Ch.6 Basics of Set-Constrained and Unconstrained Optimization

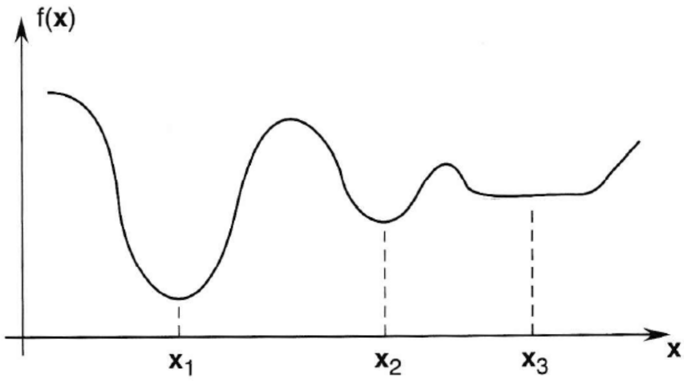
6.1 Introduction

- In this chapter, we consider the optimization problem

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega \end{array}$$

- The function $f : \Re^n \to \Re$ that we wish to minimize is a real-valued function, and is called the objective function, or cost function.
- The variables $x_1, \cdots, x_n$ are referred to as decision variables.
- The set Ω is a subset of $\Re^n$, called the constraint or feasible set.
- There are also optimization problems that require maximization of the objective function. But maximizing f is equivalent to minimizing $-f$.
- Because the decision variables are constrained to be in the constraint set Ω, the above one is a general form of a constrained optimization problem.

- If $\Omega = \Re^n$, then the problem is an unconstrained optimization problem.
- The constraint $x \in \Omega$ is called a set constraint.
- **Definition 6.1 Local minimizer** *Suppose that $f : \Re^n \to \Re$ is a real-valued function defined on some set $\Omega \subset \Re^n$. A point $x^* \in \Omega$ is a local minimizer of f over Ω if there exists $\epsilon > 0$ such that $f(x) \geq f(x^*)$ for all $x \in \Omega \backslash \{x^*\}$ and $\|x - x^*\| < \epsilon$.*
- Global minimizer: A point $x^* \in \Omega$ is a global minimizer of f over Ω if $f(x) \geq f(x^*)$ for all $x \in \Omega \backslash \{x^*\}$.
- If we replace "$\geq$" with "$>$", then we have a strict local minimizer and a strict global minimizer.
  ♣ Fig. 6.1 on page 74.



6.2 Conditions for Local Minimizers

- The first-order derivative of f, denoted Df, is

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

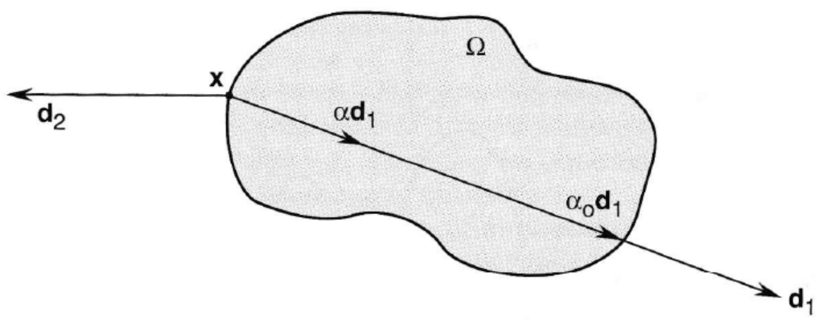
- Gradient ∇f is just the transpose of Df: $\nabla f = (Df)^T$.
- The second derivative of $f : \Re^n \to \Re$ (also called the Hessian of f) is

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

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

$$\begin{aligned} Df(x) &= (\nabla f(x))^T = \left[\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x)\right] \\ &= [5 + x_2 - 2x_1,\ 8 + x_1 - 4x_2] \\ F(x) &= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix} \end{aligned}$$

- **Definition 6.2 Feasible direction** *A vector $d \in \Re^n$, $d \neq 0$, is a feasible direction as $x \in \Omega$ if there exists $\alpha_0 > 0$ such that $x + \alpha d \in \Omega$ for all $\alpha \in [0, \alpha_0]$.*
♣ Fig. 6.2 on page 76.



