Therefore,

$$\begin{aligned}
|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle| &= \left|\int_0^1 \langle f'(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle d\tau\right| \\
&\leq \int_0^1 |\langle f'(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle| d\tau \\
&\leq \int_0^1 \|f'(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - f'(\boldsymbol{x})\|_2 \|\boldsymbol{y} - \boldsymbol{x}\|_2 d\tau \\
&\leq \int_0^1 \tau L \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 d\tau = \frac{L}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2.
\end{aligned}$$

∎

Consider a function $f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n)$. Let us fix $\boldsymbol{x}_0 \in \mathbb{R}^n$, and define two quadratic functions:

$$\begin{aligned}
\phi_1(\boldsymbol{x}) &= f(\boldsymbol{x}_0) + \langle f'(\boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0\rangle - \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2, \\
\phi_2(\boldsymbol{x}) &= f(\boldsymbol{x}_0) + \langle f'(\boldsymbol{x}_0), \boldsymbol{x} - \boldsymbol{x}_0\rangle + \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2.
\end{aligned}$$

Then the graph of the function f is located between the graphs of ϕ_1 and ϕ_2:

$$\phi_1(\boldsymbol{x}) \leq f(\boldsymbol{x}) \leq \phi_2(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^n.$$

Lemma 3.5 Let $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$. Then for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\|f'(\boldsymbol{y}) - f'(\boldsymbol{x}) - f''(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})\|_2 \leq \frac{M}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2,$$

$$|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle - \frac{1}{2}\langle f''(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle| \leq \frac{M}{6}\|\boldsymbol{y} - \boldsymbol{x}\|_2^3.$$

Lemma 3.6 Let $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, with $\|f''(\boldsymbol{x}) - f''(\boldsymbol{y})\|_2 \leq M\|\boldsymbol{x} - \boldsymbol{y}\|_2$. Then

$$f''(\boldsymbol{x}) - M\|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I} \preceq f''(\boldsymbol{y}) \preceq f''(\boldsymbol{x}) + M\|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I}.$$

Proof:

Since $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$, $\|f''(\boldsymbol{y}) - f''(\boldsymbol{x})\|_2 \leq M\|\boldsymbol{y} - \boldsymbol{x}\|_2$. This means that the eigenvalues of the symmetric matrix $f''(\boldsymbol{y}) - f''(\boldsymbol{x})$ satisfy:

$$|\lambda_i(f''(\boldsymbol{y}) - f''(\boldsymbol{x}))| \leq M\|\boldsymbol{y} - \boldsymbol{x}\|_2, \quad i = 1, 2, \ldots, n.$$

Therefore,

$$-M\|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I} \preceq f''(\boldsymbol{y}) - f''(\boldsymbol{x}) \preceq M\|\boldsymbol{y} - \boldsymbol{x}\|_2 \boldsymbol{I}.$$

∎

3.1 Exercises

1. Prove Lemma 3.5.

4 Optimality Conditions for Differentiable Functions in $\mathbb{R}^n$

Let $f(\boldsymbol{x})$ be differentiable at $\bar{\boldsymbol{x}}$. Then for $\boldsymbol{y} \in \mathbb{R}^n$, we have

$$f(\boldsymbol{y}) = f(\bar{\boldsymbol{x}}) + \langle f'(\bar{\boldsymbol{x}}), \boldsymbol{y} - \bar{\boldsymbol{x}} \rangle + o(\|\boldsymbol{y} - \bar{\boldsymbol{x}}\|_2),$$

where $o(r)$ is some function of $r > 0$ such that

$$\lim_{r \to 0} \frac{1}{r} o(r) = 0, \; o(0) = 0.$$

Let $\boldsymbol{s}$ be a direction in $\mathbb{R}^n$ such that $\|\boldsymbol{s}\|_2 = 1$. Consider the local decrease (or increase) of $f(\boldsymbol{x})$ along $\boldsymbol{s}$:

$$\Delta(\boldsymbol{s}) = \lim_{\alpha \to 0} \frac{1}{\alpha} \left[f(\bar{\boldsymbol{x}} + \alpha \boldsymbol{s}) - f(\bar{\boldsymbol{x}}) \right].$$

Since $f(\bar{\boldsymbol{x}} + \alpha \boldsymbol{s}) - f(\bar{\boldsymbol{x}}) = \alpha \langle f'(\bar{\boldsymbol{x}}), \boldsymbol{s} \rangle + o(\|\alpha \boldsymbol{s}\|_2)$, we have $\Delta(\boldsymbol{s}) = \langle f'(\bar{\boldsymbol{x}}), \boldsymbol{s} \rangle$.

