Example 5.4 The function $-\log x$ is convex on $(0, +\infty)$. Let $a, b \in (0, +\infty)$ and $0 \le \theta \le 1$. Then, from the definition of the convexity, we have

$$-\log(\theta a + (1-\theta)b) \le -\theta \log a - (1-\theta)\log b.$$

If we take the exponential of both sides, we obtain

$$a^\theta b^{1-\theta} \le \theta a + (1-\theta)b.$$

For $\theta = \frac{1}{2}$, we have the arithmetic-geometric mean inequality: $\sqrt{ab} \le \dfrac{a+b}{2}$.

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$, $p > 1$, and q such that $\frac{1}{p} + \frac{1}{q} = 1$. Consider

$$a = \frac{|[\boldsymbol{x}]_i|^p}{\sum_{j=1}^n |[\boldsymbol{x}]_j|^p}, \ b = \frac{|[\boldsymbol{y}]_i|^q}{\sum_{j=1}^n |[\boldsymbol{y}]_j|^q}, \ \theta = \frac{1}{p}, \text{ and } (1-\theta) = \frac{1}{q}.$$

Then we have

$$\left(\frac{|[\boldsymbol{x}]_i|^p}{\sum_{j=1}^n |[\boldsymbol{x}]_j|^p} \right)^{\frac{1}{p}} \left(\frac{|[\boldsymbol{y}]_i|^q}{\sum_{j=1}^n |[\boldsymbol{y}]_j|^q} \right)^{\frac{1}{q}} \le \frac{|[\boldsymbol{x}]_i|^p}{p\sum_{j=1}^n |[\boldsymbol{x}]_j|^p} + \frac{|[\boldsymbol{y}]_i|^q}{q\sum_{j=1}^n |[\boldsymbol{y}]_j|^q}.$$

and summing over i, we obtain the Hölder inequality:

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \le \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q$$

where $\|\boldsymbol{x}\|_p := \left(\sum_{i=1}^n |[\boldsymbol{x}]_i|^p \right)^{\frac{1}{p}}$.

Theorem 5.5 (Jensen's inequality) A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for any positive integer m, the following condition is valid

$$\left. \begin{array}{l} \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n \\ \alpha_1, \alpha_2, \ldots, \alpha_m \ge 0 \\ \sum_{i=1}^m \alpha_i = 1 \end{array} \right\} \Rightarrow f\left(\sum_{i=1}^m \alpha_i \boldsymbol{x}_i \right) \le \sum_{i=1}^m \alpha_i f(\boldsymbol{x}_i).$$

Proof:
Left for exercise. ∎

Theorem 5.6 Let $\{f_i\}_{i \in I}$ be a family of (finite or infinite) functions which are bounded from above and $f_i \in \mathcal{F}(\mathbb{R}^n)$. Then, $f(\boldsymbol{x}) := \sup_{i \in I} f_i(\boldsymbol{x})$ is convex on $\mathbb{R}^n$.

Proof:
For each $i \in I$, since $f_i \in \mathcal{F}(\mathbb{R}^n)$, its epigraph $E_i = \{(\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \mid f_i(\boldsymbol{x}) \le y\}$ is convex on $\mathbb{R}^{n+1}$ by Theorem 5.2. Also their intersection

$$\bigcap_{i \in I} E_i = \bigcap_{i \in I} \left\{ (\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \mid f_i(\boldsymbol{x}) \le y \right\} = \left\{ (\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \;\middle|\; \sup_{i \in I} f_i(\boldsymbol{x}) \le y \right\}$$

is convex by Exercise 2 of Section 1, which is exactly the epigraph of $f(\boldsymbol{x})$. ∎

5.2 Differentiable Convex Functions

Theorem 5.7 Let f be a continuously differentiable function. The following conditions are equivalent:

1. $f \in \mathcal{F}^1(\mathbb{R}^n)$.
2. $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.
3. $\langle \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \geq 0, \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

Proof:
Left for exercise. ∎

Theorem 5.8 (First-order sufficient optimality condition) If $f \in \mathcal{F}^1(\mathbb{R}^n)$ and $\boldsymbol{\nabla f}(\boldsymbol{x}^*) = 0$, then $\boldsymbol{x}^*$ is the *global minimum* of $f(\boldsymbol{x})$ on $\mathbb{R}^n$.

