Theorem 6.5 Let f be a continuously differentiable function. The following conditions are equivalent:

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

Proof: Left for exercise. ∎

Theorem 6.6 Let f be a twice continuously differentiable function. Then $f \in \mathcal{F}^2(\mathbb{R}^n)$ if and only if

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

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

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

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

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

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

∎

Corollary 6.7 Let f be a two times continuously differentiable function. $f \in \mathcal{F}_L^{2,1}(\mathbb{R}^n)$ if and only if $\boldsymbol{O} \preceq f''(\boldsymbol{x}) \preceq L\boldsymbol{I}, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$.

Theorem 6.8 Let f be a continuously differentiable function in $\mathbb{R}^n$, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, and $\alpha \in [0,1]$. Then the following conditions are equivalent:

1. $f \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$.
2. $0 \leq f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle \leq \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2$.
3. $f(\boldsymbol{x}) + \langle f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle + \frac{1}{2L}\|f'(\boldsymbol{x}) - f'(\boldsymbol{y})\|_2^2 \leq f(\boldsymbol{y})$.
4. $0 \leq \frac{1}{L}\|f'(\boldsymbol{x}) - f'(\boldsymbol{y})\|_2^2 \leq \langle f'(\boldsymbol{x}) - f'(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle$.
5. $0 \leq \langle f'(\boldsymbol{x}) - f'(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_2^2$.
6. $f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) + \frac{\alpha(1-\alpha)}{2L}\|f'(\boldsymbol{x}) - f'(\boldsymbol{y})\|_2^2 \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})$.
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 the definition of convex function and Lemma 3.4.

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

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

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

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

$\boxed{3\Rightarrow 4}$ 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 $\|f'(\boldsymbol{x}) - f'(\boldsymbol{y})\|_2 \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_2$. Also from Theorem 6.5, $f(\boldsymbol{x})$ is convex.

$\boxed{2\Rightarrow 5}$ Adding two copies of 2 with $\boldsymbol{x}$ and $\boldsymbol{y}$ interchanged, we obtain 5.

$\boxed{5\Rightarrow 2}$

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

The non-negativity follows from Theorem 6.5.

$\boxed{3\Rightarrow 6}$ Denote $\boldsymbol{x}_\alpha = \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}$. From 3,

$$\begin{aligned}f(\boldsymbol{x}) &\geq f(\boldsymbol{x}_\alpha) + \langle f'(\boldsymbol{x}_\alpha), (1-\alpha)(\boldsymbol{x} - \boldsymbol{y})\rangle + \frac{1}{2L}\|f'(\boldsymbol{x}) - f'(\boldsymbol{x}_\alpha)\|_2^2 \\ f(\boldsymbol{y}) &\geq f(\boldsymbol{x}_\alpha) + \langle f'(\boldsymbol{x}_\alpha), \alpha(\boldsymbol{y} - \boldsymbol{x})\rangle + \frac{1}{2L}\|f'(\boldsymbol{y}) - f'(\boldsymbol{x}_\alpha)\|_2^2.\end{aligned}$$

Multiplying the first inequality by α, the second by $1-\alpha$, and summing up, we have

$$\alpha f(\boldsymbol{x}) + (1-\alpha)f(\boldsymbol{y}) \geq f(\boldsymbol{x}_\alpha) + \frac{\alpha}{2L}\|f'(\boldsymbol{x}) - f'(\boldsymbol{x}_\alpha)\|_2^2 + \frac{1-\alpha}{2L}\|f'(\boldsymbol{y}) - f'(\boldsymbol{x}_\alpha)\|_2^2.$$

Finally, using the inequality

$$\alpha\|\boldsymbol{b} - \boldsymbol{d}\|_2^2 + (1-\alpha)\|\boldsymbol{c} - \boldsymbol{d}\|_2^2 \geq \alpha(1-\alpha)\|\boldsymbol{b} - \boldsymbol{c}\|_2^2$$

we have the result.