Using the Cauchy-Schwartz inequality $-\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 \leq \langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2$,

$$\Delta(\boldsymbol{s}) = \langle f'(\bar{\boldsymbol{x}}), \boldsymbol{s} \rangle \geq -\|f'(\bar{\boldsymbol{x}})\|_2.$$

Choosing the direction $\bar{\boldsymbol{s}} = -f'(\bar{\boldsymbol{x}}) / \|f'(\bar{\boldsymbol{x}})\|_2$,

$$\Delta(\bar{\boldsymbol{s}}) = -\left\langle f'(\bar{\boldsymbol{x}}), \frac{f'(\bar{\boldsymbol{x}})}{\|f'(\bar{\boldsymbol{x}})\|_2} \right\rangle = -\|f'(\bar{\boldsymbol{x}})\|_2.$$

Thus, the direction $-f'(\bar{\boldsymbol{x}})$ is the direction of the *fastest local decrease of* $f(\boldsymbol{x})$ *at point* $\bar{\boldsymbol{x}}$.

Theorem 4.1 (First-order necessary optimality condition) Let $\boldsymbol{x}^*$ be a local minimum of the differentiable function $f(\boldsymbol{x})$. Then

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

Proof:

Let $\boldsymbol{x}^*$ be the local minimum of $f(\boldsymbol{x})$. Then, there is $r > 0$ such that for all $\boldsymbol{y}$ with $\|\boldsymbol{y} - \boldsymbol{x}^*\|_2 \leq r$, $f(\boldsymbol{y}) \geq f(\boldsymbol{x}^*)$.

Since f is differentiable,

$$f(\boldsymbol{y}) = f(\boldsymbol{x}^*) + \langle f'(\boldsymbol{x}^*), \boldsymbol{y} - \boldsymbol{x}^* \rangle + o(\|\boldsymbol{y} - \boldsymbol{x}^*\|_2) \geq f(\boldsymbol{x}^*).$$

Dividing by $\|\boldsymbol{y} - \boldsymbol{x}^*\|_2$, and taking the limit $\boldsymbol{y} \to \boldsymbol{x}^*$,

$$\langle f'(\boldsymbol{x}^*), \boldsymbol{s} \rangle \geq 0, \quad \forall \boldsymbol{s} \in \mathbb{R}^n, \quad \|\boldsymbol{s}\|_2 = 1.$$

Consider the opposite direction $-\boldsymbol{s}$, and then we conclude that

$$\langle f'(\boldsymbol{x}^*), \boldsymbol{s} \rangle = 0, \quad \forall \boldsymbol{s} \in \mathbb{R}^n, \quad \|\boldsymbol{s}\|_2 = 1.$$

Choosing $\boldsymbol{s} = \boldsymbol{e}_i \quad (i = 1, 2, \ldots, n)$, we conclude that $f'(\boldsymbol{x}^*) = 0$. ∎

Remark 4.2 For the first-order sufficient optimality condition, we need convexity for the function $f(\boldsymbol{x})$.

Corollary 4.3 Let $\boldsymbol{x}^*$ be a local minimum of a differentiable function $f(\boldsymbol{x})$ subject to linear equality constraints

$$\boldsymbol{x} \in \mathcal{L} \equiv \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\} \neq \emptyset,$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $m < n$.

Then, there exists a vector of multipliers $\boldsymbol{\lambda}^*$ such that

$$f'(\boldsymbol{x}^*) = \boldsymbol{A}^T \boldsymbol{\lambda}^*.$$

Proof:

Consider the vectors $\boldsymbol{u}_i$ ($i = 1, 2, \ldots, k$) with $k \geq n - m$ which form an orthonormal basis of the null space of $\boldsymbol{A}$. Then, $\boldsymbol{x} \in \mathcal{L}$ can be represented as

$$\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{t}) \equiv \boldsymbol{x}^* + \sum_{i=1}^{k} t_i \boldsymbol{u}_i, \quad \boldsymbol{t} \in \mathbb{R}^k.$$

Moreover, the point $\boldsymbol{t} = \boldsymbol{0}$ is the local minimal solution of the function $\phi(\boldsymbol{t}) = f(\boldsymbol{x}(\boldsymbol{t}))$.

From Theorem 4.1, $\phi'(\boldsymbol{0}) = \boldsymbol{0}$. That is,

$$\frac{d\phi}{dt_i}(\boldsymbol{0}) = \langle f'(\boldsymbol{x}^*), \boldsymbol{u}_i \rangle = 0, \quad i = 1, 2, \ldots, k.$$

Now there is $\boldsymbol{t}^*$ and $\boldsymbol{\lambda}^*$ such that

$$f'(\boldsymbol{x}^*) = \sum_{i=1}^{k} t_i^* \boldsymbol{u}_i + \boldsymbol{A}^T \boldsymbol{\lambda}^*.$$

For each $i = 1, 2, \ldots, k$,

$$\langle f'(\boldsymbol{x}^*), \boldsymbol{u}_i \rangle = t_i^* = 0.$$

Therefore, we have the result. ∎

The following type of result is called *theorems of the alternative*, and are closed related to duality theory in optimization.

