7 Sugradients of Convex Functions

Definition 7.1 *Let D be a convex set in $\Re^n$.*

(a) *A function $f : D \to \Re$ is said to be* **convex** *if*

$$\boxed{f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)}$$

for all $x, y \in D,\ \ \lambda \in [0, 1]$.

(b) *The function is said to be* **strictly convex** *on D if*

$$\boxed{f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)}$$

for all distinct $x, y \in D,\ \ \lambda \in (0, 1)$.

(c) *A function $g : D \to \Re$ is said to be* **concave (strictly concave)** *if $-g$ is* **convex (strictly convex)** *on D.*

For convex functions, the line segment joining $f(x)$ and $f(y)$ lies above the graph of f in the interval $[x, y]$.

For concave functions, the line segment joining $f(x)$ and $f(y)$ lies below the graph of f in the interval $[x, y]$.

Proposition 7.1 *If $f_1, f_2 : D \to \Re$ are convex functions, then*

$$f_1 + f_2 \text{ is convex on } D;$$
$$\alpha f_1 \text{ is convex for } \alpha \geq 0;$$
$$\alpha f_1 \text{ is concave for } \alpha < 0.$$

Proposition 7.2 $f(x) = \|x\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$, $x \in \Re^n$ *is a convex function.*

Definition 7.2 *Let S be a nonempty set in $\Re^n$, and $f : S \to \Re$. The gradient vector of f at x is the column vector*

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)^T.$$

The Hessian of f at x is the $n \times n$ matrix

$$H(x) = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n x_1} & \frac{\partial^2 f(x)}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

Theorem 7.1 *Suppose that $f : S \to \Re$ has* **continuous second partial derivatives** *in S. Suppose that the line segment $[x, y]$ is contained in the interior of S. Then there exists $w \in [x, y]$ such that*

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2}(y - x)^T H(w)(y - x).$$

Theorem 7.2 *Suppose that $S \subseteq \Re^n$ is a nonempty open convex set and $f : S \to \Re$ has continuous second partial derivatives in S. Then, f is convex on S if and only if the Hessian matrix is positive semidefinite at each point in S.*

For example

$$f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + x_1 + x_2$$

$$\nabla f(x) = \begin{pmatrix} x_1 + 1 \\ x_2 + 1 \end{pmatrix}$$

$$\nabla^2 f(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \succeq 0\,.$$

Therefore, f is convex.

Corollary 7.1 *Suppose that $S \subseteq \Re^n$ is a nonempty open convex set and $f : S \to \Re$ has continuous second partial derivaties in S. Then for any $x, y \in S$,*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)\,.$$

Definition 7.3 *Let D be a nonempty convex set in $\Re^n$, and let $f : D \to \Re$ be convex. Then ξ is called a* **subgradient of** f *at $\bar{x} \in D$ if*

$$f(x) \geq f(\bar{x}) + \xi^T (x - \bar{x}) \qquad \text{for all } x \in D$$

Similarly, let $f : D \in \Re$ be concave, then ξ is called a subgradient of f at $\bar{x} \in D$ if

$$f(x) \leq f(\bar{x}) + \xi^T (x - \bar{x}) \qquad \text{for all } x \in D$$

The collection of subgradients of f at $\bar{x}$ is called the **subdifferential of** f *at $\bar{x}$, denoted by $\partial f(\bar{x})$. Obviously, $\partial f(\bar{x})$ is a convex set.*

Theorem 7.3 *Let S be a nonempty convex set in $\Re^n$, and let $f : S \to \Re$ be convex. Then, for $\bar{x} \in \operatorname{int} S$, $\partial f(\bar{x}) \neq \emptyset$. In particular, if $S = \Re^n, \partial f(\bar{x}) \neq \emptyset \ \forall \bar{x} \in \Re^n$.*

Theorem 7.4 *Let S be a nonempty convex set in $\Re^n$, and let $f : S \to \Re$ be convex. Suppose that f is differentiable at $\bar{x} \in \operatorname{int} S$. Then $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.*

Proof. First, since $\partial f(\bar{x}) \neq \emptyset$, let $\xi \in \partial f(\bar{x})$. Then for any d and λ sufficiently small,

$$f(\bar{x} + \lambda d) \geq f(\bar{x}) + \lambda \xi^T d$$

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x})^T d + \lambda ||d|| \alpha(\bar{x}; \lambda d)$$

where $\alpha(\bar{x}; \lambda d) \to 0$ as $\|\lambda d\| \to 0$.

