Chapter 6　Basics Concepts of Optimization

Optimization Problem

optimization problem: $\min f(\boldsymbol{x})$, s.t. $\boldsymbol{x} \in \Omega$

- $f : \mathbb{R}^n \to \mathbb{R}$ is called the objective/cost function.
- $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^\top \in \mathbb{R}^n$ is called the decision variables.
- $\Omega \subseteq \mathbb{R}^n$ is called the constraint/feasible set.

★ If f has many minimizers, finding one of them will suffice.

★ $\boxed{\max f(\boldsymbol{x}), \ \text{s.t. } \boldsymbol{x} \in \Omega} \iff \boxed{-\min -f(\boldsymbol{x}), \ \text{s.t. } \boldsymbol{x} \in \Omega}$

Definition (unconstrained optimization)

If $\Omega = \mathbb{R}^n$, then $\min f(\boldsymbol{x})$, s.t. $\boldsymbol{x} \in \mathbb{R}^n$ is unconstrained optimization.

★ representation of constraint Ω

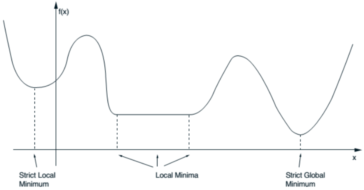
$$\begin{aligned}\Omega =& \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}_i(\boldsymbol{x}) = 0, i = 1, \cdots, m; \ \boldsymbol{g}_j(\boldsymbol{x}) \le 0, j = 1, \cdots, p\} \\ =& \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}(\boldsymbol{x}) = 0, \ \ \boldsymbol{g}(\boldsymbol{x}) \le 0\},\end{aligned}$$

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ (resp. $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$) is composed by all $\boldsymbol{h}_i$ (resp. $\boldsymbol{g}_i$).

Definition (minimizer of optimization $\min f(\boldsymbol{x})$, s.t. $\boldsymbol{x} \in \Omega$)

- $\boldsymbol{x}^* \in \Omega$ is a local minimizer if: $\exists \varepsilon > 0$ such that $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \Omega \backslash \{\boldsymbol{x}^*\}$ and $\|\boldsymbol{x} - \boldsymbol{x}^*\| < \varepsilon$.
- $\boldsymbol{x}^* \in \Omega$ is a global minimizer if: $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \Omega \backslash \{\boldsymbol{x}^*\}$.

★ If "$\geq$" in the above definitions are "$>$", then it is called strict local/global minimizer.



Definition (notation for minimizer of $\min f(\boldsymbol{x})$, s.t. $\boldsymbol{x} \in \Omega$)

$$\boldsymbol{x}^* \in \Omega \text{ is a global minimizer} \Longrightarrow \begin{cases} f(\boldsymbol{x}^*) = \min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x}), \\ \boldsymbol{x}^* = \operatorname{argmin}_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x}). \end{cases}$$

If $\Omega = \mathbb{R}^n$, then $\Longrightarrow f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}} f(\boldsymbol{x})$ or $\boldsymbol{x}^* = \operatorname{argmin}_{\boldsymbol{x}} f(\boldsymbol{x})$ for short.

Optimization Problem

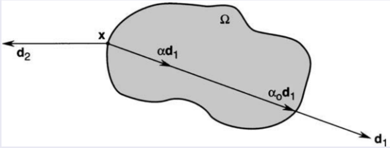
Example

$f(x) = (x+1)^2 + 3$, then $\text{argmin}_x f(x) = -1$ and $\text{argmin}_{x \geq 0} f(x) = 0$.

★ a minimizer may lie either in $\text{int}(\Omega)$ or $\text{bd}(\Omega)$.

Definition (feasible direction)

A vector $0 \neq \boldsymbol{d} \in \mathbb{R}^n$ is a feasible direction at $\boldsymbol{x} \in \Omega$ if: $\exists \alpha_0 > 0$ such that $\boldsymbol{x} + \alpha \boldsymbol{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$.