Corollary 4.4 Given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{c} \in \mathbb{R}^n$, $\eta \in \mathbb{R}$, either

$$\left\{ \begin{array}{c} \langle \boldsymbol{c}, \boldsymbol{x} \rangle < \eta \\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array} \right. \text{ has a solution } \boldsymbol{x} \in \mathbb{R}^n, \tag{1}$$

or

$$\left(\begin{array}{c} \left\{ \begin{array}{c} \langle \boldsymbol{b}, \boldsymbol{\lambda} \rangle > 0 \\ \boldsymbol{A}^T \boldsymbol{\lambda} = \boldsymbol{0} \end{array} \right. \\ \text{or} \\ \left\{ \begin{array}{c} \langle \boldsymbol{b}, \boldsymbol{\lambda} \rangle \geq \eta \\ \boldsymbol{A}^T \boldsymbol{\lambda} = \boldsymbol{c} \end{array} \right. \end{array} \right) \text{ has a solution } \boldsymbol{\lambda} \in \mathbb{R}^m, \tag{2}$$

but never both

Proof:

Let us first show that if $\exists \boldsymbol{x} \in \mathbb{R}^n$ satisfying (1), $\nexists \boldsymbol{\lambda} \in \mathbb{R}^m$ satisfying (2). Let us assume by contradiction that $\exists \boldsymbol{\lambda}$. Then $\langle \boldsymbol{\lambda}, \boldsymbol{A}\boldsymbol{x} \rangle = \langle \boldsymbol{\lambda}, \boldsymbol{b} \rangle$ and in the homogeneous case it gives $0 = \langle \boldsymbol{\lambda}, \boldsymbol{b} \rangle > 0$ and in the non-homogeneous case it gives $\eta > \langle \boldsymbol{c}, \boldsymbol{x} \rangle = \langle \boldsymbol{\lambda}, \boldsymbol{b} \rangle \geq \eta$. Both of cases are impossible.

Now, let us assume that $\nexists \boldsymbol{x} \in \mathbb{R}^n$ satisfying (1). If additionally $\nexists \boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, it means that the columns of the matrix $\boldsymbol{A}$ do not spam the vector $\boldsymbol{b}$. Therefore, there is $\boldsymbol{0} \neq \boldsymbol{\lambda} \in \mathbb{R}^m$ which is orthogonal to all of these columns and $\langle \boldsymbol{b}, \boldsymbol{\lambda} \rangle \neq 0$. Selecting the correct sign, we constructed a $\boldsymbol{\lambda}$ which satisfies the homogeneous system of (2). Now, if for all $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ we have $\langle \boldsymbol{c}, \boldsymbol{x} \rangle \geq \eta$, it means that the minimization of the function $f(\boldsymbol{x}) = \langle \boldsymbol{c}, \boldsymbol{x} \rangle$ subject to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has an optimal solution $\boldsymbol{x}^*$ with $f(\boldsymbol{x}^*) \geq \eta$ (since $\exists \boldsymbol{x} \in \mathbb{R}^n$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, if $n = m$, take $\boldsymbol{\lambda} = \boldsymbol{A}^{-T}\boldsymbol{c}$. Otherwise, we can assume $n > m$ w.l.o.g.). From Corollary 4.3, $\exists \boldsymbol{\lambda} \in \mathbb{R}^m$ such that $\boldsymbol{A}^T \boldsymbol{\lambda} = \boldsymbol{c}$, and $\langle \boldsymbol{b}, \boldsymbol{\lambda} \rangle = \langle \boldsymbol{x}^*, \boldsymbol{A}^T \boldsymbol{\lambda} \rangle = \langle \boldsymbol{x}^*, \boldsymbol{c} \rangle \geq \eta$. ∎

If $f(\boldsymbol{x})$ is twice differentiable at $\bar{\boldsymbol{x}} \in \mathbb{R}^n$, then for $\boldsymbol{y} \in \mathbb{R}^n$, we have

$$f'(\boldsymbol{y}) = f'(\bar{\boldsymbol{x}}) + f''(\bar{\boldsymbol{x}})(\boldsymbol{y} - \bar{\boldsymbol{x}}) + \boldsymbol{o}(\|\boldsymbol{y} - \bar{\boldsymbol{x}}\|_2),$$

where $\boldsymbol{o}(r)$ is such that $\lim_{r \to 0} \|\boldsymbol{o}(r)\|_2 / r = 0$ and $\boldsymbol{o}(0) = 0$.

