

1 BASICS OF SET CONSTRAINED AND UNCONSTRAINED OPTIMIZATION

1.1 INTRODUCTION

- Consider the **optimization problem**

$$\begin{aligned} &\text{minimize } f(\mathbf{x}) \\ &\text{subject to } \mathbf{x} \in \Omega. \end{aligned}$$

- The function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ that we wish to minimize is a real-valued function called the **objective function**.
- The vector \mathbf{x} is an n -vector of **independent variables**: $\mathbf{x} = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$.
- The set Ω is a subset of \mathbb{R}^n called the **constraint set** or **feasible set**, which takes the form $\Omega = \{\mathbf{x}: \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$, where \mathbf{h} and \mathbf{g} are given functions.
- The **minimizer** of f over Ω is a vector \mathbf{x} which results in the smallest value of the objective function.
- Definition 1.1** Suppose that $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a real-valued function defined on some set $\Omega \subset \mathbb{R}^n$. A point $\mathbf{x}^* \in \Omega$ is a **local minimizer** of f over Ω if there exists $\varepsilon > 0$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega \setminus \{\mathbf{x}^*\}$ and $\|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$. A point $\mathbf{x}^* \in \Omega$ is a **global minimizer** of f over Ω if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega \setminus \{\mathbf{x}^*\}$.
- If \mathbf{x}^* is a global minimizer of f over Ω , we write $f(\mathbf{x}^*) = \min_{\mathbf{x} \in \Omega} f(\mathbf{x})$ and $\mathbf{x}^* = \underset{\mathbf{x} \in \Omega}{\operatorname{argmin}} f(\mathbf{x})$.

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

$$\mathbf{x}^* = \underset{x \in \Omega}{\operatorname{argmin}} f(x) = 0.$$

Note: If $\Omega = \mathbb{R}$, then $\mathbf{x}^* = \underset{x \in \Omega}{\operatorname{argmin}} f(x) = -1$.

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 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, denoted $Df(\mathbf{x})$, is

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

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

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

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

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

and

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

- 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** A vector $\mathbf{d} \in \mathbb{R}^n$, $\mathbf{d} \neq \mathbf{0}$, is a **feasible direction** at $\mathbf{x} \in \Omega$ if there exists $\alpha_0 > 0$ such that $\mathbf{x} + \alpha \mathbf{d} \in \Omega$ for all $\alpha \in [0, \alpha_0]$.
- Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be a real-valued function and let \mathbf{d} be a feasible direction at $\mathbf{x} \in \Omega$. The **directional derivative** of f in the direction \mathbf{d} , denoted $\partial f / \partial \mathbf{d}$, is the real-valued function defined by

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

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

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

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

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

- **Theorem 1.1 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 \mathbf{x}^* is a local minimizer of f over Ω , then for any feasible direction \mathbf{d} at \mathbf{x}^* , we have

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

- **Corollary 1.1 Interior Case.** Let Ω be a subset of \mathbb{R}^n and $f \in C^1$ a real-valued function on Ω . If \mathbf{x}^* is a local minimizer of f over Ω and if \mathbf{x}^* is an **interior point** of Ω , then

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

Example 1.2.2. Consider the problem

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

- Is the first-order necessary condition (FONC) for a local minimizer satisfied at $\mathbf{x} = [1, 3]^\top$?
- Is the FONC for a local minimizer satisfied at $\mathbf{x} = [0, 3]^\top$?
- Is the FONC for a local minimizer satisfied at $\mathbf{x} = [1, 0]^\top$?
- Is the FONC for a local minimizer satisfied at $\mathbf{x} = [0, 0]^\top$?

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

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

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

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

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

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

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

- **Theorem 1.2 Second-Order Necessary Condition (SONC).** Let $\Omega \subset \mathbb{R}^n$, $f \in C^2$ a function on Ω , \mathbf{x}^* a local minimizer of f over Ω , and \mathbf{d} a feasible direction at \mathbf{x}^* . If $\mathbf{d}^\top \nabla f(\mathbf{x}^*) = 0$, then

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

where \mathbf{F} is the Hessian of f .