★ if $\boldsymbol{x} \in \text{int}(\Omega)$, all feasible directions at $\boldsymbol{x}$ is $\mathbb{R}^n$.

★ if $\boldsymbol{x} \in \text{bd}(\Omega)$, the feasible direction at $\boldsymbol{x}$ may be cone, hyperplane, halfspace, empty,...

Feasible Direction

Definition (linearized feasible direction of Ω)

If $\boldsymbol{x} \in \mathrm{bd}(\Omega)$, the linearized feasible direction of Ω at $\boldsymbol{x}$ is

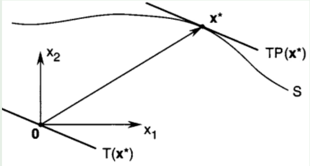
$$0 \neq \boldsymbol{d} = \lim_{\boldsymbol{x}' \to \boldsymbol{x}, \boldsymbol{x}' \in \Omega} (\boldsymbol{x}' - \boldsymbol{x}).$$

All linearized feasible direction of Ω at $\boldsymbol{x}$ is denoted by $T(\boldsymbol{x})$.

★ $T(\boldsymbol{x})$ is usually a cone in $\mathbb{R}^n$.

Example (equality constraint)

If $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{h}_i(\boldsymbol{x}) = 0, i = 1, \dots, m\}$ with all $\boldsymbol{h}_i : \mathbb{R}^n \to \mathbb{R}$ being differentiable, then $T(\boldsymbol{x}) = \{\boldsymbol{d} \in \mathbb{R}^n \mid \boldsymbol{D}\boldsymbol{h}_i(\boldsymbol{x})\boldsymbol{d} = 0, i = 1, \dots, m\}$ is a subspace.

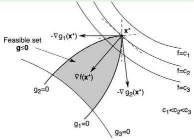


Question: if all $\boldsymbol{h}_i(\boldsymbol{x}) = \boldsymbol{a}_i^\top \boldsymbol{x} - b_i$, $T(\boldsymbol{x}) = ?$

Feasible Direction

Example (inequality constraint)

If $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{g}_j(\boldsymbol{x}) \le 0, j = 1, \dots, p\}$
with $\boldsymbol{g}_j : \mathbb{R}^n \to \mathbb{R}$ being differentiable, then
$T(\boldsymbol{x}) = \{\boldsymbol{d} \in \mathbb{R}^n \mid \boldsymbol{D}\boldsymbol{g}_j(\boldsymbol{x})\boldsymbol{d} \le 0, j \in J(\boldsymbol{x})\}$
is a cone, where $J(\boldsymbol{x}) = \{j \mid \boldsymbol{g}_j(\boldsymbol{x}) = 0, j = 1, \dots, p\}$.



Definition (directional derivative)

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{d}$ be a feasible direction at $\boldsymbol{x} \in \Omega$. The directional derivative of f w.r.t. $\boldsymbol{d}$ is defined by $\frac{\partial f}{\partial \boldsymbol{d}} = \lim_{\alpha \to 0} \frac{f(\boldsymbol{x}+\alpha\boldsymbol{d}) - f(\boldsymbol{x})}{\alpha}$.

- ★ if f is differentiable, then $\frac{\partial f}{\partial \boldsymbol{d}} = \nabla f(\boldsymbol{x})^\top \boldsymbol{d}$.
- ★ if $\boldsymbol{d}$ is a unit vector, then $\frac{\partial f}{\partial \boldsymbol{d}}$ is the increasing rate of $f(\boldsymbol{x})$ along direction $\boldsymbol{d}$.

$f(\boldsymbol{x}) = x_1 x_2 x_3$, and $\boldsymbol{d} = [\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}]^\top$. $\frac{\partial f}{\partial \boldsymbol{d}}$=?