Proof:
Left for exercise. ∎

Lemma 5.9 If $f \in \mathcal{F}^1(\mathbb{R}^m), \ \boldsymbol{b} \in \mathbb{R}^m$, and $\boldsymbol{A} : \mathbb{R}^n \to \mathbb{R}^m$, then

$$\phi(\boldsymbol{x}) = f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) \in \mathcal{F}^1(\mathbb{R}^n).$$

Proof:
Left for exercise. ∎

Example 5.10 The following functions are differentiable and convex:

1. $f(x) = e^x$
2. $f(x) = |x|^p, \quad p > 1$
3. $f(x) = \frac{x^2}{1+|x|}$
4. $f(x) = |x| - \ln(1 + |x|)$
5. $f(\boldsymbol{x}) = \sum_{i=1}^m e^{\alpha_i + \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle}$
6. $f(\boldsymbol{x}) = \sum_{i=1}^m |\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle - b_i|^p, \quad p > 1$

Theorem 5.11 Let f be a twice continuously differentiable function. Then $f \in \mathcal{F}^2(\mathbb{R}^n)$ if and only if

$$\boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}) \succeq \boldsymbol{O}, \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Proof:
Let $f \in \mathcal{F}^2(\mathbb{R}^n)$, and denote $\boldsymbol{x}_\tau = \boldsymbol{x} + \tau \boldsymbol{s}, \ \tau > 0$. Then, from the previous result

$$\begin{aligned} 0 &\leq \frac{1}{\tau^2} \langle \boldsymbol{\nabla f}(\boldsymbol{x}_\tau) - \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{x}_\tau - \boldsymbol{x} \rangle = \frac{1}{\tau} \langle \boldsymbol{\nabla f}(\boldsymbol{x}_\tau) - \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{s} \rangle \\ &= \frac{1}{\tau} \int_0^\tau \langle \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x} + \lambda \boldsymbol{s}) \boldsymbol{s}, \boldsymbol{s} \rangle d\lambda \\ &= \frac{F(\tau) - F(0)}{\tau} \end{aligned}$$

where $F(\tau) = \int_0^\tau \langle \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x} + \lambda \boldsymbol{s}) \boldsymbol{s}, \boldsymbol{s} \rangle d\lambda$. Therefore, tending τ to 0, we get $0 \leq F'(0) = \langle \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{s}, \boldsymbol{s} \rangle$, and we have the result.

Conversely, $\forall \boldsymbol{x} \in \mathbb{R}^n$,

$$\begin{aligned} f(\boldsymbol{y}) &= f(\boldsymbol{x}) + \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \int_0^1 \int_0^\tau \langle \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle d\lambda d\tau \\ &\geq f(\boldsymbol{x}) + \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle. \end{aligned}$$

∎

5.3 Differentiable Convex Functions with Lipschitz Continuous Gradients

Corollary 5.12 Let f be a two times continuously differentiable function. $f \in \mathcal{F}_L^{2,1}(\mathbb{R}^n)$ if and only if $\boldsymbol{O} \preceq \boldsymbol{\nabla}^2 \boldsymbol{f}(\boldsymbol{x}) \preceq L\boldsymbol{I}, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$.