- **Corollary 1.2 Interior Case.** Let \mathbf{x}^* be an interior point of $\Omega \subset \mathbb{R}^n$. If \mathbf{x}^* is a local minimizer of $f: \Omega \rightarrow \mathbb{R}$, $f \in C^2$, then $\nabla f(\mathbf{x}^*) = \mathbf{0}$, and $\mathbf{F}(\mathbf{x}^*)$ is positive semidefinite ($\mathbf{F}(\mathbf{x}^*) \succcurlyeq 0$); that is, for all $\mathbf{d} \in \mathbb{R}^n$, $\mathbf{d}^\top \mathbf{F}(\mathbf{x}^*) \mathbf{d} \geq 0$.

Example 1.6 Consider a function of one variable $f(x) = x^3$, $f: \mathbb{R} \rightarrow \mathbb{R}$. 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 1.7 Consider a function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, where $f(\mathbf{x}) = x_1^2 - x_2^2$. Does the point $\mathbf{x} = [0, 0]^\top$ satisfy both the FONC and SONC?

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

The Hessian matrix of f is

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

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

- **Theorem 1.3 Second-Order Sufficient Condition (SOSC), Interior Case.** Let $f \in C^2$ be defined on a region in which \mathbf{x}^* is an interior point. Suppose that

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

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

Example 1.8 Consider a function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, where $f(\mathbf{x}) = x_1^2 + x_2^2$. Does the point $\mathbf{x} = [0, 0]^\top$ satisfy the FONC, SONC, and SOSC?

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

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

satisfies the FONC, SONC, and SOSC.

APPENDIX

- **Rules of differentiation with respect to a vector.**
- **Positive definite matrix determination.**
- **Rayleigh inequality**
- **Theorem 1.1 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 \mathbf{x}^* is a local minimizer of f over Ω , then for any feasible direction \mathbf{d} at \mathbf{x}^* , we have

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

Proof. Define $\mathbf{x}(\alpha) = \mathbf{x}^* + \alpha \mathbf{d} \in \Omega$. Note that $\mathbf{x}(0) = \mathbf{x}^*$. Define the composite function

$$\phi(\alpha) = f(\mathbf{x}(\alpha)).$$

Then, by Taylor's theorem,

$$f(\mathbf{x}^* + \alpha \mathbf{d}) - f(\mathbf{x}^*) = \phi(\alpha) - \phi(0) = \phi'(0)\alpha + o(\alpha) = \alpha \mathbf{d}^\top \nabla f(\mathbf{x}(0)) + o(\alpha),$$

where $\alpha \geq 0$ and $o(\alpha)$ means that $\lim_{\alpha \rightarrow 0} o(\alpha)/\alpha = 0$. Thus, if $\phi(\alpha) \geq \phi(0)$, that is, $f(\mathbf{x}^* + \alpha \mathbf{d}) \geq f(\mathbf{x}^*)$ for sufficiently small values of $\alpha > 0$, then we have to have $\mathbf{d}^\top \nabla f(\mathbf{x}(0)) \geq 0$.

- **Corollary 1.1 Interior Case.** Let Ω be a subset of \mathbb{R}^n and $f \in C^1$ a real-valued function on Ω . If \mathbf{x}^* is a local minimizer of f over Ω and if \mathbf{x}^* is an interior point of Ω , then

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

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

- **Theorem 1.2 Second-Order Necessary Condition (SONC).** Let $\Omega \subset \mathbb{R}^n$, $f \in C^2$ a function on Ω , \mathbf{x}^* a local minimizer of f over Ω , and \mathbf{d} a feasible direction at \mathbf{x}^* . If $\mathbf{d}^\top \nabla f(\mathbf{x}^*) = 0$, then

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

where \mathbf{F} is the Hessian of f .

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

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

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

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

that is,

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

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

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

- **Corollary 1.2 Interior Case.** Let \mathbf{x}^* be an interior point of $\Omega \subset \mathbb{R}^n$. If \mathbf{x}^* is a local minimizer of $f: \Omega \rightarrow \mathbb{R}$, $f \in C^2$, then

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

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

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

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

- **Theorem 1.3 Second-Order Sufficient Condition (SOSC), Interior Case.** Let $f \in C^2$ be defined on a region in which \mathbf{x}^* is an interior point. Suppose that

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

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

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

By Taylor's theorem and assumption 1,

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

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

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

which completes the proof.