THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
MATH4230 2024-25
Lecture 1
January 7, 2025 (Tuesday)

1 Introduction

Usually, we want to solve some problems related to maximum or minimum, and subject to some constraints, such as the following:

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

In this course, we will mainly learn the following topics:

- Existence of solution
- Characterization of the optimizer
- Convex optimization or duality
- Numerical method and applications

Let us first focus on "Existence of solution", and it is natural to ask the following questions:

1. When is the existence of the minimizer x^*?
2. If x^* exists, which properties that it has to satisfy?
3. How to find the minimizer x^* and the optimal value $f(x^*) = \min_{x \in K} f(x)$ numerically?

In the first lesson, we will focus on the section of "existence of optimal solution". The main reference for the first part of our course is:

D. Michael Patricksson:
`An introduction to continuous optimization Foundations and Fundamental Algorithms (Section 4)`

2 Existence of Optimal Solution

Definition 1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and $K \subseteq \mathbb{R}^n$ be a non-empty set. We call the value $\ell \in [-\infty, +\infty)$ the **infimum** of $f(x)$ on K if ℓ satisfies

1. $\ell \leq f(x)$, for all $x \in K$,
2. there exists a sequence $(x_n)_{n \geq 1} \subseteq K$ such that $\lim_{n \to \infty} f(x_n) = \ell$.

Then, we denote

$$\boxed{\inf_{x \in K} f(x) := \ell}$$

Definition 2. If $\inf_{x \in K} f(x) > -\infty$ and there exists $x^* \in K$ such that $f(x^*) = \inf_{x \in K} f(x)$, we say

1. $f^* := f(x^*)$ is the minimum value of the problem $\inf_{x \in K} f(x)$, denoted as $\min_{x \in K} f(x)$ (P)
2. x^* is a solution to the minimization problem (P).

 Prepared by Max Shung

Definition 3. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called **coercive** if

$$\lim_{\|x\| \to +\infty} f(x) = +\infty \ .$$

Example 1. Let $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$, where $\mathbf{b} \in \mathbb{R}^n$ is a vector, $c \in \mathbb{R}$ is a constant and A is an $n \times n$ symmetric positive-definite matrix. Then $f(\mathbf{x})$ is coercive.

Proof. As A is positive-definite matrix, so we have $\mathbf{x}^T A\mathbf{x} > 0$ for any $\mathbf{x} \in \mathbb{R}^n$.
Moreover, if λ is an eigenvalue of A, then we have

$$\mathbf{x}^T A\mathbf{x} = \lambda \mathbf{x}^T\mathbf{x} = \lambda\|\mathbf{x}\|^2 > 0$$

So, any eigenvalues of A must be strictly positive. Denote

$$\lambda_{\text{min}} := \min\{\lambda : A\mathbf{x} = \lambda\mathbf{x},\ \mathbf{x} \neq \mathbf{0}\} > 0$$

then we have

$$\mathbf{x}^T A\mathbf{x} \geq \lambda_{\text{min}}\|\mathbf{x}\|^2.$$

Moreover, as $\mathbf{b}$ and c are constants, so we have

$$f(\mathbf{x}) \geq \underbrace{\lambda_{\text{min}}\|\mathbf{x}\|^2}_{\text{Dominated}} + \mathbf{b}^T\mathbf{x} + c \to +\infty$$

as $\|\mathbf{x}\| \to +\infty$. This proves that f is corecive. $\square$

Proposition 1. *Let $f_i : \mathbb{R} \to \mathbb{R}$ be bounded below and coercive functions for each $i = 1, 2, \ldots, n$. Then*

$$f(\mathbf{x}) = \sum_{i=1}^{n} f_i(x_i)$$

*is also **coercive** for all $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.*

Proof. Without loss of generality, we may assume $f_i(x_i) \geq 0$ for all $x_i \in \mathbb{R}$, $i = 1, 2, \ldots, n$.
As f_i is bounded below, that is, there exists some $m \in \mathbb{R}$ such that $f_i \geq m$ for all i, we may replace

$$g = f_i - m \geq 0$$

if necessarily. Then, using this fact, we have

$$f(\mathbf{x}) = \sum_{i=1}^{n} f_i(x_i) \geq \max_{1 \leq i \leq n} f_i(x_i) \quad (\because f_i \geq 0,\ \forall i)$$

