*) + \sum_{i=1}^{\ell}\lambda_i \nabla g_i(x^*) + \sum_{j=1}^{m}\mu_j\nabla h_j(x^*) = \mathbf{0} \end{cases}$$

For the qualification conditions that we had mentioned before, we have two conditions:

<u>Mangasarian Fromovitz Qualification condition:</u>

① the family of vectors $\{\nabla h_1(x),\dots,\nabla h_m(x)\}$ is linearly independent.

② there exists a vector $v\in\mathbb{R}^n$ satisfying

$$\langle \nabla h_j(x^*), v\rangle = 0,\ \forall j=1,\dots,m$$

and

$$\langle \nabla g_i(x^*), v\rangle < 0,\ \forall i\in I(x) := \{k : g_k(x)=0\}.$$

Then the constraint K is qualified at $x\in K$.

<u>Abadie's Condition:</u>

$$T_K(x) = \Big\{v\in\mathbb{R}^n : \ \exists (s_k, v_k)\to(0^+, v) \ \text{ and } \ x+s_kv_k\in K\Big\}$$
$$D = \left\{ v\in\mathbb{R}^n : \begin{array}{c} \langle\nabla g_i(x), v\rangle \le 0,\ \forall i=1,\dots,\ell \ \text{ satisfying } \ g_i(x)=0 \\ \langle \nabla h_j(x), v\rangle = 0,\ \ \forall j=1,\dots,m \end{array}\right\}$$

If $T_K(x)=D$, then the constraint K is qualified at $x\in K$.

2 Application

Example 1. Solve the following problem:

$$\min_{\substack{x^2+y^2+z^2\le 1\\ x\ge 0}} x+y+z.$$

Solution. Letting $f(x,y,z) = x+y+z$. Clearly, now $n=3$ and we only have the inequality constraints, put $g_1(x,y,z) = x^2+y^2+z^2-1$ and $g_2(x,y,z) = -x$.
Now, first, we consider

$$K = \{(x,y,z)\in\mathbb{R}^3 : g_1(x,y,z)\le 0,\ g_2(x,y,z)\le 0\}$$

is compact, and $f(x,y,z)$ is continuous, so there exists a minimizer, say (x^*,y^*,z^*). Then, we compute

$$\nabla g_1(x,y,z) = \begin{pmatrix}2x\\2y\\2z\end{pmatrix},\quad \nabla g_2(x,y,z) = \begin{pmatrix}-1\\0\\0\end{pmatrix}$$

Let $v = \begin{pmatrix}v_1\\v_2\\v_3\end{pmatrix}$ be such that $v_1>0$, then we have $\langle v, \nabla g_2\rangle = -v_1 < 0$.
Now, to choose v_2, v_3, we consider the following cases:

- **Case 1:** $y^* = z^* = 0$,
  Then $\langle v, \nabla g_1(x,y,z)\rangle = 2v_1\cdot x^* \ge 0$.
  So, the Mangasarian-Fromovitz Qualification condition is not satisfied.
  In this case, we cannot apply the KKT theorem and thus the solution is $(x^*,y^*,z^*) = (0,0,0)$.

- **Case 2:** $y^*z^*\neq 0$
  Then there exists (v_2,v_3) such that

  $$\langle v, \nabla g_1(x,y,z)\rangle = 2\cdot(v_1x^*+v_2y^*+v_3z^*) < 0$$

  So, the Mangasarian-Fromovitz Qualification condition is satisfied.
  As the qualification holds, we can apply the KKT theorem, so there exists $\lambda_1,\lambda_2\ge 0$ such that

  $$\lambda_1 g_1(x,y,z) = \lambda_2 g_2(x,y,z) = 0$$

  and

  $$\begin{pmatrix}1\\1\\1\end{pmatrix} + \lambda_1\begin{pmatrix}2x^*\\2y^*\\2z^*\end{pmatrix} + \lambda_2\begin{pmatrix}-1\\0\\0\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$
  $$\implies \begin{cases}1+2\lambda_1x^*-\lambda_2 = 0\\ 1+2\lambda_1y^* = 0\\ 1+2\lambda_1z^* = 0\end{cases}$$

  From the above, we can deduce that $\lambda_1\neq 0$, so we have $y^* = z^* = -\dfrac{1}{2\lambda_1}$.
  Putting back to the first equation, we have $x^* = \dfrac{\lambda_2-1}{2\lambda_1}\ge 0$.

These imply that $\lambda_1 > 0$ and $\lambda_2 \geq 1 > 0$, so $g_1(x^*, y^*, z^*) = g_2(x^*, y^*, z^*) = 0$.
So, we have $x^* = 0$ and $(x^*)^2 + (y^*)^2 + (z^*)^2 - 1 = 0$.
Putting all together, we have $x^* = 0$ and $y^* = z^* = -\dfrac{\sqrt{2}}{2}$.
By comparing the values of f at $(0, 0, 0)$ and $(0, -\sqrt{2}/2, -\sqrt{2}/2)$, we conclude that

$$f\left(0, -\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right) = -\sqrt{2} < 0 = f(0, 0, 0)$$

and so the optimizer is $(x^*, y^*, z^*) = \left(0, -\dfrac{\sqrt{2}}{2}, -\dfrac{\sqrt{2}}{2}\right)$. ◀

Let us finish this part by introducing one more example.