Ans: $\because \quad \nabla f(\boldsymbol{x}) = [x_2 x_3, x_1 x_3, x_1 x_2] \Longrightarrow \frac{\partial f}{\partial \boldsymbol{d}} = \frac{x_2 x_3 + x_1 x_3 + \sqrt{2} x_1 x_2}{2}$.

First-Order Conditions for Local Minimizers

Theorem (first-order necessary condition, FONC)

Let $\Omega \subseteq \mathbb{R}^n$ be a set and $f \in \mathcal{C}^1(\Omega)$. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω, then $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0$ for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$.

proof. By the first-order Taylor expansion

$$f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{d} + o(\alpha).$$

$$\therefore \quad \nabla f(\boldsymbol{x}^*)^\top \boldsymbol{d} = \frac{f(\boldsymbol{x}^*+\alpha\boldsymbol{d}) - f(\boldsymbol{x}^*)}{\alpha} - \frac{o(\alpha)}{\alpha} \geq 0.$$

Corollary (particular case of FONC: interior case or $\Omega = \mathbb{R}^n$)

Let $f \in \mathcal{C}^1(\Omega)$ and $\boldsymbol{x}^*$ be a local minimizer. If $\Omega = \mathbb{R}^n$ or $\boldsymbol{x}^* \in \mathrm{int}(\Omega)$, then $\nabla f(\boldsymbol{x}^*) = 0$. (Hint of proof: $\boldsymbol{d}$ is all vectors in $\mathbb{R}^n$.)

Theorem (first-order sufficient condition, FOSC)

Let $\Omega \subseteq \mathbb{R}^n$ be a set and $f \in \mathcal{C}^1(\Omega)$. If $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) > 0$ for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$, then $\boldsymbol{x}^*$ is a strict local minimizer of f over Ω.

Second-Order Conditions for Local Minimizers

Theorem (second-order necessary condition, SONC)

Let $\Omega \subseteq \mathbb{R}^n$ *be a set and* $f \in \mathcal{C}^2(\Omega)$. $\boldsymbol{x}^*$ *is a local minimizer of* f *over* Ω *and* $\boldsymbol{d}$ *is a feasible direction. If* $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, *then* $\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0$.

proof. By the second-order Taylor expansion

$$\therefore\ f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha \underbrace{\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{d}}_{=0} + \frac{\alpha^2}{2}\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} + o(\alpha^2).$$

Corollary (particular case of SONC: $\Omega = \mathbb{R}^n$ or interior case)

Let $f \in \mathcal{C}^2(\Omega)$ *and* $\boldsymbol{x}^*$ *is a local minimizer. If* $\Omega = \mathbb{R}^n$ *or* $\boldsymbol{x}^* \in \mathrm{int}(\Omega)$, $\nabla f(\boldsymbol{x}^*) = 0$, *then* $\boldsymbol{F}(\boldsymbol{x}^*) \succeq 0$.

Conditions for Local Optima (I): $\Omega = \mathbb{R}$ or $x^* \in \text{int}(\Omega)$

$$\boxed{\min f(x), \quad \text{s.t. } x \in \mathbb{R}}$$

Optim Info: differentiable objective $f : \mathbb{R} \to \mathbb{R}$.

Condition:

$$\boxed{f'(x^*) \neq 0} \Longrightarrow \text{ neither min nor max}$$

$$\boxed{f'(x^*) = 0} \xRightarrow{f''(x^*)} \begin{cases} > 0 : \text{strict local min} \\ < 0 : \text{strict local max} \\ = 0 : \text{not sure} \end{cases}$$

Generalized condition

$$\boxed{\begin{array}{c} f^{(i)}(x^*) = 0 \\ i = 1, \cdots, k-1 \end{array}} \xRightarrow{f^{(k)}(x^*) \neq 0} \begin{cases} > 0 \text{ and } k \text{ is even : strict local min} \\ < 0 \text{ and } k \text{ is even : strict local max} \\ k \text{ is odd : neither min nor max} \end{cases}$$

Conditions for Local Optima (II): $\Omega = \mathbb{R}^n$ or $\boldsymbol{x}^* \in \text{int}(\Omega)$

Optim info: differentiable objective $f : \mathbb{R}^n \to \mathbb{R}$.