On the other hand, if $\|\mathbf{x}\| \to +\infty$, then $\max\limits_{1 \leq i \leq n} |x_i| \to +\infty$. Now, it remains to show $f(\mathbf{x}) \to +\infty$ to conclude that f is coercive. Suppose on contradictory that $\lim\limits_{\|\mathbf{x}\| \to +\infty} f(\mathbf{x}) \neq +\infty$, then there exists a sequence of points $(\mathbf{x}^m)_{m \geq 1}$ such that

$$\lim_{m \to +\infty} f(\mathbf{x}^m) < +\infty \quad \text{as} \quad \|\mathbf{x}^m\| \to +\infty$$

Since $f \geq \max\limits_{1 \leq i \leq n} f_i$, so this follows that $\overline{\lim\limits_{m \to +\infty}} f_i(x_i^m) < +\infty$ for all $i = 1, \ldots, n$.
As $\|\mathbf{x}^m\| \to +\infty$, there exists a subsequence $(m_k)_{m \geq 1}$ and i so that $|x_i^{m_k}| \to +\infty$.
Using the given fact that f_i are coercive for all $i = 1, 2, \ldots, n$, so

$$\overline{\lim_{m \to +\infty}} f_i(x_i) \geq \lim_{k \to +\infty} f_i(x_i^{m_k}) = +\infty > \lim f(\mathbf{x}^m)$$

Contradiction occurs, so we have $\lim\limits_{\|\mathbf{x}\| \to +\infty} f(\mathbf{x}) = +\infty$, i.e. f is also coercive. $\square$

Theorem 2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function, and $K \subseteq \mathbb{R}^n$ be a non-empty subset. Then*

1. *If K is compact, then (P) has an optimal solution $\mathbf{x}^* \in K$*
2. *If K is closed, and $f(\mathbf{x})$ is coercive, then (P) has an optimal solution $\mathbf{x}^* \in K$.*

Proof. 1. If K is compact, consider a sequence $(\mathbf{x}_n)_{n\geq 1} \subseteq K$ be a minimizing sequence of (P), that is

$$\lim_{n\to+\infty} f(\mathbf{x}_n) = \inf_{\mathbf{x}\in K} f(\mathbf{x})$$

Then there exists a subsequence $(\mathbf{x}_{n_k})_{k\geq 1}$ such that $\lim_{k\to+\infty} \mathbf{x}_{n_k} = \mathbf{x}^* \in K$ (by closeness), and

$$f(\mathbf{x}^*) = \lim_{k\to+\infty} f(\mathbf{x}_{n_k}) = \inf_{\mathbf{x}\in K} f(\mathbf{x})$$

(this step aims to concretely define $\mathbf{x}^*$.)

Alternative Proof: Applying Extreme-Value Theorem for f over compact K, then prove the existence by intermediate value theorem.

2. If $f(\mathbf{x})$ is coercive, so by definition we have

$$\lim_{\|\mathbf{x}\|\to+\infty} f(\mathbf{x}) = +\infty$$

Then, we have

$$\inf_{\mathbf{x}\in K} f(\mathbf{x}) = \inf_{\mathbf{x}\in K_n} f(\mathbf{x})$$

where $K_n := \{\mathbf{x} \in K : \|\mathbf{x}\| \leq n\}$ for sufficiently large enough $n \in \mathbb{N}$.
Since K_n is closed (as its complement is open), and bounded in $\mathbb{R}^n$, hence it is compact.
Therefore, there exists $\mathbf{x}^* \in K_n \subseteq K$ such that

$$f(\mathbf{x}^*) = \inf_{\mathbf{x}\in K_n} f(\mathbf{x}) = \inf_{\mathbf{x}\in K} f(\mathbf{x}).$$

□

Example 2. The set

$$K = \{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \leq 0,\ h(\mathbf{x}) = 0\}$$

is closed if g and h are continuous.

Proof. It suffices to prove its complement is open, that is

$$K' := \mathbb{R}^n \setminus K = \{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) > 0 \text{ or } h(\mathbf{x}) \neq 0\}$$

Now, we aim to find open balls of all $\mathbf{x}$ contained in K'.
Let $\varepsilon > 0$ be arbitrary. By the continuity of g, if $g(\mathbf{x}) > 0$, there exists some $\delta > 0$ such that for any $\mathbf{y}$ satisfying $\|\mathbf{y} - \mathbf{x}\| < \delta$, then we have

$$|g(\mathbf{y}) - g(\mathbf{x})| < \varepsilon$$

i.e. $g(\mathbf{y}) > g(\mathbf{x}) - \varepsilon > 0$, so there exists an open ball $B_\delta(\mathbf{x}) \subseteq K_1' := \{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) > 0\}$.
Moreover, by continuity of h as well, there exists some $\delta' > 0$ such that for any $\mathbf{y} \in B_{\delta'}(\mathbf{x})$, then

$$|h(\mathbf{y}) - h(\mathbf{x})| < \varepsilon$$

By choosing $\varepsilon = |h(\mathbf{x})|/2 > 0$, we have $|h(\mathbf{y})| > |h(\mathbf{x})|/2 > 0$, so this proves there exists an open ball $B_{\delta'}(\mathbf{x}) \subseteq K_2' := \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) \neq 0\}$. Since $K' = K_1' \cup K_2'$, and from the above, there exists $\delta^* := \min\{\delta, \delta'\}$ such that $B_{\delta^*}(\mathbf{x}) \subseteq K'$. This proves that K' is open, that is equivalently showing that K is closed. □

