Introduction to Optimization

Part I

Unconstrained Optimization

Chia-Hui Chang (張嘉惠)

National Central University

2008.9.17

Optimization Problem

- Definition

$$\text{mimimize} \quad f(x)$$
$$\text{subject to} \quad x \in \Omega$$

- *Objective function: f : $R^n \to R$*
- *Decision variables: $x=[x_1, x_2, \ldots, x_n]^T \in R^n$*
- *Feasible set: $\Omega \subseteq R^n$*

- ***Local minimizer***
  - *If $\exists\, \varepsilon > 0$ such that $f(x) \geq f(x^*)$, $\forall x \in \Omega$ and $||x - x^*|| < \varepsilon$*

- ***Global minimizer***
  - *If $f(x) \geq f(x^*)\ \forall x \in \Omega$*

$$f(\mathbf{x}) \geq f(\mathbf{x}^*), \forall \mathbf{x} \in \Omega$$

Preliminary

- First-order derivative · Second-order derivative

    (Tangent vector)

$$Df \equiv \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right].$$

$$F(x) = D^2 f \equiv \begin{bmatrix} \frac{\partial f}{\partial x_1^2} & \cdots & \frac{\partial f}{\partial x_n \partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial f}{\partial x_n \partial x_n} \end{bmatrix}.$$

  - Gradient and Hessian matrix *F(x)*

$$\nabla f = (Df)^T$$

Example 6.1 Let $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$. Then,

$$Df(\boldsymbol{x}) = (\nabla f(\boldsymbol{x}))^T = \left[\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \frac{\partial f}{\partial x_2}(\boldsymbol{x})\right] = [5 + x_2 - 2x_1, 8 + x_1 - 4x_2],$$

and

$$\boldsymbol{F}(\boldsymbol{x}) = D^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix}.$$

Feasible Direction

- Definition 6.2
  - A vector ***d*** is a feasible direction at ***x*** if

$$\exists \alpha_0 \text{ such that } \mathbf{x} + \alpha \mathbf{d} \in \Omega \; \forall \alpha \in [0, \alpha_0]$$

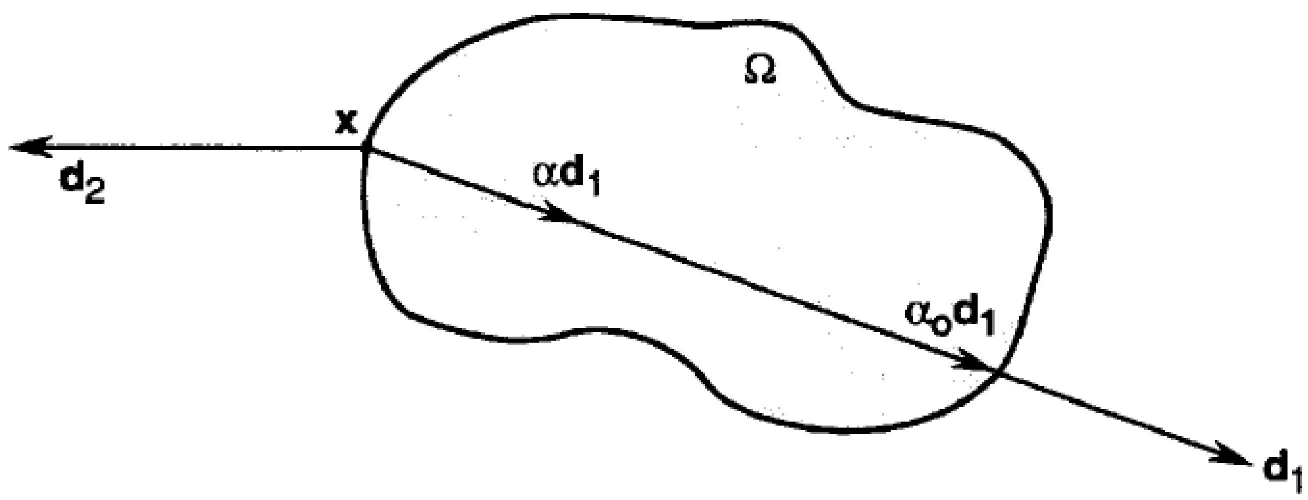

Figure 6.2 Two-dimensional illustration of feasible directions; $\boldsymbol{d}_1$ is a feasible direction, $\boldsymbol{d}_2$ is not a feasible direction

