THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
MATH4230 2024-25
Lecture 2
January 8, 2025 (Wednesday)

1 Recall

Recall the minimization problem that we had introduced yesterday as follows:

$$\boxed{\inf_{x\in K} f(x) \quad \text{subject to } x \in K \subseteq \mathbb{R}^n,\ f:\mathbb{R}^n \to \mathbb{R}} \tag{P}$$

From the previous lecture 1, we had discuss "Existence of minimizer". Today, let us keep track on the part "First order necessary condition".

2 First order necessary condition of the optimizer x^*

Recall the condition in the previous lecture as follows:

- **Euler's first order condition**

  > If $f(\mathbf{x})$ is **continuously** differentiable, $\emptyset \neq K$ is an open set in $\mathbb{R}^n$ and $\mathbf{x}^* \in K$ is an optimal solution to (P), then
  >
  > $$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

Today, we introduce the following conditions:

- **The Karush-Kuhn-Tucker (KKT) conditions**
  Consider the following nonlinear optimization problem:

  $$(K) : \begin{cases} \text{minimize } f(x) \\ \text{subject to} \\ g_i(x) \leq 0,\ i = 1, 2, \ldots, \ell \\ h_j(x) = 0,\ j = 1, 2, \ldots, m \end{cases}$$

  where $x \in K = \{x \in \mathbb{R}^n : g_i(x) \leq 0,\ i = 1, 2, \ldots, \ell, h_j(x) = 0,\ j = 1, 2, \ldots, m\}$ is the optimization variable from a subset of $\mathbb{R}^n$.
  Then, there exists $p_0, p_1, \ldots, p_\ell \geq 0$, $q_1, q_2, \ldots, q_m \in \mathbb{R}$ such that the following holds:

  - $p_0 \nabla f(x^*) + \sum_{i=1}^{\ell} p_i \nabla g_i(x^*) + \sum_{j=1}^{m} q_j \nabla h_j(x^*) = \mathbf{0}$
  - $(p_1, p_2, \ldots, p_\ell, q_1, \ldots, q_m) \neq \mathbf{0}$
  - $\sum_{i=1}^{\ell} p_i g_i(x^*) = 0 \iff p_i \cdot g_i(x^*) = 0,\ \forall i = 1, \ldots, \ell$

    $$\iff p_i = 0 \ \vee g_i(x^*) = 0\ \forall i = 1, \ldots, \ell$$

In the following, let us do a proof on the above theorem.

Proof. We will complete the proof by following the procedures:

1. Let
$$f_N(x) := f(x) + \|x - x^*\|^2 + \frac{N}{2}\left(\sum_{i=1}^{\ell} \max(0, g_i(x))^2 + \sum_{j=1}^{m} h_j^2(x)\right)$$
be a C^1-penalization function (see Wikipedia). Then, we have
$$\begin{aligned} f_N(x^*) &= f(x^*) + 0 + \frac{N}{2}\left(\sum_{i=1}^{\ell} \max(0, \underbrace{g_i(x^*)}_{\leq 0})^2 + \sum_{j=1}^{m} \underbrace{h_j^2(x^*)}_{0}\right) \\ &= f(x^*) \end{aligned}$$
and $f_N(x) > f(x)$ for any $x \neq x^*$.
Note. $x \mapsto \max(0, x)$ is not C^1 globally, but $x \mapsto [\max(0, x)]^2 \in C^1$ globally.

2. Now, we claim that there exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$, there exists $N_\varepsilon \in \mathbb{N}$ such that $f_{N_\varepsilon}(x) > f_{N_\varepsilon}(x^*)$ for all $x \in \mathbb{R}^n$ with $\|x - x^*\| = \varepsilon$.

3. Then, the question becomes
$$\min_{x \in \mathbb{R}^n} f_{N_\varepsilon}(x) \quad \text{subject to} \quad \|x - x^*\| < \varepsilon$$
Suppose there exists a solution x_ε such that $\|x_\varepsilon - x^*\| < \varepsilon$, then by the Euler's condition, it follows that
$$\begin{aligned} 0 &\overset{(*)}{=} \nabla f_{N_\varepsilon}(x_\varepsilon) \\ &= \nabla f(x_\varepsilon) + 2(x_\varepsilon - x^*) + N_\varepsilon\left(\sum_{i=1}^{\ell} \max(0, g_i(x_\varepsilon)) \cdot \nabla g_i(x_\varepsilon) + \sum_{j=1}^{m} h_j(x_\varepsilon) \cdot \nabla h_j(x_\varepsilon)\right) \end{aligned}$$