- Directional derivative of f in the direction d, denoted $\partial f/\partial d$, is

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

- If $\|d\| = 1$, then $\partial f/\partial d$ is the rate of increase of f at x in the direction d.

- Applying the chain rule yields,

$$\begin{aligned} \frac{\partial f}{\partial d}(x) &= \left. \frac{d}{d\alpha} f(x+\alpha d) \right|_{\alpha=0} = \nabla f(x)^T d \\ &= < \nabla f(x), d >= d^T \nabla f(x) \end{aligned}$$

- Example 6.2 Define $f : \Re^3 \to \Re$ by $f(x) = x_1 x_2 x_3$,

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

  The directional derivative of f in the direction d is

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

- **Theorem 6.1 First-Order Necessary Condition(FONC)** *Let Ω be a subset of $\Re^n$ and $f \in C^1$ a real-valued function on Ω. If x^* is a local* *minimizer of f over Ω, then for any feasible direction d at x^*, we have*

$$d^T \nabla f(x^*) \geq 0$$

- If x^* is a local minimizer, then the rate of increase of f at x^* in any feasible direction d in Ω is nonnegative.
- **Corollary 6.1 Interior case** *Let Ω be a subset of $\Re^n$ and $f \in C^1$ a real-valued function on Ω. If x^* is a local minimizer of f over Ω and if x^* is an interior point of Ω, then*

$$\nabla f(x^*) = 0$$

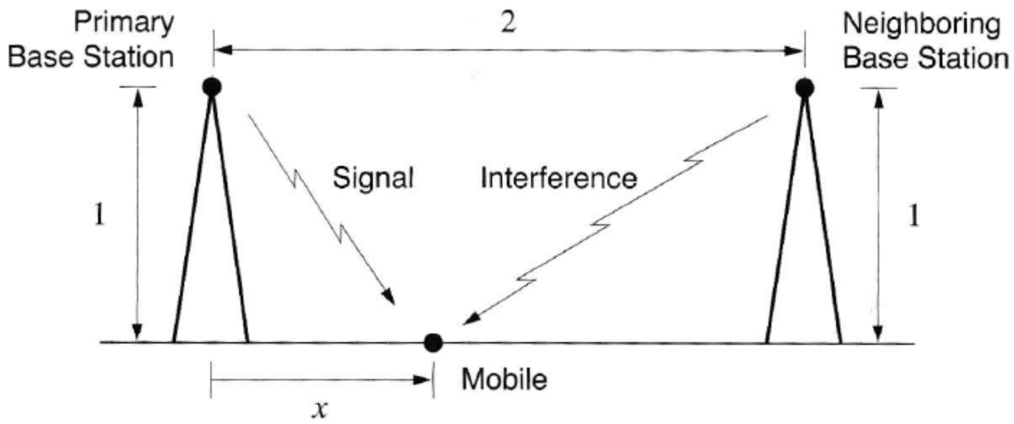
- Example 6.3 Consider the problem

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

(1) Is the first-order necessary condition (FONC) for a local minimizer satisfied at $x = [1,3]^T$? (2) $x = [0,3]^T$? (3) $x = [1,0]^T$? (4) $x = [0,0]^T$?
Sol. (1) We have $\nabla f(x) = [2x_1, x_2 + 3]^T = [2,6]^T$. The point $x = [1,3]^T$ is an interior point of $\Omega = \{x : x_1, x_2 \geq 0\}$. Hence, the FONC requires $\nabla f(x) = 0$. The point $x = [1,3]^T$ does not satisfy the FONC for a local minimizer.
(2) At $x = [0,3]^T$, we have $\nabla f(x) = [0,6]^T$, and hence

$d^T \nabla f(x) = 6d_2$. For d to be feasible at x, we need $d_1 \geq 0$, and d_2 can take an arbitrary value in $\Re$. For example $d = [1, -1]^T$ is a feasible direction, but $d^T \nabla f(x) = -6 < 0$.
(3) At $x = [1, 0]^T$, we have $\nabla f(x) = [2, 3]^T$, and hence $d^T \nabla f(x) = 2d_1 + 3d_2$. For d to be feasible, we need $d_2 \geq 0$, and d_1 can take an arbitrary value in $\Re$. For example, $d = [-5, 1]^T$, $d^T \nabla f(x) = -7 < 0$.
(4) At $x = [0, 0]^T$, we have $\nabla f(x) = [0, 3]^T$, and hence $d^T \nabla f(x) = 3d_2$. For d to be feasible, we need $d_2 \geq 0$ and $d_1 \geq 0$. $x = [0, 0]^T$ satisfies the FONC for a local minimizer.

