MA 5037: Optimization Methods and Applications
Optimization Over a Convex Set

Suh-Yuh Yang (楊肅煜)

Department of Mathematics, National Central University
Jhongli District, Taoyuan City 320317, Taiwan

First version: July 28, 2018/Last updated: June 17, 2025

Convex optimization problem

- *We will consider the constrained optimization problem (P):*

$$(P) \qquad \min f(\boldsymbol{x}) \quad \text{s.t. } \boldsymbol{x} \in C,$$

  where f is a continuously differentiable function and $C \subseteq \mathbb{R}^n$ is a closed and convex set.

- *For an unconstrained optimization problem, the stationary points of continuously differentiable functions are points that the gradient vanishes. It was shown that stationarity is a necessary condition for a point to be an unconstrained local optimum point.*

- **Definition:** *(stationary points of constrained problems) Let f be a continuously differentiable function over a closed convex set $C \subseteq \mathbb{R}^n$. Then $\boldsymbol{x}^* \in C$ is called a stationary point of problem (P) if*

$$\boxed{\nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \forall\, \boldsymbol{x} \in C}$$

- Stationarity actually means that there are no feasible descent directions of f at $\boldsymbol{x}^*$. This suggests that stationarity is in fact a necessary condition for a local minimum of (P).

Stationarity as a necessary optimality condition

- **Theorem:** *Let f be a continuously differentiable function over a closed convex set $C \subseteq \mathbb{R}^n$, and let $\boldsymbol{x}^* \in C$ be a local minimum of (P). Then $\boldsymbol{x}^*$ is a stationary point of problem (P).*

  Proof: Assume in contradiction that $\boldsymbol{x}^*$ is not a stationary point of (P). Then $\exists\, \boldsymbol{x} \in C$ such that $\nabla f(\boldsymbol{x}^*)^\top(\boldsymbol{x} - \boldsymbol{x}^*) < 0 \Rightarrow f'(\boldsymbol{x}^*; \boldsymbol{d}) < 0$, where $\boldsymbol{d} := \boldsymbol{x} - \boldsymbol{x}^*$. It follows that $\exists\, \varepsilon \in (0,1)$ s.t. $f(\boldsymbol{x}^* + t\boldsymbol{d}) < f(\boldsymbol{x}^*)$ for all $t \in (0, \varepsilon)$. Since C is convex, $\boldsymbol{x}^* + t\boldsymbol{d} = \boldsymbol{x}^* + t(\boldsymbol{x} - \boldsymbol{x}^*) = (1-t)\boldsymbol{x}^* + t\boldsymbol{x} \in C$. Therefore, $f(\boldsymbol{x}^*)$ is not a local minimum. This is a contradiction! $\square$

- **Note:** *If $C = \mathbb{R}^n$, then the stationary points of problem (P) are the points $\boldsymbol{x}^*$ satisfying $\nabla f(\boldsymbol{x}^*)^\top(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0$, for all $\boldsymbol{x} \in \mathbb{R}^n$. Plugging $\boldsymbol{x} = \boldsymbol{x}^* - \nabla f(\boldsymbol{x}^*)^\top$ into the above inequality, we obtain $-\|\nabla f(\boldsymbol{x}^*)\|^2 \geq 0$, and hence $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.*

  Therefore, it follows that the notion of a stationary point of a function and a stationary point of a minimization problem coincide when the problem is unconstrained.

Stationarity over $C = \mathbb{R}^n_+$

Consider the optimization problem:

$$(Q) \qquad \min f(\boldsymbol{x}) \quad \text{s.t. } x_i \geq 0,\ i = 1, 2, \cdots, n,$$

where f is a continuously differentiable function over $\mathbb{R}^n_+$. By definition, a vector $\boldsymbol{x}^* \in \mathbb{R}^n_+$ is a stationary point of problem (Q) if and only if

$$\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x} - \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x}^* \geq 0, \quad \forall\, \boldsymbol{x} \geq \boldsymbol{0}. \qquad (\star)$$