4. Now, we construct
$$\rho_\varepsilon := \sqrt{1 + N_\varepsilon^2 \sum_{i=1}^{\ell} \max(0, g_i(x_\varepsilon))^2 + N_\varepsilon^2 \sum_{j=1}^{m} h_j(x_\varepsilon)^2}$$
Also, put $p_0^\varepsilon := \dfrac{1}{\rho_\varepsilon} \geq 0$, $p_i^\varepsilon := \dfrac{N_\varepsilon \max(0, g_i(x_\varepsilon))}{\rho_\varepsilon} \geq 0$ and $q_j^\varepsilon := \dfrac{N_\varepsilon h_j(x_\varepsilon)}{\rho_\varepsilon} \in \mathbb{R}$ so that
$$\begin{aligned} \|(p_1^\varepsilon, \cdots, p_\ell^\varepsilon, q_1^\varepsilon, \cdots, q_m^\varepsilon)\|^2 &= \frac{1}{\rho_\varepsilon^2} + \frac{N_\varepsilon^2 \cdot \sum_{i=1}^{\ell} \max(0, g_i(x_\varepsilon))^2}{\rho_\varepsilon^2} + \frac{N_\varepsilon^2 \cdot \sum_{j=1}^{m} h_j(x_\varepsilon)^2}{\rho_\varepsilon^2} \\ &= \frac{1}{\rho_\varepsilon^2}\left(\sqrt{1 + N_\varepsilon^2 \sum_{i=1}^{\ell} \max(0, g_i(x_\varepsilon))^2 + N_\varepsilon^2 \sum_{j=1}^{m} h_j(x_\varepsilon)^2}\right)^2 \\ &= 1 \end{aligned}$$
From $(*)$, multiplying p_0^ε on both sides yields:
$$p_0^\varepsilon \nabla f(x_\varepsilon) + 2p_0^\varepsilon(x_\varepsilon - x^*) + \sum_{i=1}^{\ell} p_i^\varepsilon \nabla g_i(x_\varepsilon) + \sum_{j=1}^{m} q_j^\varepsilon \nabla h_j(x_\varepsilon) = 0$$

So, we can choose a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ such that $\varepsilon \searrow 0$ and

$$(p_0^{\varepsilon_n}, p_1^{\varepsilon_n}, \cdots, p_\ell^{\varepsilon_n}, q_1^{\varepsilon_n}, \cdots, q_m^{\varepsilon_n}) \to \left(\underbrace{p_0}_{\geq 0}, \underbrace{p_1}_{\geq 0}, \cdots, \underbrace{p_\ell}_{\geq 0}, \underbrace{q_1}_{\in\mathbb{R}}, \cdots, \underbrace{q_m}_{\in\mathbb{R}} \right) \neq \mathbf{0}$$

and $x_{\varepsilon_n} \to x^*$ because $\|x_{\varepsilon_n} - x^*\| < \varepsilon_n \to 0$. Bringing all together, we have

$$0 = p_0^{\varepsilon_n} \nabla f(x_{\varepsilon_n}) + 2p_0^{\varepsilon_n} \underbrace{(x_{\varepsilon_n} - x^*)}_{\to \mathbf{0}} + \sum_{i=1}^{\ell} p_i^{\varepsilon_n} \nabla g_i(x_{\varepsilon_n}) + \sum_{j=1}^{m} q_j^{\varepsilon_n} \nabla h_j(x_{\varepsilon_n})$$

$$\longrightarrow p_0 \nabla f(x^*) + \sum_{i=1}^{\ell} p_i \nabla g_i(x^*) + \sum_{j=1}^{m} q_j \nabla h_j(x^*)$$

$$\boxed{p_0 \nabla f(x^*) + \sum_{i=1}^{\ell} p_i \nabla g_i(x^*) + \sum_{j=1}^{m} q_j \nabla h_j(x^*) = \mathbf{0}}$$

Now, it remains to show item 3 in the KKT condition in page 1.
If $g_i(x^*) < 0$, then from $p_i^\varepsilon = \dfrac{N_\varepsilon \max(0, g_i(x^*))}{\rho_\varepsilon} = 0$ for small enough $\varepsilon > 0$.
This gives $p_i = 0$ for each $i = 1, 2, \ldots, \ell$.
Similarly, if $p_i > 0$ but $g_i(x^*) \neq 0$ for all $i = 1, 2, \ldots, \ell$, then

$$0 = \sum_{i=1}^{\ell} p_i g_i(x^*) < \underbrace{p_1}_{>0} \underbrace{g_1(x^*)}_{<0} < 0$$

Contradiction arises. Thus, this proves that

$$\sum_{i=1}^{\ell} p_i g_i(x^*) = 0 \iff p_i \cdot g_i(x^*) = 0, \ \forall i = 1, \ldots, \ell$$

$$\iff p_i = 0 \ \vee g_i(x^*) = 0 \ \forall i = 1, \ldots, \ell$$

□

Remarks. To complete the whole proof, we need to look back on the claim that we stated in item 2. Please refer to lecture 5 for the detail proof of the claim.

— End of Lecture 2 —