Chong, Edwin Kah Pin., and Stanislaw H. Żak. *An Introduction to Optimization*. Fourth edition., John Wiley & Sons, Inc., 2013.

1 BASICS OF SET CONTSTRAINED AND UNCONSTRAINED OPTIMIZATION

1.1 INTRODUCTION

- Consider the **optimization problem**

$$\begin{gathered}\text{minimize } f(\boldsymbol{x}) \\ \text{subject to } \boldsymbol{x} \in \Omega.\end{gathered}$$

- The function $f: \mathbb{R}^n \to \mathbb{R}$ that we wish to minimize is a real-valued function called the **objective function**.

- The vector $\boldsymbol{x}$ is an n-vector of **independent variables**: $\boldsymbol{x} = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$.

- The set Ω is a subset of $\mathbb{R}^n$ called the **constraint set** or **feasible set**, which takes the form $\Omega = \{\boldsymbol{x}: \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$, where $\boldsymbol{h}$ and $\boldsymbol{g}$ are given functions.

- The **minimizer** of f over Ω is a vector $\boldsymbol{x}$ which results in the smallest value of the objective function.

- **Definition 1.1.1** Suppose that $f: \mathbb{R}^n \to \mathbb{R}$ is a real-valued function defined on some set $\Omega \subset \mathbb{R}^n$. A point $\boldsymbol{x}^* \in \Omega$ is a **local minimizer** of f over Ω if there exists $\varepsilon > 0$ such that $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \Omega \setminus \{\boldsymbol{x}^*\}$ and $\|\boldsymbol{x} - \boldsymbol{x}^*\| < \varepsilon$. A point $\boldsymbol{x}^* \in \Omega$ is a **global minimizer** of f over Ω if $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$ for all $\boldsymbol{x} \in \Omega \setminus \{\boldsymbol{x}^*\}$.

- If $\boldsymbol{x}^*$ is a global minimizer of f over Ω, we write $f(\boldsymbol{x}^*) = \min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x})$ and $\boldsymbol{x}^* = \operatorname{argmin}_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x})$.

Example 1.1.1. Suppose $f: \mathbb{R} \to \mathbb{R}$ is given by $f(x) = (x+1)^2 + 3$. Find $\arg\min_{x \in \Omega} f(x)$ where $\Omega = \{x: x \geq 0\}$.

$$x^* = \arg\min_{x \in \Omega} f(x) = 0.$$

Note: If $\Omega = \mathbb{R}$, then $x^* = \arg\min_{x \in \Omega} f(x) = -1$.

1.2 CONDITIONS FOR LOCAL MINIMIZERS

- **Global** minimizers are, in general, **difficult** to find. Therefore, in practice, we often have to be **satisfied** with finding local minimizers.

- The **first-order derivative** of $f: \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}$, denoted $Df(\boldsymbol{x})$, is

$$Df(\boldsymbol{x}) \triangleq \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \dots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right].$$

- The **gradient** $\nabla f(\boldsymbol{x})$ is the transpose of $Df(\boldsymbol{x})$; that is, $\nabla f(\boldsymbol{x}) = \left(Df(\boldsymbol{x})\right)^{\top}$.
- The **second derivative** of $f: \mathbb{R}^n \to \mathbb{R}$ (also called the Hessian of f)) is

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

Example 1.2.1. Let $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$. Find $Df(\boldsymbol{x})$ and $D^2 f(\boldsymbol{x})$.

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

and

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

- Given an optimization problem with constraint set Ω, a minimizer may lie either in the **interior** or on the **boundary** of Ω. To study the case where it lies on the boundary, we need the notion of **feasible directions**.
- **Definition 1.2.1** A vector $\boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{d} \neq \boldsymbol{0}$, is a **feasible direction** at $\boldsymbol{x} \in \Omega$ if there exists $\alpha_0 > 0$ such that $\boldsymbol{x} + \alpha\boldsymbol{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$.
- Let $f: \mathbb{R}^n \to \mathbb{R}$ be a real-valued function and let $\boldsymbol{d}$ be a feasible direction at $\boldsymbol{x} \in \Omega$. The **directional derivative** of f in the direction $\boldsymbol{d}$, denoted $\partial f / \partial \boldsymbol{d}$, is the real-valued function defined by

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{d}} = \lim_{\alpha \to 0} \frac{f(\boldsymbol{x} + \alpha\boldsymbol{d}) - f(\boldsymbol{x})}{\alpha} = \langle \nabla f(\boldsymbol{x}), \boldsymbol{d} \rangle = \boldsymbol{d}^{\top} \nabla f(\boldsymbol{x}).$$

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

Example 1.2.2. Define $f: \mathbb{R}^3 \to \mathbb{R}$ by $f(\boldsymbol{x}) = x_1x_2x_3$, and let $\boldsymbol{d} = \left[1/2, 1/2, 1/\sqrt{2}\right]^{\top}$. Find the directional derivative of f in the direction $\boldsymbol{d}$.