Theorem 4.5 (Second-order necessary optimality condition) Let $\boldsymbol{x}^*$ be a local minimum of a twice continuously differentiable function $f(\boldsymbol{x})$. Then

$$f'(\boldsymbol{x}^*) = 0, \qquad f''(\boldsymbol{x}^*) \succeq \boldsymbol{O}.$$

Proof:

Since $\boldsymbol{x}^*$ is a local minimum of $f(\boldsymbol{x})$, $\exists r > 0$ such that for all $\boldsymbol{y} \in \mathbb{R}^n$ which satisfy $\|\boldsymbol{y}-\boldsymbol{x}^*\|_2 \leq r$, $f(\boldsymbol{y}) \geq f(\boldsymbol{x}^*)$.

From Theorem 4.1, $f'(\boldsymbol{x}^*) = 0$. Then

$$f(\boldsymbol{y}) = f(\boldsymbol{x}^*) + \frac{1}{2}\langle f''(\boldsymbol{x}^*)(\boldsymbol{y} - \boldsymbol{x}^*), \boldsymbol{y} - \boldsymbol{x}^*\rangle + o(\|\boldsymbol{y} - \boldsymbol{x}^*\|_2^2) \geq f(\boldsymbol{x}^*).$$

And $\langle f''(\boldsymbol{x}^*)\boldsymbol{s}, \boldsymbol{s}\rangle \geq 0,\ \forall \boldsymbol{s} \in \mathbb{R}^n$ with $\|\boldsymbol{s}\|_2 = 1$. ∎

Theorem 4.6 (Second-order sufficient optimality condition) Let the function $f(\boldsymbol{x})$ be twice continuously differentiable on $\mathbb{R}^n$, and let $\boldsymbol{x}^*$ satisfy the following conditions:

$$f'(\boldsymbol{x}^*) = 0, \quad f''(\boldsymbol{x}^*) \succ \boldsymbol{O}.$$

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

Proof:

In a small neighborhood of $\boldsymbol{x}^*$, function $f(\boldsymbol{x}^*)$ can be represented as:

$$f(\boldsymbol{y}) = f(\boldsymbol{x}^*) + \frac{1}{2}\langle f''(\boldsymbol{x}^*)(\boldsymbol{y} - \boldsymbol{x}^*), \boldsymbol{y} - \boldsymbol{x}^*\rangle + o(\|\boldsymbol{y} - \boldsymbol{x}^*\|_2^2).$$

Since $o(r)/r \to 0$, there is a $\bar{r} > 0$ such that for all $r \in [0, \bar{r}]$,

$$|o(r)| \leq \frac{r}{4}\lambda_1(f''(\boldsymbol{x}^*)),$$

where $\lambda_1(f''(\boldsymbol{x}^*))$ is the smallest eigenvalue of the matrix $f''(\boldsymbol{x}^*)$ which is positive. Then

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}^*) + \frac{1}{2}\lambda_1(f''(\boldsymbol{x}^*))\|\boldsymbol{y} - \boldsymbol{x}^*\|_2^2 + o(\|\boldsymbol{y} - \boldsymbol{x}^*\|_2^2).$$

Considering that $\bar{r} < 1$, $|o(r^2)| \leq r^2/4\lambda_1(f''(\boldsymbol{x}^*))$ for $r \in [0, \bar{r}]$, finally

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}^*) + \frac{1}{4}\lambda_1(f''(\boldsymbol{x}^*))\|\boldsymbol{y} - \boldsymbol{x}^*\|_2^2 > f(\boldsymbol{x}^*).$$

∎

4.1 Exercises

1. In view of Theorem 4.6, find a twice continuously differentiable function on $\mathbb{R}^n$ which satisfies $f'(\boldsymbol{x}^*) = 0, \quad f''(\boldsymbol{x}^*) \succeq \boldsymbol{O}$, but $\boldsymbol{x}^*$ is not a local minimum of $f(\boldsymbol{x})$.

5 Algorithms for Minimizing Unconstrained Functions

5.1 General Minimization Problem and Terminologies

Definition 5.1 We define the *general minimization problem* as follows

$$\begin{cases} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & f_j(\boldsymbol{x}) \ \&\ 0, \quad j = 1, 2, \ldots, m \\ & \boldsymbol{x} \in S, \end{cases} \tag{3}$$

where $f : \mathbb{R}^n \to \mathbb{R},\ f_j : \mathbb{R}^n \to \mathbb{R}$ $(j = 1, 2, \ldots, m)$, the symbol & could be $=$, $\geq$, or $\leq$, and $S \subseteq \mathbb{R}^n$.