Directional Derivative

- Directional derivative
  - Let $f : R^n \to R$ be a real-valued function and let **d** be a feasible direction at $x \in \Omega$. The direction derivative of f in the direction **d**, denoted $\partial f / \partial \mathrm{d}$, is the real-valued function defined by

$$\frac{\partial f}{\partial \mathbf{d}} = \lim_{\alpha \to 0} \frac{f(x + \alpha \mathbf{d}) - f(x)}{\alpha}.$$

$$\frac{\partial f}{\partial \mathbf{d}} = \frac{d}{d\alpha} f(x + \alpha \mathbf{d})\Big|_{\alpha=0} = Df(x + \alpha \mathbf{d})\mathbf{d}\Big|_{\alpha=0}.$$

$$\frac{\partial f}{\partial \mathbf{d}} = \nabla f(x)^T \mathbf{d}.$$

Speed

Example 6.2 Define $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(\boldsymbol{x}) = x_1 x_2 x_3$, and let

$$\boldsymbol{d} = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right]^T.$$

The directional derivative of f in the direction $\boldsymbol{d}$ is

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})^T \boldsymbol{d} = [x_2 x_3, x_1 x_3, x_1 x_2] \begin{bmatrix} 1/2 \\ 1/2 \\ 1/\sqrt{2} \end{bmatrix} = \frac{x_2 x_3 + x_1 x_3 + \sqrt{2} x_1 x_2}{2}.$$

Note that because $\|\boldsymbol{d}\| = 1$, the above is also the rate of increase of f at $\boldsymbol{x}$ in the direction $\boldsymbol{d}$. ■

First-Order Necessary Condition

- Theorem 6.1
  - If ***x**** is a local minimizer of f over Ω, then for any feasible direction ***d*** at ***x****, we have $\mathbf{d}^T \nabla f \geq 0$

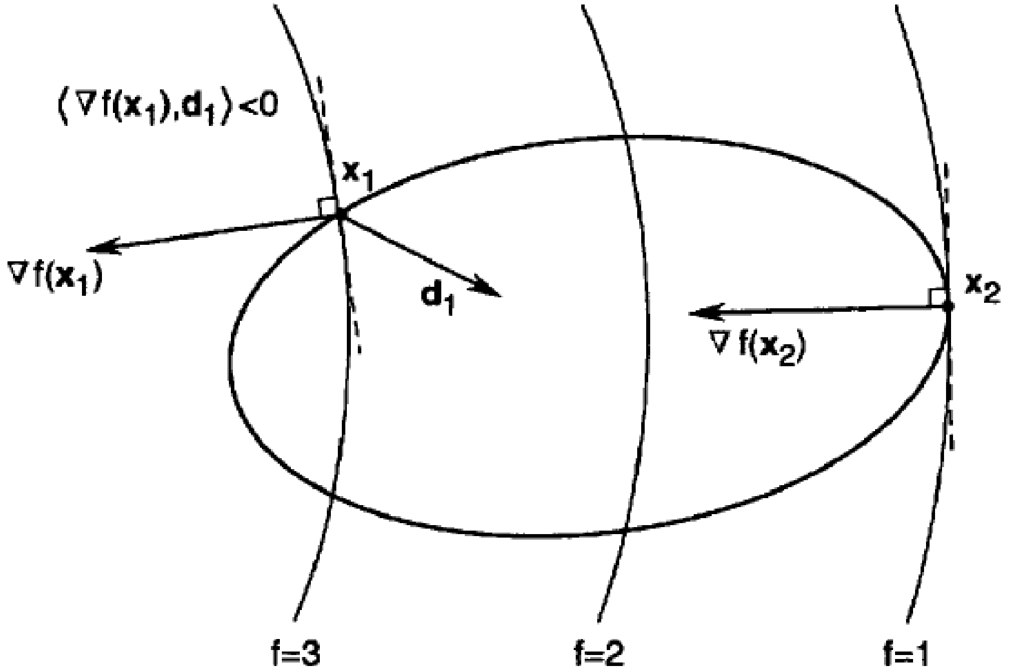

Figure 6.3 Illustration of the FONC for the constrained case; x_1 does not satisfy the FONC, x_2 satisfies the FONC

First-Order Necessary Condition (Cont.)