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{d}} = \nabla f(\boldsymbol{x})^\top \boldsymbol{d} = [x_2x_3, x_1x_3, x_1x_2] \begin{bmatrix} 1/2 \\ 1/2 \\ 1/\sqrt{2} \end{bmatrix} = \frac{x_2x_3 + x_1x_3 + \sqrt{2}x_1x_2}{2}.$$

Note that $\|\boldsymbol{d}\| = \sqrt{(1/2)^2 + (1/2)^2 + \left(1/\sqrt{2}\right)^2} = 1$. The above is also the rate of increase of f at $\boldsymbol{x}$ in the direction $\boldsymbol{d}$.

- **Theorem 1.2.2 First-Order Necessary Condition (FONC).** Let Ω be a subset of $\mathbb{R}^n$ and $f \in C^1$ (i.e., the first derivative exists and is continuous) a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$, we have

$$\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0.$$

- **Corollary 1.2.3 Interior Case.** Let Ω be a subset of $\mathbb{R}^n$ and $f \in C^1$ a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω and if $\boldsymbol{x}^*$ is an **interior point** of Ω, then

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

Example 1.2.3. Consider the problem

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

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

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

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

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

(a) At $\boldsymbol{x} = [1,3]^\top$, we have $\nabla f(\boldsymbol{x}) = [2x_1, x_2 + 3]^\top = [2,6]^\top$. The point $\boldsymbol{x} = [1,3]^\top$ is an interior point of $\Omega = \{\boldsymbol{x}: x_1 \geq 0, x_2 \geq 0\}$. Hence, the FONC requires that $\nabla f(\boldsymbol{x}) = 0$.

The point $\boldsymbol{x} = [1,3]^\top$ does not satisfy the FONC for a local minimizer.

(b) At $\boldsymbol{x} = [0,3]^\top$, we have $\nabla f(\boldsymbol{x}) = [2x_1, x_2 + 3]^\top = [0,6]^\top$, and hence $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}) = 6d_2$, where $\boldsymbol{d} = [d_1, d_2]^\top$. For $\boldsymbol{d}$ to be feasible at $\boldsymbol{x}$, we need $d_1 \geq 0$ and d_2 can take an arbitrary value in $\mathbb{R}$.

The point $\boldsymbol{x} = [0,3]^\top$ does not satisfy the FONC for a minimizer because d_2 is allowed to be less than zero. For example, $\boldsymbol{d} = [1,-1]^\top$ is a feasible direction, but $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}) = -6 < 0$.

(c) At $\boldsymbol{x} = [1,0]^\top$, we have $\nabla f(\boldsymbol{x}) = [2x_1, x_2 + 3]^\top = [2,3]^\top$, and hence $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}) = 2d_1 + 3d_2$, where $\boldsymbol{d} = [d_1, d_2]^\top$. For $\boldsymbol{d}$ to be feasible at $\boldsymbol{x}$, we need $d_2 \geq 0$ and d_1 can take an arbitrary value in $\mathbb{R}$.

The point $\boldsymbol{x} = [1,0]^\top$ does not satisfy the FONC for a minimizer because d_1 is allowed to be less than zero. For example, $\boldsymbol{d} = [-5,1]^\top$ is a feasible direction, but $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}) = -7 < 0$.

(d) At $\boldsymbol{x} = [0,0]^\top$, we have $\nabla f(\boldsymbol{x}) = [2x_1, x_2 + 3]^\top = [0,3]^\top$, and hence $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}) = 3d_2$, where $\boldsymbol{d} = [d_1, d_2]^\top$. For $\boldsymbol{d}$ to be feasible at $\boldsymbol{x}$, we need $d_1 \geq 0$ and $d_2 \geq 0$. The point $\boldsymbol{x} = [0,0]^\top$ satisfies the FONC because $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}) = 3d_2 \geq 0$.

- **Theorem 1.2.4 Second-Order Necessary Condition (SONC).** Let $\Omega \subset \mathbb{R}^n$, $f \in C^2$ a function on Ω, $\boldsymbol{x}^*$ a local minimizer of f over Ω, and $\boldsymbol{d}$ a feasible direction at $\boldsymbol{x}^*$. If $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, then
$$\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*) \boldsymbol{d} \geq 0,$$
where $\boldsymbol{F}$ is the Hessian of f.