Proof:
Left for exercise. ∎

Theorem 5.13 Let f be a continuously differentiable function on $\mathbb{R}^n$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, and $\alpha \in [0, 1]$. Then the following conditions are equivalent:

1. $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$.
2. $0 \leq f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \leq \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$.
3. $f(\boldsymbol{x}) + \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2L} \|\boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y})\|_2^2 \leq f(\boldsymbol{y})$.
4. $0 \leq \frac{1}{L} \|\boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y})\|_2^2 \leq \langle \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle$.
5. $0 \leq \langle \boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \leq L \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$.
6. $f(\alpha \boldsymbol{x} + (1 - \alpha)\boldsymbol{y}) + \frac{\alpha(1-\alpha)}{2L} \|\boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y})\|_2^2 \leq \alpha f(\boldsymbol{x}) + (1 - \alpha) f(\boldsymbol{y})$.
7. $0 \leq \alpha f(\boldsymbol{x}) + (1 - \alpha) f(\boldsymbol{y}) - f(\alpha \boldsymbol{x} + (1 - \alpha)\boldsymbol{y}) \leq \alpha(1 - \alpha)\frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$.

Proof:
$\boxed{1 \Rightarrow 2}$ It follows from Lemmas 5.7 and 3.6.
$\boxed{2 \Rightarrow 3}$ Fix $\boldsymbol{x} \in \mathbb{R}^n$, and consider the function $\phi(\boldsymbol{y}) = f(\boldsymbol{y}) - \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} \rangle$. Clearly $\phi(\boldsymbol{y})$ satisfies 2. Also, $\boldsymbol{y}^* = \boldsymbol{x}$ is a minimal solution. Therefore from 2,

$$\begin{aligned} \phi(\boldsymbol{x}) &= \phi(\boldsymbol{y}^*) \leq \phi\left(\boldsymbol{y} - \frac{1}{L} \boldsymbol{\nabla \phi}(\boldsymbol{y})\right) \leq \phi(\boldsymbol{y}) + \frac{L}{2} \left\| \frac{1}{L} \boldsymbol{\nabla \phi}(\boldsymbol{y}) \right\|_2^2 + \langle \boldsymbol{\nabla \phi}(\boldsymbol{y}), -\frac{1}{L} \boldsymbol{\nabla \phi}(\boldsymbol{y}) \rangle \\ &= \phi(\boldsymbol{y}) + \frac{1}{2L} \|\boldsymbol{\nabla \phi}(\boldsymbol{y})\|_2^2 - \frac{1}{L} \|\boldsymbol{\nabla \phi}(\boldsymbol{y})\|_2^2 = \phi(\boldsymbol{y}) - \frac{1}{2L} \|\boldsymbol{\nabla \phi}(\boldsymbol{y})\|_2^2. \end{aligned}$$

Since $\boldsymbol{\nabla \phi}(\boldsymbol{y}) = \boldsymbol{\nabla f}(\boldsymbol{y}) - \boldsymbol{\nabla f}(\boldsymbol{x})$, finally we have

$$f(\boldsymbol{x}) - \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{x} \rangle \leq f(\boldsymbol{y}) - \langle \boldsymbol{\nabla f}(\boldsymbol{x}), \boldsymbol{y} \rangle - \frac{1}{2L} \|\boldsymbol{\nabla f}(\boldsymbol{y}) - \boldsymbol{\nabla f}(\boldsymbol{x})\|_2^2.$$

$\boxed{3 \Rightarrow 4}$ Adding two copies of 3 with $\boldsymbol{x}$ and $\boldsymbol{y}$ interchanged, we obtain 4.
$\boxed{4 \Rightarrow 1}$ Applying the Cauchy-Schwarz inequality to 4, we obtain $\|\boldsymbol{\nabla f}(\boldsymbol{x}) - \boldsymbol{\nabla f}(\boldsymbol{y})\|_2 \leq L \|\boldsymbol{x} - \boldsymbol{y}\|_2$. Also from Theorem 5.7, $f(\boldsymbol{x})$ is convex.
$\boxed{2 \Rightarrow 5}$ Adding two copies of 2 with $\boldsymbol{x}$ and $\boldsymbol{y}$ interchanged, we obtain 5.
$\boxed{5 \Rightarrow 2}$

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