Example 2. Solve the following problem

$$\min_{\substack{x^2+y^2+z^2=1 \\ x+y+z\leq 0}} x + 2y + 3z.$$

Solution. Letting $f(x, y, z) = x + 2y + 3z$ and $n = 3$. Let $g(x, y, z) = x + y + z$ and $h(x, y, z) = x^2 + y^2 + z^2$.
Now, first, we consider

$$K = \{(x, y, z) \in \mathbb{R}^3 : g(x, y, z) \leq 0,\ h(x, y, z) = 0\}$$

is compact, and $f(x, y, z)$ is continuous, so there exists a minimizer, say (x^*, y^*, z^*).
Secondly, we compute

$$\nabla f = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},\ \nabla g = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},\ \nabla h = \begin{pmatrix} 2x^* \\ 2y^* \\ 2z^* \end{pmatrix}.$$

① For $(x^*, y^*, z^*) \neq \mathbf{0}$, then $\{\nabla h(x^*, y^*.z^*)\}$ is linearly independent.

② There exists $v \in \mathbb{R}^3$ such that

$$\langle v, \nabla h(x^*, y^*, z^*)\rangle = \left\langle v, 2\begin{pmatrix} x^* \\ y^* \\ z^* \end{pmatrix}\right\rangle = 0$$

and

$$\langle v, \nabla g\rangle = \left\langle v, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\rangle \neq 0$$

If $\langle v, \nabla g\rangle < 0$, the qualification condition is automatically satisfied.
If $\langle v, \nabla g\rangle > 0$, then we can replace v by $-v$ so that $\langle -v, \nabla g\rangle < 0$ and $\langle -v, \nabla h(x^*, y^*, z^*)\rangle = 0$ so that the qualification condition holds.

So, the M-F condition holds. By the KKT theorem, there exist $\lambda \geq 0$ and $\mu \in \mathbb{R}$ such that

$$\begin{cases} \lambda g(x^*, y^*, z^*) = 0 \\ \nabla f(x^*, y^*, z^*) + \lambda \nabla g(x^*, y^*, z^*) = 0 + \mu \nabla h(x^*, y^*, z^*) = \mathbf{0} \end{cases}$$

$$\Longrightarrow \begin{cases} \lambda(x^* + y^* + z^*) = 0 & (1) \\ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 2x^* \\ 2y^* \\ 2z^* \end{pmatrix} = \mathbf{0} & (2) \end{cases}$$

By solving equations in (2), we can conclude $\mu \neq 0$ and

$$x^* = -\frac{1+\lambda}{\mu},\ y^* = -\frac{2+\lambda}{\mu},\ z^* = -\frac{3+\lambda}{\mu}.$$

Hence, putting all into (1) and gives

$$\lambda \cdot (x^* + y^* + z^*) = -\frac{\lambda(6+3\lambda)}{\mu} = 0 \implies 3\lambda(2+\lambda) = 0$$

Therefore, we have $\lambda = 0$ or $\lambda = 2$, we consider the following cases:

- **Case 1:** $\lambda = 0$
  Then, we have $x^* = -\frac{1}{\mu}, y^* = -\frac{2}{\mu}$ and $z^* = -\frac{3}{\mu}$.
  Putting $(x^*)^2 + (y^*)^2 + (z^*)^2 = 1$, we have $1 = \frac{1}{\mu^2} + \frac{4}{\mu^2} + \frac{9}{\mu^2} \implies \mu^2 = 14$.
  So, in this case, we can solve $x^* = -\frac{1}{\sqrt{14}}, y^* = -\frac{2}{\sqrt{14}}, z^* = -\frac{3}{\sqrt{14}}$.
- **Case 2:** $\lambda = 2$
  Then, we have $x^* = \frac{1}{\mu},\ y^* = 0$ and $z^* = -\frac{1}{\mu}$.
  Putting $(x^*)^2 + (y^*)^2 + (z^*)^2 = 1$, we have

  $$\begin{aligned} \frac{1}{\mu^2} + \frac{1}{\mu^2} &= 1 \\ \mu &= \sqrt{2} \end{aligned}$$

  So, we solve $x^* = \frac{\sqrt{2}}{2},\ y^* = 0,\ z^* = -\frac{\sqrt{2}}{2}$.