$$\left(\begin{array}{l} -\alpha(1-\alpha)\|\boldsymbol{b} - \boldsymbol{c}\|_2^2 \geq -\alpha(1-\alpha)(\|\boldsymbol{b} - \boldsymbol{d}\|_2 + \|\boldsymbol{c} - \boldsymbol{d}\|)_2^2 \\ \text{Therefore} \\ \alpha\|\boldsymbol{b} - \boldsymbol{d}\|_2^2 + (1-\alpha)\|\boldsymbol{c} - \boldsymbol{d}\|_2^2 - \alpha(1-\alpha)(\|\boldsymbol{b} - \boldsymbol{d}\|_2 + \|\boldsymbol{c} - \boldsymbol{d}\|_2)^2 \\ = (\alpha\|\boldsymbol{b} - \boldsymbol{d}\|_2 - (1-\alpha)\|\boldsymbol{c} - \boldsymbol{d}\|_2)^2 \geq 0 \end{array}\right)$$

$\boxed{6\Rightarrow 3}$ Dividing both sides by $1-\alpha$ and tending α to 1, we obtain 3.

$\boxed{2\Rightarrow 7}$ From 2,

$$\begin{aligned}f(\boldsymbol{x}) &\leq f(\boldsymbol{x}_\alpha) + \langle f'(\boldsymbol{x}_\alpha), (1-\alpha)(\boldsymbol{x} - \boldsymbol{y})\rangle + \frac{L}{2}(1-\alpha)^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \\ f(\boldsymbol{y}) &\leq f(\boldsymbol{x}_\alpha) + \langle f'(\boldsymbol{x}_\alpha), \alpha(\boldsymbol{y} - \boldsymbol{x})\rangle + \frac{L}{2}\alpha^2\|\boldsymbol{x} - \boldsymbol{y}\|_2^2\end{aligned}$$

Multiplying the first inequality by α, the second by $1-\alpha$, and summing up, we have

$$\alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \le f(\boldsymbol{x}_\alpha) + \frac{L}{2}\left(\alpha(1-\alpha)^2 + (1-\alpha)\alpha^2\right) \|\boldsymbol{x}-\boldsymbol{y}\|_2^2.$$

The non-negativity follows from Theorem 6.5.

$\boxed{7 \Rightarrow 2}$ Dividing both sides by $1-\alpha$ and tending α to 1, we obtain 2. The non-negativity follows from Theorem 6.5. ∎

6.1 Convex Functions

Theorem 6.9 $f \in \mathcal{F}(\mathbb{R}^n)$ if and only if its epigraph $E := \{(\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \mid f(\boldsymbol{x}) \le y\}$ is a convex.

Proof:

$\boxed{\Rightarrow}$ Let $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2) \in E$. Then for any $0 \le \alpha \le 1$, we have

$$f(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) \le \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2) \le \alpha y_1 + (1-\alpha) y_2$$

and therefore $(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2, \alpha y_1 + (1-\alpha) y_2) \in E$.

$\boxed{\Leftarrow}$ Let $(\boldsymbol{x}_1, f(\boldsymbol{x}_1)), (\boldsymbol{x}_2, f(\boldsymbol{x}_2)) \in E$. By the convexity of E, for any $0 \le \alpha \le 1$,

$$f(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) \le \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2)$$

and therefore, $f \in \mathcal{F}(\mathbb{R}^n)$. ∎

Theorem 6.10 If $f \in \mathcal{F}(\mathbb{R}^n)$, then its λ-*level set* $L_\lambda := \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \le \lambda\}$ is convex for each $\lambda \in \mathbb{R}$. But the converse is not true.

Proof:

For any $\lambda \in \mathbb{R}$, let $\boldsymbol{x}, \boldsymbol{y} \in L_\lambda$. Then for $\forall \alpha \in (0,1)$, since $f \in \mathcal{F}(\mathbb{R}^n)$, $f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \le \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \le \alpha\lambda + (1-\alpha)\lambda = \lambda$. Therefore, $\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in L_\lambda$.

For the converse, $L_\lambda = \{x \in \mathbb{R} \mid f(x) = x^3 \le \lambda\}$ is convex for all $\lambda \in \mathbb{R}$, but $f \notin \mathcal{F}(\mathbb{R})$. ∎

Theorem 6.11 (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

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

Proof:

Left for exercise. ∎

Example 6.12 The function $-\log x$ is convex in $(0, +\infty)$. Let $a, b \in (0, +\infty)$ and $0 \le \alpha \le 1$. Then, from the Jensen's inequality we have

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

If we take the exponential of both sides, we obtain

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

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