Condition:

$$\boxed{\nabla f(\boldsymbol{x}^*) \neq 0 \implies \text{neither min nor max}}$$

$$\boxed{\nabla f(\boldsymbol{x}^*) = 0} \xRightarrow{\text{eig}(F(\boldsymbol{x}^*))} \begin{cases} [+] : \text{strict local min} \\ [-] : \text{strict local max} \\ [+,-] \text{ or } [+,-,0] : \text{saddle} \\ [+,0] : \text{must not max} \\ [-,0] : \text{must not min} \end{cases}$$

Conditions for Local Optima (III): $\Omega \subset \mathbb{R}^n$

Optim info: objective f, constraint $\Omega \subset \mathbb{R}^n$, feasible directions $\mathcal{F}(\boldsymbol{x}^*)$[1].

First-order condition:

$$\boxed{\forall \boldsymbol{d} \in \mathcal{F}(\boldsymbol{x}^*), \boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \neq 0} \Longrightarrow \begin{cases} > 0 : \text{strict local min} \\ < 0 : \text{strict local max} \\ > 0 \ \& < 0 : \text{neither min nor max} \end{cases}$$

$$\boxed{\exists \boldsymbol{d} \in \mathcal{F}(\boldsymbol{x}^*), \boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0} \Longrightarrow \text{not sure}$$

Second-order condition: $\exists \boldsymbol{d} \in \mathcal{F}(\boldsymbol{x}^*), \boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$

$$\boxed{\forall \boldsymbol{d} \in \mathcal{F}(\boldsymbol{x}^*), \boldsymbol{d}^\top F(\boldsymbol{x}^*)\boldsymbol{d} \neq 0} \Longrightarrow \begin{cases} > 0 : \text{strict local min} \\ < 0 : \text{strict local max} \\ > 0 \ \& < 0 : \text{neither min nor max} \end{cases}$$

$$\boxed{\exists \boldsymbol{d} \in \mathcal{F}(\boldsymbol{x}^*), \boldsymbol{d}^\top F(\boldsymbol{x}^*)\boldsymbol{d} = 0} \Longrightarrow \boldsymbol{x}^* \text{ is not sure.}$$

[1] $\mathcal{F}(\boldsymbol{x}^*) \approx T(\boldsymbol{x}^*)$

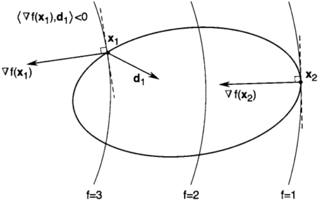
Conditions for Local Minimizers

Example ($\min_{x\in\mathbb{R}} f(x) = \frac{1+(2-x)^2}{1+x^2}$)

proof. $\because$ unconstrained optimization $\xRightarrow{FONC} 0 = f'(x) = \frac{4(x^2-2x-1)}{(1+x^2)^2}$

$\therefore\ x_1^* = 1-\sqrt{2}$ or $x_2^* = 1+\sqrt{2}$. $\because\ f''(x_1^*) = -8.24 < 0, f''(x_2^*) = 0.242.$

$\therefore\ x_1^*$ is strict local max, x_2^* is strict local min.



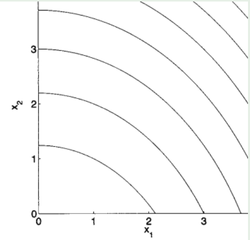
Example

Consider the optimization problem

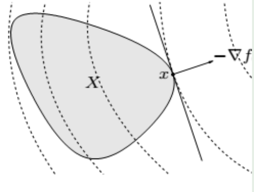
$$\min x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$$
$$\text{s.t. } x_1 \geq 0,\ x_2 \geq 0.$$

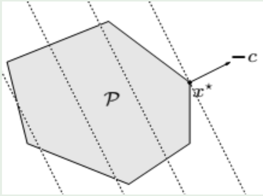
do the following points satisfy the first-order. necessary condition?