- Example 6.4 A simplified model of a celluar wireless system.
  ♣ Fig. 6.5 on page 80.



We are interested in finding the position of the mobile that maximizes the signal-to-interference ratio, which is the ratio of the received signal power from the primary

basestation to the received signal power from the neighboring basestation.
The signal-to-interference ratio is

$$\begin{aligned} f(x) &= \frac{1+x^2}{1+(2-x)^2} \\ f'(x) &= \frac{4(x^2-2x-1)}{1+(2-x)^2} \end{aligned}$$

By the FONC, at the optimal position x^*, we have $f'(x^*) = 0$. Hence, either $x^* = 1-\sqrt{2}$ or $x^* = 1+\sqrt{2}$. Finally, $x^* = 1-\sqrt{2}$ is the optimal position.

- **Theorem 6.2 Second-Order Necessary Condition(SONC)** *Let $\Omega \subset \Re^n$, $f \in C^2$ a function on Ω, x^* a local minimizer of f over Ω, and d a feasible direction at x^*. If $d^T \nabla f(x^*) = 0$, then*

$$d^T F(x^*) d \geq 0,$$

where F is the Hessian of f.

- **Corollary 6.2 Interior case** *Let x^* be an interior point of $\Omega \subset \Re^n$. If x^* is a local minimizer of $f : \Omega \to \Re$, $f \in C^2$, then*

$$\nabla f(x^*) = 0$$

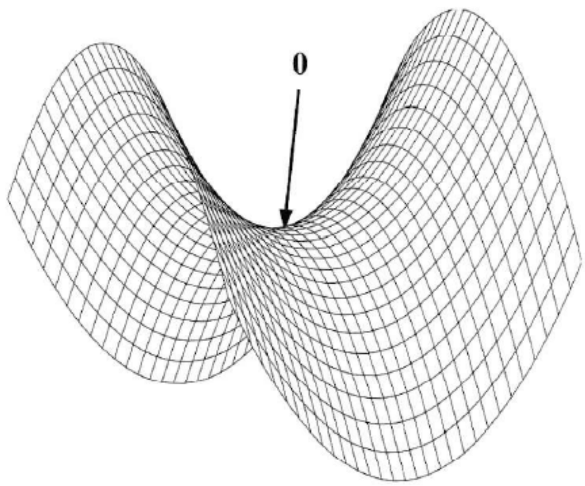
and $F(x^*)$ is positive semidefinite ($F(x^*) \geq 0$); that is, for all $d \in \Re^n$,

$$d^T F(x^*) d \geq 0.$$

- Example 6.7 Consider a function $f : \Re^2 \to \Re$, where $f(x) = x_1^2 - x_2^2$. The FONC requires that $\nabla f(x) = [2x_1, -2x_2]^T = 0$. Thus, $x = [0, 0]^T$ satisfies the FONC. The Hessian matrix of f is

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

  The Hessian matrix is indefinite. Thus $x = [0, 0]^T$ does not satisfy the SONC, and hence it is not a minimizer. ♣ Fig. 6.7 on page 82.



- **Theorem 6.3 SOSC, Interior case** *Let $f \in C^2$ be defined on a region in which x^* is an interior point. Suppose that*

  1. $\nabla f(x^*) = 0$: and

  2. $F(x^*) > 0$.

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

- Example 6.8 Let $f(x) = x_1^2 + x_2^2$. We have $\nabla f(x) = [2x_1, 2x_2]^T = 0$ if and only if $x = [0, 0]^T$. For all $x \in \Re^2$, we have

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

The point $x = [0, 0]^T$ satisfies the FONC, SONC. It is a strict local minimizer.

♣ Fig. 6.8 on page 83.