- **Corollary 1.2.5 Interior Case.** Let $\boldsymbol{x}^*$ be an interior point of $\Omega \subset \mathbb{R}^n$. If $\boldsymbol{x}^*$ is a local minimizer of $f: \Omega \to \mathbb{R}$, $f \in C^2$, then $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$, and $\boldsymbol{F}(\boldsymbol{x}^*)$ is positive semidefinite ($\boldsymbol{F}(\boldsymbol{x}^*) \succcurlyeq 0$); that is, for all $\boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*) \boldsymbol{d} \geq 0$.

Example 1.2.4 Consider a function of one variable $f(x) = x^3$, $f: \mathbb{R} \to \mathbb{R}$. Does the point $x = 0$ satisfy both the FONC and SONC?

$x = 0$ is an interior point. Then, $\nabla f(x^*) = 3x^2$ and $F(x^*) = 6x$. Hence, both the FONC and SONC are satisfied: $\nabla f(0) = 0$ and $F(0) = 0$.

Example 1.2.5 Consider a function $f: \mathbb{R}^2 \to \mathbb{R}$, where $f(\boldsymbol{x}) = x_1^2 - x_2^2$. Does the point $\boldsymbol{x} = [0,0]^\top$ satisfy both the FONC and SONC?

The FONC requires that $\nabla f(\boldsymbol{x}) = [2x_1, -2x_2]^\top = 0$. Thus, $\boldsymbol{x} = [0,0]^\top$ satisfies the FONC.

The Hessian matrix of f is

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

The Hessian matrix is indefinite; that is, for some $\boldsymbol{d}_1 \in \mathbb{R}^2$ we have $\boldsymbol{d}_1^\top \boldsymbol{F} \boldsymbol{d}_1 > 0$ (e.g., $\boldsymbol{d}_1 = [1,0]^\top$); and, for some $\boldsymbol{d}_2$, we have $\boldsymbol{d}_2^\top \boldsymbol{F} \boldsymbol{d}_2 < 0$ (e.g., $\boldsymbol{d}_2 = [0,1]^\top$). Thus, $\boldsymbol{x} = [0,0]^\top$ does not satisfy the SONC, and hence it is not a minimizer.

- **Theorem 1.2.6 Second-Order Sufficient Condition (SOSC), Interior Case.** Let $f \in C^2$ be defined on a region in which $\boldsymbol{x}^*$ is an interior point. Suppose that

  1. $\nabla f(\boldsymbol{x}^*) = 0$.
  2. $\boldsymbol{F}(\boldsymbol{x}^*) \succ 0$.

  Then, $\boldsymbol{x}^*$ is a strict local minimizer of f.

Example 1.2.6 Consider a function $f: \mathbb{R}^2 \to \mathbb{R}$, where $f(\boldsymbol{x}) = x_1^2 + x_2^2$. Does the point $\boldsymbol{x} = [0, 0]^\top$ satisfy the FONC, SONC, and SOSC?

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

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

satisfies the FONC, SONC, and SOSC.

APPENDEX

- **Theorem 1.2.2 First-Order Necessary Condition (FONC).** Let Ω be a subset of $\mathbb{R}^n$ and $f \in C^1$ (i.e., the first derivative exists and is continuous) a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$, we have
$$\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0.$$

  Proof. Define $\boldsymbol{x}(\alpha) = \boldsymbol{x}^* + \alpha \boldsymbol{d} \in \Omega$. Note that $x(0) = x^*$. Define the composite function
$$\phi(\alpha) = f(\boldsymbol{x}(\alpha)).$$
  Then, by Taylor's theorem,
$$f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) - f(\boldsymbol{x}^*) = \phi(\alpha) - \phi(0) = \phi'(0)\alpha + o(\alpha) = \alpha \boldsymbol{d}^\top \nabla f(\boldsymbol{x}(0)) + o(\alpha),$$
  where $\alpha \geq 0$ and $o(\alpha)$ means that $\lim_{\alpha \to 0} o(\alpha)/\alpha = 0$. Thus, if $\phi(\alpha) \geq \phi(0)$, that is, $f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) \geq f(\boldsymbol{x}^*)$ for sufficiently small values of $\alpha > 0$, then we have to have $d^\top \nabla f(\boldsymbol{x}(0)) \geq 0$.

