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 satisfies $\|\boldsymbol{y} - \boldsymbol{x}^*\| \leq r$, $f(\boldsymbol{y}) \geq f(\boldsymbol{x}^*)$.

From Theorem 1.3.1, $f'(\boldsymbol{x}^*) = 0$. Then

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

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

Theorem 1.3.4 (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:

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

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

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

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

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

∎

1.4 Lipschitz continuously differentiable functions

Definition 1.4.1 Let Q be a subset of $\mathbb{R}^n$. We denote by $C_L^{k,p}(Q)$ the class of functions with the following properties:

- Any $f \in C_L^{k,p}(Q)$ is k times continuously differentiable on Q.
- Its pth derivative is Lipschitz continuous on Q with the constant L:

$$\|f^{(p)}(\boldsymbol{x}) - f^{(p)}(\boldsymbol{y})\| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in Q.$$

If $f_1 \in C_L^{k,p}(Q)$, $f_2 \in C_L^{k,p}(Q)$, and $\alpha, \beta \in \mathbb{R}$, then for $L_3 = |\alpha|L_1 + |\beta|L_2$, $\alpha f_1 + \beta f_2 \in C_{L_3}^{k,p}(Q)$.

Lemma 1.4.2 Let $f \in C^2(\mathbb{R}^n)$. Then $f \in C_L^{2,1}(\mathbb{R}^n)$ if and only if $\|f''(\boldsymbol{x})\| \leq L, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$.

Proof: For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$\begin{aligned} f'(\boldsymbol{y}) &= f'(\boldsymbol{x}) + \int_0^1 f''(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x}) d\tau \\ &= f'(\boldsymbol{x}) + \left(\int_0^1 f''(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) d\tau \right) (\boldsymbol{y} - \boldsymbol{x}). \end{aligned}$$

Since $\|f''(\boldsymbol{x})\| \leq L$,

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

On the other hand, for $\boldsymbol{s} \in \mathbb{R}^n$, and $\alpha \in \mathbb{R}^*$,

$$\|f'(\boldsymbol{x} + \alpha \boldsymbol{s}) - f'(\boldsymbol{x})\| \leq \alpha L \|\boldsymbol{s}\|.$$

Dividing both sides by α and taking the limit to zero,

$$\|f''(\boldsymbol{x})\boldsymbol{s}\| \leq L\|\boldsymbol{s}\|, \quad \boldsymbol{s} \in \mathbb{R}^n.$$

Therefore, $\|f''(\boldsymbol{x})\| \leq L$. ∎

Example 1.4.3

1. The linear function $f(\boldsymbol{x}) = \alpha + \langle \boldsymbol{a}, \boldsymbol{x} \rangle \in C_0^{1,1}(\mathbb{R}^n)$ since
$$f'(\boldsymbol{x}) = \boldsymbol{a}, \quad f''(\boldsymbol{x}) = \boldsymbol{O}.$$
2. The quadratic function $f(\boldsymbol{x}) = \alpha + \langle \boldsymbol{a}, \boldsymbol{x} \rangle + 1/2 \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \rangle$ with $\boldsymbol{A} = \boldsymbol{A}^T$ belongs to $C_L^{1,1}(\mathbb{R}^n)$ where
$$f'(\boldsymbol{x}) = \boldsymbol{a} + \boldsymbol{A}\boldsymbol{x}, \quad f''(\boldsymbol{x}) = \boldsymbol{A}, \quad L = \|\boldsymbol{A}\|.$$
3. The function $f(\boldsymbol{x}) = \sqrt{1 + x^2} \in C_1^{1,1}(\mathbb{R})$ since
$$f'(x) = \frac{x}{\sqrt{1 + x^2}}, \quad f''(\boldsymbol{x}) = \frac{1}{(1 + x^2)^{3/2}} \leq 1.$$

Lemma 1.4.4 Let $f \in C_L^{1,1}(\mathbb{R}^n)$. Then for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$|f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle| \leq \frac{L}{2} \|\boldsymbol{y} - \boldsymbol{x}\|^2.$$

Proof: For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

$$\begin{aligned} f(\boldsymbol{y}) &= f(\boldsymbol{x}) + \int_0^1 \langle f'(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})), \boldsymbol{y} - \boldsymbol{x} \rangle d\tau \\ &= f(\boldsymbol{x}) + \langle f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \int_0^1 \langle f'(\boldsymbol{x} + \tau(\boldsymbol{y} - \boldsymbol{x})) - f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle d\tau. \end{aligned}$$

Therefore,

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

∎

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

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

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 1.4.5 Let $f \in C_M^{2,2}(\mathbb{R}^n)$. Then for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have

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

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

Proof: Left for exercise. ∎

Lemma 1.4.6 Let $f \in C_M^{2,2}(\mathbb{R}^n)$, with $\|f''(\boldsymbol{x}) - f''(\boldsymbol{y})\|_2 \leq M\|\boldsymbol{x} - \boldsymbol{y}\|$. Then

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

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

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

Therefore,

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

∎

1.5 The gradient method

Gradient Method	
Choose:	$\boldsymbol{x}_0 \in \mathbb{R}^n$
Iterate:	$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - h_k f'(\boldsymbol{x}_k), \quad k = 0, 1, \ldots$

We consider three strategies for the step-size choice:

1. **Constant Step**

   The sequence $\{h_k\}_{k=0}^{\infty}$ is chosen in *advance*. For example

   $$h_k = h > 0,$$

   $$h_k = \frac{h}{\sqrt{k+1}}.$$

   This is the simplest strategy.

2. **Exact Line Search (Cauchy Step-Size)**

   The sequence $\{h_k\}_{k=0}^{\infty}$ is chosen such that

   $$h_k = \text{argmin}_{h \geq 0} f(\boldsymbol{x}_k - h f'(\boldsymbol{x}_k)).$$

   This choice is only theoretical since even for the one dimensional case, it is very difficult and expensive.

3. **Goldstein-Armijo Rule**

   Find a sequence $\{h_k\}_{k=0}^{\infty}$ such that

   $$\begin{aligned} \alpha \langle f'(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k+1} \rangle &\leq f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}), \\ \beta \langle f'(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k+1} \rangle &\geq f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}), \end{aligned}$$

   where $0 < \alpha < \beta < 1$ are fixed parameters.

   Since $f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k - h_k f'(\boldsymbol{x}_k))$,

   $$f(\boldsymbol{x}_k) - \beta h_k \|f'(\boldsymbol{x}_k)\|^2 \leq f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{x}_k) - \alpha h_k \|f'(\boldsymbol{x}_k)\|^2.$$

   The acceptable steps exist unless $f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k - h f'(\boldsymbol{x}_k))$ is not bounded below.

Consider the problem

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$$

with $f \in C_L^{1,1}(\mathbb{R}^n)$, and $f(\boldsymbol{x})$ is bounded from below.