- Corollary 6.1 Interior Case
  - If ***x**** is a local minimizer of f over Ω, and if ***x**** is an interior point of Ω, we have $\nabla f(\mathbf{x}^*) = 0$

- Note: Necessary but not sufficient

Example 6.3 Consider the problem

$$\begin{aligned} \text{minimize} \quad & x_1^2 + 0.5x_2^2 + 3x_2 + 4.5 \\ \text{subject to} \quad & x_1, x_2 \geq 0. \end{aligned}$$

Questions:

a. Is the first-order necessary condition (FONC) for a local minimizer satisfied at $\boldsymbol{x} = [1, 3]^T$?

b. Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = [0, 3]^T$?

c. Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = [1, 0]^T$?

d. Is the FONC for a local minimizer satisfied at $\boldsymbol{x} = [0, 0]^T$?

Second-Order Necessary Condition

- Theorem 6.2
  - Let ***x**** be a local minimizer of f over *Ω*, and ***d*** is a feasible direction at ***x****. We have

$$\text{If } \mathbf{d}^T \nabla f = 0, \text{ then } \mathbf{d}^T F(\mathbf{x}^*)\mathbf{d} \geq 0.$$

Acceleration

- Corollary 6.2 ***Interior Case***
  - Let ***x**** be an interior point of *Ω*. If ***x**** is an local minimizer of *Ω*, then

$$\nabla f(\mathbf{x}^*) = 0, \text{ and Hession of } f(x) \text{ is semidefinite;}$$

$$\text{i.e. } \forall \mathbf{d} \in \mathrm{R}^n, \mathbf{d}^T F(\mathbf{x}^*)\mathbf{d} \geq 0.$$

Example:

Example 6.5 Consider a function of one variable $f(x) = x^3$, $f : \mathbb{R} \to \mathbb{R}$. Because $f'(0) = 0$, and $f''(0) = 0$, the point $x = 0$ satisfies both the FONC and SONC. However, $x = 0$ is not a minimizer (see Figure 6.6). ∎

Example 6.6 Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(\boldsymbol{x}) = x_1^2 - x_2^2$. The FONC requires that $\nabla f(\boldsymbol{x}) = [2x_1, -2x_2]^T = \boldsymbol{0}$. Thus, $\boldsymbol{x} = [0, 0]^T$ satisfies the FONC. The Hessian matrix of f is

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}.$$

Indefinite

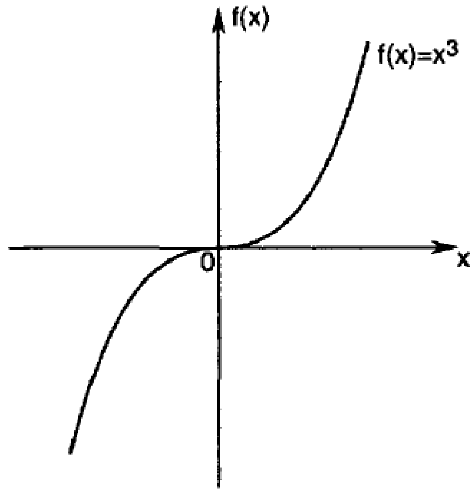

Figure 6.6 The point 0 satisfies the FONC and SONC, but is not a minimizer

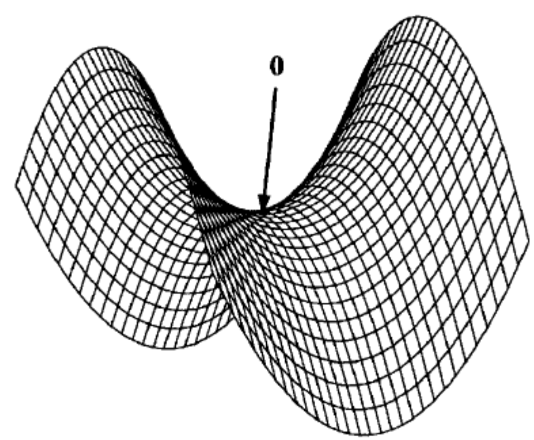

Figure 6.7 Graph of $f(\boldsymbol{x}) = x_1^2 - x_2^2$. The point $\boldsymbol{0}$ satisfies the FONC but not SONC; this point is not a minimizer