- **Corollary 1.2.3 Interior Case.** Let Ω be a subset of $\mathbb{R}^n$ and $f \in C^1$ a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω and if $\boldsymbol{x}^*$ is an interior point of Ω, then
$$\nabla f(\boldsymbol{x}^*) = 0.$$

  Proof. Suppose that f has a local minimizer $\boldsymbol{x}^*$ that is an interior point of Ω. Because $\boldsymbol{x}^*$ is an interior point of Ω, the set of feasible directions at $\boldsymbol{x}^*$ is the whole of $\mathbb{R}^n$. Thus, for any $\boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0$ and $-\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \geq 0$. Hence, $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$ for all $d \in \mathbb{R}^n$, which implies that $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$.

- **Theorem 1.2.4 Second-Order Necessary Condition (SONC).** Let $\Omega \subset \mathbb{R}^n$, $f \in C^2$ a function on Ω, $\boldsymbol{x}^*$ a local minimizer of f over Ω, and $\boldsymbol{d}$ a feasible direction at $\boldsymbol{x}^*$. If $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$, then

$$\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0,$$

where $\boldsymbol{F}$ is the Hessian of f.

Proof. We prove the result by contradiction. Suppose there is a feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$ such that $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) = 0$ and $\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} < 0$. Let $\boldsymbol{x}(\alpha) = \boldsymbol{x}^* + \alpha\boldsymbol{d}$ and define the composite function $\phi(\alpha) = f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) = f\big(\boldsymbol{x}(\alpha)\big)$. Then, by Taylor's theorem,

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

where, by assumption, $\phi'(0) = \boldsymbol{d}^\top \nabla \boldsymbol{F}(\boldsymbol{x}^*) = 0$ and $\phi''(0) = \boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} < 0$. For sufficiently small α,

$$\phi(\alpha) - \phi(0) = \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2) < 0,$$

that is,

$$f(\boldsymbol{x}^* + \alpha\boldsymbol{d}) < f(\boldsymbol{x}^*),$$

which contradicts the assumption that $\boldsymbol{x}^*$ is a local minimizer. Thus,

$$\phi''(0) = \boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0.$$

- **Corollary 1.2.5 Interior Case.** Let x^* be an interior point of $\Omega \subset \mathbb{R}^n$. If $\boldsymbol{x}^*$ is a local minimizer of $f: \Omega \to \mathbb{R}$, $f \in C^2$, then

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

and $\boldsymbol{F}(\boldsymbol{x}^*)$ is positive semidefinite ($\boldsymbol{F}(\boldsymbol{x}^*) \succcurlyeq 0$); that is, for all $\boldsymbol{d} \in \mathbb{R}^n$,

$$\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} \geq 0.$$

Proof. If $\boldsymbol{x}^*$ is an interior point, then all directions are feasible. The result then follows from Corollary 1.1 and Theorem 1.2.

- **Theorem 1.2.6 Second-Order Sufficient Condition (SOSC), Interior Case.** Let $f \in C^2$ be defined on a region in which $\boldsymbol{x}^*$ is an interior point. Suppose that

1. $\nabla f(\boldsymbol{x}^*) = 0$.
2. $\boldsymbol{F}(\boldsymbol{x}^*) > 0$.

Then, $\boldsymbol{x}^*$ is a strict local minimizer of f.

Proof. Because $f \in C^2$, we have $\boldsymbol{F}(\boldsymbol{x}^*) = \boldsymbol{F}^\top(\boldsymbol{x}^*)$. Using assumption 2 and Rayleigh's inequality it follows that if $\boldsymbol{d} \neq \boldsymbol{0}$, then $0 < \lambda_{\min}\big(\boldsymbol{F}(\boldsymbol{x}^*)\big)\|\boldsymbol{d}\|^2 \leq \boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d}$.

By Taylor's theorem and assumption 1,

$$f(\boldsymbol{x}^* + \boldsymbol{d}) - f(\boldsymbol{x}^*) = \frac{1}{2}\boldsymbol{d}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} + o(\|\boldsymbol{d}\|^2) \geq \frac{\lambda_{\min}\big(\boldsymbol{F}(\boldsymbol{x}^*)\big)}{2}\|\boldsymbol{d}\|^2 + o(\|\boldsymbol{d}\|^2).$$

Hence, for all $\boldsymbol{d}$ such that $\|\boldsymbol{d}\|$ is sufficiently small,

$$f(\boldsymbol{x}^* + \boldsymbol{d}) > f(\boldsymbol{x}^*),$$

which completes the proof.