We will now use the following technical result: $\boldsymbol{a}^\top \boldsymbol{x} + b \geq 0\ \forall\, \boldsymbol{x} \geq \boldsymbol{0}$ iff $\boldsymbol{a} \geq \boldsymbol{0}$ and $b \geq 0$. Thus, $(\star)$ holds iff $\nabla f(\boldsymbol{x}^*) \geq \boldsymbol{0}$ and $\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x}^* \leq 0$. Since $\boldsymbol{x}^* \geq \boldsymbol{0}$, we have $(\star)$ if and only if

$$\nabla f(\boldsymbol{x}^*) \geq \boldsymbol{0} \text{ and } x_i^* \frac{\partial f}{\partial x_i}(\boldsymbol{x}^*) = 0, \quad i = 1, 2, \cdots, n.$$

We can compactly write the above condition as follows:

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}^*) = \begin{cases} = 0 & x_i^* > 0, \\ \geq 0, & x_i^* = 0. \end{cases}$$

Stationarity over the unit-sum set

Consider the optimization problem:

$$(R) \qquad \min f(\boldsymbol{x}) \quad \text{s.t. } \boldsymbol{e}^\top \boldsymbol{x} = 1,$$

where f is a continuously differentiable function over $\mathbb{R}^n$. The following feasible set is called the unit-sum set:

$$U = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{e}^\top \boldsymbol{x} = 1\} = \{\boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1\}.$$

A point $\boldsymbol{x}^* \in U$ is a stationary point of problem (R) if and only if

$$(I) \qquad \nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \ \forall\, \boldsymbol{x} \text{ satisfying } \boldsymbol{e}^\top \boldsymbol{x} = 1.$$

We will show that condition (I) is equivalent to

$$(II) \qquad \frac{\partial f}{\partial x_1}(\boldsymbol{x}^*) = \frac{\partial f}{\partial x_2}(\boldsymbol{x}^*) = \cdots = \frac{\partial f}{\partial x_n}(\boldsymbol{x}^*).$$

(II) $\Rightarrow$ (I): Assume that $\boldsymbol{x}^* \in U$ satisfies (II). Then for any $\boldsymbol{x} \in U$,

$$\nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) = \frac{\partial f}{\partial x_1}(\boldsymbol{x}^*) \left(\sum_{i=1}^n x_i - \sum_{i=1}^n x_i^* \right) = \frac{\partial f}{\partial x_1}(\boldsymbol{x}^*)(1 - 1) = 0.$$

We have thus shown that (I) is satisfied.

Stationarity over the unit-sum set (cont'd)

(I) $\Rightarrow$ (II): Take $\boldsymbol{x}^* \in U$ that satisfies (I). Suppose in contradiction that (II) does not hold. Then $\exists\, i \neq j$ such that $\dfrac{\partial f}{\partial x_i}(\boldsymbol{x}^*) > \dfrac{\partial f}{\partial x_j}(\boldsymbol{x}^*)$. Define the vector $\boldsymbol{x} \in U$ as

$$x_k = \begin{cases} x_k^* & k \notin \{i,j\}, \\ x_i^* - 1 & k = i, \\ x_j^* + 1 & k = j. \end{cases}$$

Then we have

$$\begin{aligned} \nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) &= \frac{\partial f}{\partial x_i}(\boldsymbol{x}^*)(x_i - x_i^*) + \frac{\partial f}{\partial x_j}(\boldsymbol{x}^*)(x_j - x_j^*) \\ &= -\frac{\partial f}{\partial x_i}(\boldsymbol{x}^*) + \frac{\partial f}{\partial x_j}(\boldsymbol{x}^*) < 0, \end{aligned}$$

which is a contradiction to the assumption that (I) is satisfied. Hence, we have (II).