Subtracting the equation from the inequality, we obtain

$$0 \geq \lambda[\xi - \nabla f(\bar{x})]^T d - \lambda ||d|| \alpha(\bar{x}; \lambda d),$$

which implies $[\xi - \nabla f(\bar{x})]^T d \leq 0$. Let $d = \xi - \nabla f(\bar{x})$. $\Longrightarrow \xi = \nabla f(\bar{x})$.

Theorem 7.5 *Let $f : \Re^n \to \Re$ be a convex function. Then*

$$f(\bar{x}) \leq f(x) \text{ for all } x \in \Re^n \text{ iff } 0 \in \partial f(\bar{x}).$$

Example 1.

$$f(x) = |x|, \; \partial f(x) = \begin{cases} \{1\} & \text{if } x > 0 \\ \{-1\} & \text{if } x < 0 \\ [-1, 1] & \text{if } x = 0 \end{cases}$$

Example 2. $f(x) = \sqrt{x_1^2 + x_2^2}$, $x \in \Re^2$. Then

(i) $\bar{x} \neq 0$,

$$\partial f(\bar{x}) = \{\nabla f(\bar{x})\} = \left(\frac{x_1}{\sqrt{x_1^2 + x_2^2}}, \frac{x_2}{\sqrt{x_1^2 + x_2^2}} \right)^T.$$

(ii) $\bar{x} = (0,0)^T$. Let $\xi \in \partial f(\bar{x})$. Then

$$\begin{aligned}
& f(x) - f(\bar{x}) \geq \xi^T(x - \bar{x}) \quad \forall x \in \Re^2 \\
\Longleftrightarrow & \sqrt{x_1^2 + x_2^2} \geq \xi_1 x_1 + \xi_2 x_2 \quad \forall\, x \in \Re^2 \\
\Longleftrightarrow & \sqrt{x_1^2 + x_2^2} \geq \xi_1 x_1 + \xi_2 x_2 \quad \forall\, \|x\| \leq 1 \\
\Longleftrightarrow & \sqrt{x_1^2 + x_2^2} \geq \xi_1 x_1 + \xi_2 x_2 \quad \forall\, \|x\| = 1 \\
\Longleftrightarrow & 1 \geq \xi_1 x_1 + \xi_2 x_2 \quad \forall\, \|x\| = 1 \\
\Longleftrightarrow & \xi_1^2 + \xi_2^2 \leq 1\,.
\end{aligned}$$

Therefore,

$$\partial f(\bar{x}) = \{\xi \in \Re^2 \,|\, \xi_1^2 + \xi_2^2 \leq 1\}\,.$$

Theorem 7.6 *Let S be a nonempty convex set in $\Re^n$, and let $f, g : S \to \Re$ be convex functions. Then $h(x) = \max\{f(x), g(x)\}$ is a convex function.*

Proof. $\forall\, x, y \in S, \lambda \in [0,1]$,

$$\begin{aligned}
& h(\lambda x + (1-\lambda)y) \\
\leq\ & \max\{\lambda f(x) + (1-\lambda) f(y),\ \lambda g(x) + (1-\lambda) g(y)\} \\
\leq\ & \lambda \max\{\, f(x), g(x)\,\} + (1-\lambda) \max\{f(y), g(y)\} \\
=\ & \lambda h(x) + (1-\lambda) h(y)\,.
\end{aligned}$$

Theorem 7.7 *Suppose that $f, g : \Re^n \to \Re$ are continuously differentiable convex functions. Let*

$$h(x) = \max\{\, f(x), g(x)\,\}\,.$$

Then

$$\partial h(x) = \begin{cases} \{\,\nabla f(x)\,\} & \text{if } f(x) > g(x) \\ \operatorname{conv}\{\,\nabla f(x), \nabla g(x)\,\} & \text{if } f(x) = g(x) \\ \{\,\nabla g(x)\,\} & \text{if } f(x) < g(x)\,. \end{cases}$$