Proposition 3. *Let $K \subseteq \mathbb{R}^n$ be a bounded and open subset, and $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function on $\bar{K}$ (the closure of K). If there exists $\mathbf{x}_0 \in K$ such that $f(\mathbf{x}_0) \le f(\mathbf{x})$, for all $\mathbf{x} \in \partial K = \bar{K} \setminus K$, then (P) has an optimal solution $\mathbf{x}^* \in K$.*

Proof. Since there exists such $\mathbf{x}_0 \in K$ such that $f(\mathbf{x}_0) \le f(\mathbf{x})$, for all $\mathbf{x} \in \partial K$, so it follows that

$$\inf_{\mathbf{x}\in K} f(\mathbf{x}) = \inf_{\mathbf{x}\in \bar{K}} f(\mathbf{x})$$

As $\bar{K}$ is closed set (after taking closure), and bounded (given), so it is compact.
Moreover, as f is continuous on $\bar{K}$, so by the Extreme-Value Theorem, there exists $\mathbf{x}^* \in \bar{K}$ such that

$$f(\mathbf{x}^*) = \inf_{\mathbf{x}\in \bar{K}} f(\mathbf{x})$$

Now, it remains to conclude that $\mathbf{x}^* \in K$ but not on ∂K. We consider the following cases:

- **Case 1:** $\mathbf{x}^* \in \partial K$
  From $f(\mathbf{x}_0) \le f(\mathbf{x})$ for all $\mathbf{x} \in \partial K$ consequences

  $$f(\mathbf{x}_0) \le f(\mathbf{x}^*) = \inf_{\mathbf{x}\in \bar{K}} f(\mathbf{x})$$

  this implies that $\mathbf{x}_0 \in K$ is a solution to (P).

- **Case 2:** $\mathbf{x}^* \in K$
  From the above $f(\mathbf{x}^*) = \inf\limits_{\mathbf{x}\in \bar{K}} f(\mathbf{x})$, it follows that $\mathbf{x}^* \in K$ is a solution to (P).

In any case, there exists $\mathbf{x}^* \in K$ is a solution to (P). $\square$

3 First order necessary condition of the optimizer $\mathbf{x}^*$

After discussing the existence of optimal solution to (P), it is natural to ask what properties that optimizers have. So, we introduce the following theorem:

Theorem 4. *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*

$$\nabla f(\mathbf{x}^*) = \mathbf{0}$$

*This is called the **Euler's first order condition**.*

Proof. We will prove by contradiction. Suppose that $\nabla f(\mathbf{x}^*) = \begin{pmatrix} \partial_{x_1} f(\mathbf{x}^*) \\ \vdots \\ \partial_{x_n} f(\mathbf{x}^*) \end{pmatrix} \neq \mathbf{0}$, then there exists some $\mathbf{v} \in \mathbb{R}^n$ such that

$$\langle \mathbf{v}, \nabla f(\mathbf{x}^*) \rangle > 0.$$

Let $\varepsilon > 0$. Define $\mathbf{x}_\varepsilon := \mathbf{x}^* - \varepsilon \mathbf{v}$, then applying the taylor-expansion of f about $\mathbf{x}_\varepsilon$, we have

$$\begin{aligned} f(\mathbf{x}_\varepsilon) &= f(\mathbf{x}^*) + \langle \mathbf{x}_\varepsilon - \mathbf{x}^*, \nabla f(\mathbf{x}^*) \rangle + O(\varepsilon^2) \\ &= f(\mathbf{x}^*) - \varepsilon \langle \mathbf{v}, \nabla f(\mathbf{x}^*) \rangle + O(\varepsilon^2) \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, taking $\varepsilon \to 0^+$ yields $f(\mathbf{x}_\varepsilon) < f(\mathbf{x}^*)$ and $\mathbf{x}_\varepsilon \in K$.
This contradicts the fact that $\mathbf{x}^*$ is an optimal solution because $f(\mathbf{x}^*) \neq \inf\limits_{\mathbf{x}\in K} f(\mathbf{x})$. $\square$

— End of Lesson 1 —