Stationarity over the unit-ball

Consider the optimization problem:

$$(S) \qquad \min f(\boldsymbol{x}) \quad \text{s.t. } \|\boldsymbol{x}\| \leq 1,$$

where f is a continuously differentiable function over $B[\mathbf{0}, 1]$. Then

A point $\boldsymbol{x}^* \in B[\mathbf{0}, 1]$ is a stationary point of (S)

$$\iff \nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \forall \|\boldsymbol{x}\| \leq 1$$

$$\iff \min\{\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x} - \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x}^* : \|\boldsymbol{x}\| \leq 1\} \geq 0 \qquad (\star)$$

Claim: $\forall\, \boldsymbol{a} \in \mathbb{R}^n$ *the optimal value of* $\min\{\boldsymbol{a}^\top \boldsymbol{x} : \|\boldsymbol{x}\| \leq 1\}$ *is "$-\|\boldsymbol{a}\|$".*
Proof: The case of $\boldsymbol{a} = \mathbf{0}$ is trivial. Assume that $\boldsymbol{a} \neq \mathbf{0}$, then by the CS inequality, for any $\boldsymbol{x} \in B[\mathbf{0}, 1]$, we have $\boldsymbol{a}^\top \boldsymbol{x} \geq -\|\boldsymbol{a}\|\|\boldsymbol{x}\| \geq -\|\boldsymbol{a}\|$, so that $\min\{\boldsymbol{a}^\top \boldsymbol{x} : \|\boldsymbol{x}\| \leq 1\} \geq -\|\boldsymbol{a}\|$. The lower bound is attained at $\boldsymbol{x} := -\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}$. $\square$

Returning to the characterization of stationary points, from the claim, we have $(\star)$ iff $-\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x}^* \geq \|\nabla f(\boldsymbol{x}^*)\|$. However, by the CS inequality, we have $-\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x}^* \leq \|\nabla f(\boldsymbol{x}^*)\|\|\boldsymbol{x}^*\| \leq \|\nabla f(\boldsymbol{x}^*)\|$.

Stationarity over the unit-ball (cont'd)

Finally, we can conclude that $\boldsymbol{x}^*$ is a stationary point of (S) iff

$$\|\nabla f(\boldsymbol{x}^*)\| = -\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x}^*. \qquad (\star\star)$$

Let $\boldsymbol{x}^*$ be a point satisfying $(\star\star)$. Then

- If $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$, then $(\star\star)$ holds automatically.
- If $\nabla f(\boldsymbol{x}^*) \neq \mathbf{0}$, then $\|\boldsymbol{x}^*\| = 1$ since otherwise, if $\|\boldsymbol{x}^*\| < 1$ then by the CS inequality,

$$\|\nabla f(\boldsymbol{x}^*)\| = -\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x}^* \leq \|\nabla f(\boldsymbol{x}^*)\| \|\boldsymbol{x}^*\| < \|\nabla f(\boldsymbol{x}^*)\|,$$

which is a contradiction. We therefore conclude that when $\nabla f(\boldsymbol{x}^*) \neq \mathbf{0}$, $\boldsymbol{x}^*$ is a stationary point if and only if $\|\boldsymbol{x}^*\| = 1$ and

$$\begin{aligned} &\|\nabla f(\boldsymbol{x}^*)\| \cdot \|\boldsymbol{x}^*\| = \|\nabla f(\boldsymbol{x}^*)\| = -\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{x}^* \\ &\underbrace{\Longleftrightarrow}_{\textit{by CS}} \exists\, \lambda < 0 \text{ such that } \nabla f(\boldsymbol{x}^*) = \lambda \boldsymbol{x}^*. \end{aligned}$$

In conclusion, $\boldsymbol{x}^*$ is a stationary point of (S) if and only if either $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ or $\|\boldsymbol{x}^*\| = 1$ and $\exists\, \lambda < 0$ such that $\nabla f(\boldsymbol{x}^*) = \lambda \boldsymbol{x}^*$.