Second-order Sufficient Condition

- Theorem 6.3 Interior Case.
  Suppose that $\nabla f(\mathbf{x}^*) = 0.$

  $$F(\mathbf{x}^*) > 0.$$

  Then ***x****** is a strict local minimizer of *f*.

Proof of SOSC

Taylor's theorem.

f: R $\to$R $\quad f(b) = f(a) + \frac{h}{1!} f^{(1)}(a) + \frac{h^2}{2!} f^{(2)}(a) + \ldots + \frac{h^{m-1}}{(m-1)!} f^{(m-1)}(a) + o(h^m).$

f : $R^n \to$R $\quad f(x) = f(x_0) + \frac{1}{1!} Df(x_0)(x - x_0) + \frac{1}{2!}(x - x_0)^T D^2 f(x_0)(x - x_0) + o(\|x - x_0\|^2).$

Rayleigh's Inequality. If an $n \times n$ matrix $\boldsymbol{P}$ is real symmetric positive definite, then

$$\lambda_{\min}(\boldsymbol{P})\|\boldsymbol{x}\|^2 \leq \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} \leq \lambda_{\max}(\boldsymbol{P})\|\boldsymbol{x}\|^2,$$

where $\lambda_{\min}(\boldsymbol{P})$ denotes the smallest eigenvalue of $\boldsymbol{P}$, and $\lambda_{\max}(\boldsymbol{P})$ denotes the largest eigenvalue of $\boldsymbol{P}$.

Example

Example 6.7 Let $f(\boldsymbol{x}) = x_1^2 + x_2^2$. We have $\nabla f(\boldsymbol{x}) = [2x_1, 2x_2]^T = \mathbf{0}$ if and only if $\boldsymbol{x} = [0, 0]^T$. For all $\boldsymbol{x} \in \mathbb{R}^2$, we have

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0.$$

The point $\boldsymbol{x} = [0, 0]^T$ satisfies the FONC, SONC, and SOSC. It is a strict local minimizer. Actually $\boldsymbol{x} = [0, 0]^T$ is a strict global minimizer. Figure 6.8 shows the graph of $f(\boldsymbol{x}) = x_1^2 + x_2^2$.

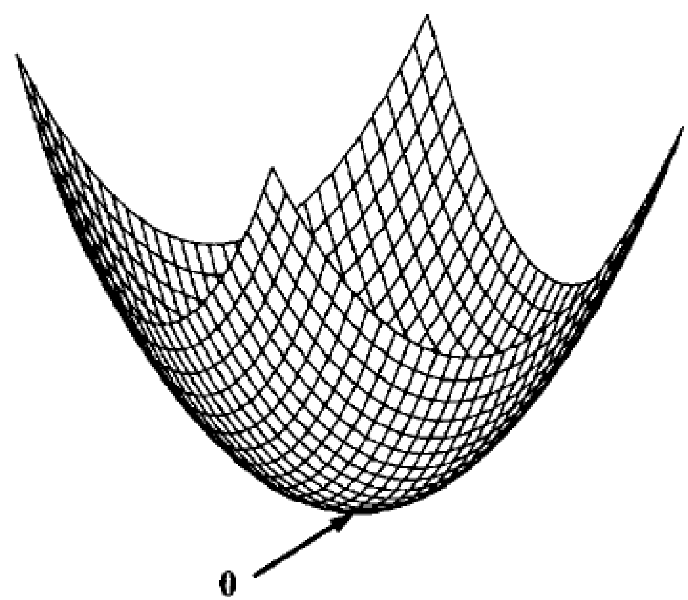

Figure 6.8 Graph of $f(\boldsymbol{x}) = x_1^2 + x_2^2$