Now, it remains to compare f at $(-1/\sqrt{14}, -2/\sqrt{14}, -3/\sqrt{14})$ and $(\sqrt{2}/2, 0, -\sqrt{2}/2)$.
We have

$$\begin{aligned} f\left(-\frac{1}{\sqrt{14}}, -\frac{2}{\sqrt{14}}, -\frac{3}{\sqrt{14}}\right) &= -\frac{1}{\sqrt{14}} - \frac{4}{\sqrt{14}} - \frac{9}{\sqrt{14}} \\ &= -\sqrt{14} \\ &< -\sqrt{2} \\ &= f\left(\frac{\sqrt{2}}{2}, 0, -\frac{\sqrt{2}}{2}\right) \end{aligned}$$

Thus, we have

$$\min_{\substack{x^2+y^2+z^2=1 \\ x+y+z\leq 0}} x + 2y + 3z = -\sqrt{14}$$

and the optimal solution is $(x^*, y^*, z^*) = \left(-\frac{1}{\sqrt{14}}, -\frac{2}{\sqrt{14}}, -\frac{3}{\sqrt{14}}\right)$. ◀

3 Convex Optimization

Today, we begin a new chapter - Convex Optimization. The main reference for the second part of our course is:

<u>**G.Lan**</u> **`First-order and Stochastic optimization Methods`** **`for Machine Learning`**

In this section, we consider the following problem

$$\inf_{x\in K} f(x) \qquad \begin{cases} f & \text{is convex} \\ K & \text{is convex} \end{cases}$$

and

$$K = \{x \in \mathbb{R}^n : g_i(x) \leq 0,\ h_j(x) = 0,\ i = 1, \ldots, \ell,\ j = 1, \ldots, m\}$$

with $g_i(\cdot)$ and $h_j(\cdot)$ are convex functions.

Convex Set

Definition 1. A set $K \subseteq \mathbb{R}^n$ is said to be **convex** if

$$\lambda x + (1-\lambda)y \in K$$

whenever $x, y \in K$ and $\lambda \in [0,1]$.

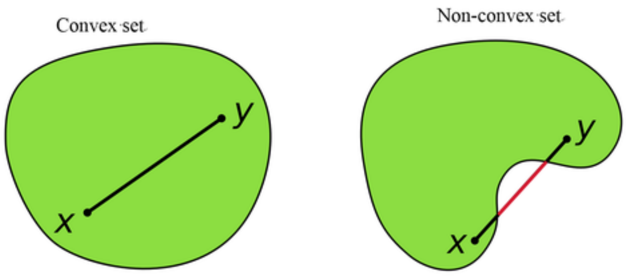

Figure 1: Example and Counterexample of Convex Set

Example 3. The following are examples of Convex sets.

① $\mathbb{R}^n$ is convex.

② $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x_1 \geq 0,\ i = 1, 2, \ldots, n\}$ is convex, because

$$\forall x, y \in \mathbb{R}^n_+ \implies \lambda x + (1-\lambda)y = (\lambda x_i + (1-\lambda)y_i)_{i=1,2,\ldots,n} \in \mathbb{R}^n_+$$

③ Balls: $K := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ is convex because

$$\forall x, y \in K \implies \|\lambda x + (1-\lambda)y\| \leq \lambda\|x\| + (1-\lambda)\|y\| \leq 1$$

④ $K := \{x \in \mathbb{R}^n : A^T x \leq b \ \text{ or } \ A^T x = b, \ \text{where } A \in \mathbb{R}^n, b \in \mathbb{R}\}$ is convex because

$$\forall x, y \in K \implies A^T(\lambda x + (1-\lambda)y) = \lambda \underbrace{A^T x}_{\leq b} + (1-\lambda) \underbrace{A^T y}_{\leq b} \leq b$$

Lemma 1. *Let $K_1, \cdots, K_m$ be convex subsets of $\mathbb{R}^n$. Then*

$$\bigcap_{i=1}^{m} K_i \quad \textit{is convex.}$$

Proof. For any $x, y \in \bigcap K_i$, we have $x \in K_i, \ y_i \in K_i$ for all $i = 1, 2, \ldots, m$.
Since each K_i is convex, so for every $\lambda \in [0, 1]$, we have

$$\begin{aligned} \lambda x + (1-\lambda)y \in K_i \ \ \forall i = 1, \ldots, m \\ \implies \lambda x + (1-\lambda)y \in \bigcap K_i \end{aligned}$$

□

Projection onto convex Set

Definition 2. Let K be a closed convex subset of $\mathbb{R}^n$. Then for any $y \in \mathbb{R}^n$, we define the closed point to y in K as

$$\operatorname{proj}_K(y) = \arg\min_{x \in K} \|y - x\|$$

and call it the projection of y onto K.

Proposition 2. *Let K be a **closed** convex subset of $\mathbb{R}^n$. Then, there exist a **unique** point $x^* \in K$ such that*

$$\|y - x^*\| = \min_{x \in K} \|y - x\|.$$

Remarks. If K is open, then the projection $\operatorname{proj}_K(y)$ may not exist.
For example, if $n = 1$, $K = (0, 1)$ and $y = 2$, but the minimizer to $\inf\limits_{x \in (0,1)} \|y - x\|$ does not exist.

Remarks. It is equivalent to consider $\inf\limits_{x \in K} \|y - x\|^2$ in the definition of $\operatorname{proj}_K(y)$.

— End of Lecture 7 —