Some stationarity conditions

We summarize the results obtained above in the following table:

feasible set	explicit stationarity condition
$\mathbb{R}^n$	$\nabla f(\mathbf{x}^*) = 0$
$\mathbb{R}^n_+$	$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) \begin{cases} = 0, & x_i^* > 0 \\ \geq 0, & x_i^* = 0 \end{cases}$
$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{e}^T\mathbf{x} = 1\}$	$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) = \cdots = \frac{\partial f}{\partial x_n}(\mathbf{x}^*)$
$B[0,1]$	$\nabla f(\mathbf{x}^*) = 0$ or $\|\mathbf{x}^*\| = 1$ and $\exists \lambda \leq 0 : \nabla f(\mathbf{x}^*) = \lambda \mathbf{x}^*$

Stationarity in convex problems

- *Stationarity is a necessary optimality condition for local optimality. However, when the objective function is additionally assumed to be convex, stationarity is a necessary and sufficient condition for optimality. See the theorem below.*

- **Theorem:** *Let f be a continuously differentiable convex function over a closed and convex set $C \subseteq \mathbb{R}^n$. Then $\boldsymbol{x}^*$ is a stationary point of*

$$(P) \qquad \min f(\boldsymbol{x}) \quad \text{s.t. } \boldsymbol{x} \in C$$

if and only if $\boldsymbol{x}^*$ is an optimal solution of (P).

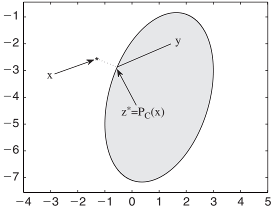
Proof: If $\boldsymbol{x}^*$ is an optimal solution of (P), then by Theorem 9.2 (page 3), it follows that $\boldsymbol{x}^*$ is a stationary point of (P). Assume that $\boldsymbol{x}^*$ is a stationary point of (P), and let $\boldsymbol{x} \in C$. Then from the gradient inequality for convex functions, we have

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) \geq f(\boldsymbol{x}^*).$$

This shown that $\boldsymbol{x}^*$ is the global minimum point of (P). $\square$

The second projection theorem

- *Geometrically, the second projection theorem states that for a given closed and convex set C, $\boldsymbol{x} \in \mathbb{R}^n$ and $y \in C$, the angle between $\boldsymbol{x} - P_C(\boldsymbol{x})$ and $\boldsymbol{y} - P_C(\boldsymbol{x})$ is greater than or equal to $90°$.*



- **Theorem:** *(second projection theorem) Let $C \subseteq \mathbb{R}^n$ be a closed convex set and let $\boldsymbol{x} \in \mathbb{R}^n$. Then $\boldsymbol{z} = P_C(\boldsymbol{x})$ if and only if $\boldsymbol{z} \in C$ and $(\boldsymbol{x} - \boldsymbol{z})^\top(\boldsymbol{y} - \boldsymbol{z}) \leq 0$ for any $\boldsymbol{y} \in C$.*

  Proof: $\boldsymbol{z} = P_C(\boldsymbol{x})$ if and only if it is the optimal solution of the problem

  $$\min\ g(\boldsymbol{y}) := \|\boldsymbol{y} - \boldsymbol{x}\|^2 \quad \text{s.t. } \boldsymbol{y} \in C.$$

  It follows that $\boldsymbol{z} = P_C(\boldsymbol{x})$ if and only if $\nabla g(\boldsymbol{z})^\top(\boldsymbol{y} - \boldsymbol{z}) \geq 0\ \forall\, \boldsymbol{y} \in C$, i.e.,

  $$(\boldsymbol{x} - \boldsymbol{z})^\top(\boldsymbol{y} - \boldsymbol{z}) \geq 0 \quad \forall\, \boldsymbol{y} \in C. \qquad \square$$

Nonexpansiveness property of P_C

Let $C \subseteq \mathbb{R}^n$ be a closed convex set. Then

(1) $\forall\, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$, $(P_C(\boldsymbol{v}) - P_C(\boldsymbol{w}))^\top (\boldsymbol{v} - \boldsymbol{w}) \geq \|P_C(\boldsymbol{v}) - P_C(\boldsymbol{w})\|^2$.

