6 Differentiable Convex Functions

6.1 Convex Functions

Theorem 6.1 $f \in \mathcal{F}(\mathbb{R}^n)$ if and only if its epigraph $E := \{(\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \mid f(\boldsymbol{x}) \leq y\}$ is a convex.

Proof:

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

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

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 $0 \leq \alpha \leq 1$,

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

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

Theorem 6.2 If $f \in \mathcal{F}(\mathbb{R}^n)$, then its λ-*level set* $L_\lambda := \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \leq \lambda\}$ 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 $f \in \mathcal{F}(\mathbb{R}^n)$, $f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \leq \alpha\lambda + (1-\alpha)\lambda = \lambda$. Therefore, $\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in L_\lambda$.

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

Theorem 6.3 (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

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

Proof:

Left for exercise. ∎

Example 6.4 The function $-\log x$ is convex in $(0, +\infty)$. Let $a, b \in (0, +\infty)$ and $0 \leq \theta \leq 1$. Then, from the Jensen's inequality we have

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

If we take the exponential of both sides, we obtain

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

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

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

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

Then we have

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

and summing over i, we obtain the Hölder inequality:

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

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

Theorem 6.5 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 in $\mathbb{R}^n$.

Proof:
For each $i \in I$, since $f_i \in \mathcal{F}(\mathbb{R}^n)$, its epigraph $E_i = \{(\boldsymbol{x}, y) \in \mathbb{R}^{n+1} \mid f_i(\boldsymbol{x}) \leq y\}$ is convex in $\mathbb{R}^{n+1}$ by Theorem 6.1. Also their intersection

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

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

6.2 Differentiable Convex Functions

Definition 6.6 A continuously differentiable function $f(\boldsymbol{x})$ is called *convex* on $\mathbb{R}^n$ (notation $\mathcal{F}^1(\mathbb{R}^n)$) if

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n.$$

if $-f(\boldsymbol{x})$ is convex, $f(\boldsymbol{x})$ is called *concave.*

Theorem 6.7 If $f \in \mathcal{F}^1(\mathbb{R}^n)$ and $f'(\boldsymbol{x}^*) = 0$, then $\boldsymbol{x}^*$ is the *global minimum* of $f(\boldsymbol{x})$ on $\mathbb{R}^n$.

Proof:
Left for exercise. ∎

Lemma 6.8 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 6.9 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 6.10 Let f be a continuously differentiable function. The following conditions are equivalent:

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

Proof:
Left for exercise. ∎

Theorem 6.11 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

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

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 $0 \le F'(0) = \langle f''(\boldsymbol{x}) \boldsymbol{s}, \boldsymbol{s} \rangle$, and we have the result.

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

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

∎

6.3 Differentiable Convex Functions with Lipschitz Continuous Gradients

Corollary 6.12 Let f be a two times continuously differentiable function. $f \in \mathcal{F}_L^{2,1}(\mathbb{R}^n)$ if and only if $\boldsymbol{O} \preceq f''(\boldsymbol{x}) \preceq L\boldsymbol{I}, \quad \forall \boldsymbol{x} \in \mathbb{R}^n$.

Theorem 6.13 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. $0 \le f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \le \frac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$.
3. $f(\boldsymbol{x}) + \langle f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{1}{2L} \|f'(\boldsymbol{x}) - f'(\boldsymbol{y})\|_2^2 \le f(\boldsymbol{y})$.
4. $0 \le \frac{1}{L} \|f'(\boldsymbol{x}) - f'(\boldsymbol{y})\|_2^2 \le \langle f'(\boldsymbol{x}) - f'(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle$.
5. $0 \le \langle f'(\boldsymbol{x}) - f'(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \rangle \le L \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$.