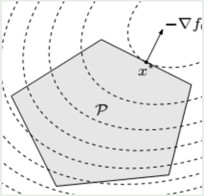
(A). $\boldsymbol{x} = [1,3]^\top$; (B). $\boldsymbol{x} = [0,3]^\top$;
(C). $\boldsymbol{x} = [1,0]^\top$; (D). $\boldsymbol{x} = [0,0]^\top$



Example (where is the minimizer?)







Conditions for Local Minimizers

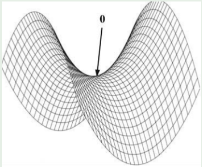
★ The necessary conditions are not sufficient.

Example ($f(x) = x^3$)

Because $f'(0) = 0$ and $f''(0) = 0$, $x^* = 0$ satisfies both 1st- and 2nd-necessary conditions. However, $x^* = 0$ is not a minimizer.

Example ($f(\boldsymbol{x}) = x_1^2 - x_2^2$)

$\because \nabla f(\boldsymbol{x}) = [2x_1, -2x_2]^\top = 0$, $\therefore \boldsymbol{x}^* = [0, 0]^\top$ satisfies 1st-order necessary condition. However, the Hessian $\boldsymbol{F}(\boldsymbol{x}^*) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \not\succeq 0$. Thus, $x^* = 0$ is not a minimizer.



Conditions for Local Minimizers

Theorem (second-order sufficient condition (interior case))

Let $\Omega \subseteq \mathbb{R}^n$ be a set and $f \in \mathcal{C}^2(\Omega)$. $\boldsymbol{x}^* \in \operatorname{int}(\Omega)$. If $\nabla f(\boldsymbol{x}^*) = 0$ and $\boldsymbol{F}(\boldsymbol{x}^*) \succ 0$, then $\boldsymbol{x}^*$ is a strict local minimizer.

proof. let $\boldsymbol{x}(\alpha) = \boldsymbol{x}^* + \alpha \boldsymbol{d}$ and $\varphi(\alpha) = f(\boldsymbol{x}^* + \alpha \boldsymbol{d})$.

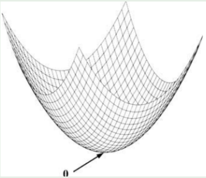
$\because\ 0 = \arg\min\limits_{\boldsymbol{x}\in\Omega} \varphi(\alpha)$,

$\therefore\ \varphi(\alpha) = \varphi(0) + \alpha\varphi'(0) + \frac{\alpha^2}{2}\varphi'(0) + o(\alpha^2)$,

$\therefore\ f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) = f(\boldsymbol{x}^*) + \alpha \underbrace{\nabla f(\boldsymbol{x}^*)^\top \boldsymbol{d}}_{=0} + \frac{\alpha^2}{2} \boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*) \boldsymbol{d} + o(\alpha^2)$.

Example ($f(\boldsymbol{x}) = x_1^2 + x_2^2$)

$\because\ \nabla f(\boldsymbol{x}) = [2x_1, 2x_2]^\top = 0$, $\therefore\ \boldsymbol{x}^* = [0,0]^\top$ satisfies 1st-order necessary condition. $\because$ the Hessian matrix $\boldsymbol{F}(\boldsymbol{x}^*) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \succ 0$. $\therefore\ \boldsymbol{x}^* = 0$ satisfies 2nd-order sufficient condition. $\therefore\ \boldsymbol{x}^* = 0$ is a minimizer.



Homework

Exercise in textbook: 6.2, 6.10, 6.16, 6.19