Proof: By the second projection theorem, for any $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in C$ we have

$$(\boldsymbol{x} - P_C(\boldsymbol{x}))^\top (\boldsymbol{y} - P_C(\boldsymbol{x})) \leq 0.$$

Substituting $\boldsymbol{x} = \boldsymbol{v}$ and $\boldsymbol{y} = P_C(\boldsymbol{w})$, $\boldsymbol{x} = \boldsymbol{w}$ and $\boldsymbol{y} = P_C(\boldsymbol{v})$, we have

$$(\boldsymbol{v} - P_C(\boldsymbol{v}))^\top (P_C(\boldsymbol{w}) - P_C(\boldsymbol{v})) \leq 0 \quad \text{and} \quad (\boldsymbol{w} - P_C(\boldsymbol{w}))^\top (P_C(\boldsymbol{v}) - P_C(\boldsymbol{w})) \leq 0.$$

Adding the two inequalities yields

$$(P_C(\boldsymbol{w}) - P_C(\boldsymbol{v}))^\top (\boldsymbol{v} - \boldsymbol{w} + P_C(\boldsymbol{w}) - P_C(\boldsymbol{v})) \leq 0,$$

showing the desired inequality. $\square$

(2) *(nonexpansiveness)* $\forall\, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$, $\|P_C(\boldsymbol{v}) - P_C(\boldsymbol{w})\| \leq \|\boldsymbol{v} - \boldsymbol{w}\|$.

Proof: Assume that $P_C(\boldsymbol{v}) \neq P_C(\boldsymbol{w})$. Then by the CS inequality we have

$$(P_C(\boldsymbol{v}) - P_C(\boldsymbol{w}))^\top (\boldsymbol{v} - \boldsymbol{w}) \leq \|P_C(\boldsymbol{v}) - P_C(\boldsymbol{w})\| \|\boldsymbol{v} - \boldsymbol{w}\|,$$

which combined with (1) yields

$$\|P_C(\boldsymbol{v}) - P_C(\boldsymbol{w})\|^2 \leq \|P_C(\boldsymbol{v}) - P_C(\boldsymbol{w})\| \|\boldsymbol{v} - \boldsymbol{w}\|,$$

showing the desired inequality. $\square$

An additional useful representation of stationarity

The next result describes an additional useful representation of stationarity in terms of the orthogonal projection operator

Theorem: *Let f be a continuously differentiable function defined on the closed and convex set $C \subseteq \mathbb{R}^n$ and $s > 0$. Then $\boldsymbol{x}^*$ is a stationary point of*

$$(P) \qquad \min f(\boldsymbol{x}) \quad \text{s.t. } \boldsymbol{x} \in C$$

if and only if

$$\boldsymbol{x}^* = P_C(\boldsymbol{x}^* - s\nabla f(\boldsymbol{x}^*)).$$

Proof: By the second projection theorem, $\boldsymbol{x}^* = P_C(\boldsymbol{x}^* - s\nabla f(\boldsymbol{x}^*))$ if and only if

$$(\boldsymbol{x}^* - s\nabla f(\boldsymbol{x}^*) - \boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) \leq 0 \quad \forall\, \boldsymbol{x} \in C$$

if and only if

$$\nabla f(\boldsymbol{x}^*)^\top (\boldsymbol{x} - \boldsymbol{x}^*) \geq 0 \quad \forall\, \boldsymbol{x} \in C.$$

That is, $\boldsymbol{x}^*$ is a stationary point of the problem (P). $\square$

The gradient projection method

The stationarity condition $\boldsymbol{x}^* = P_C(\boldsymbol{x}^* - s\nabla f(\boldsymbol{x}^*))$ naturally motivates the following algorithm for solving problem (P):

The gradient projection method:

Input: $\varepsilon > 0$, tolerance parameter.

Initialization: Pick $\boldsymbol{x}_0 \in \mathbb{R}^n$ arbitrarily.

General step: For any $k = 0, 1, \cdots$, execute the following steps

(a) Pick a stepsize t_k by a line search procedure.

(b) Set $\boldsymbol{x}_{k+1} = P_C(\boldsymbol{x}_k - t_k\nabla f(\boldsymbol{x}_k))$.

(c) if $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| \leq \varepsilon$ then stop, and $\boldsymbol{x}_{k+1}$ is the output.

Note:

(1) *In the unconstrained case, that is, when* $C = \mathbb{R}^n$, *the gradient projection method is just the gradient method.*

(2) *There are several strategies for choosing the stepsizes* t_k. *Two choices are (i) constant stepsize* $t_k = \bar{t}$ *for all k; (ii) backtracking.*