Subgradients

Instructor: Jin Zhang

Department of Mathematics
Southern University of Science and Technology

Fall 2023

Contents

Definitions

- **(subgradient)** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and let $x \in \mathrm{dom}(f)$. Then $g \in \mathbb{R}^n$ is called a subgradient of f at x if

$$f(y) \geq f(x) + \langle g, y - x \rangle \text{ for all } y \in \mathbb{R}^n.$$

- **(subdifferential)** The set of all subgradients of f at $x \in \mathrm{dom}(f)$ is called the subdifferential of f at x and is denoted by $\partial f(x)$:

$$\partial f(x) \equiv \{ g \in \mathbb{R}^n \ : \ f(y) \geq f(x) + \langle g, y - x \rangle \text{ for all } y \in \mathbb{R}^n \}.$$

  When $x \notin \mathrm{dom}(f)$, we define $\partial f(x) = \varnothing$.

f -> ∞

$\|y\| \geq 0 + \langle g, y-0 \rangle$, $\|g\|_* \|y\| \geq \langle g, y \rangle$

- Example 1 (subdifferential of norms at 0). Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \|x\|$, where $\|\cdot\|$ is the endowed norm on $\mathbb{R}^n$. Then the subdifferential of f at $x = 0$ is the dual norm unit ball:

$$\partial f(0) = \mathbb{B}_{\|\cdot\|_*}[0,1] = \{g \in \mathbb{R}^n \, : \, \|g\|_* \leq 1\}.$$

- Example 2 (subdifferential of the l_1-norm at 0). Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \|x\|_1$. Then

$$\partial f(0) = \mathbb{B}_{\|\cdot\|_\infty}[0,1] = [-1,1]^n.$$

- Definition (normal cone). Given a set $S \subseteq \mathbb{R}^n$ and a point $x \in S$, the normal cone of S at x is defined as

$$N_S(x) = \{y \in \mathbb{R}^n : \langle y, z - x \rangle \leq 0 \text{ for any } z \in S\}.$$

When $x \notin S$, we define $N_S(x) = \varnothing$.

- Example 3 (subdifferential of indicator functions). Suppose that $S \subset \mathbb{R}^n$ is nonempty and consider the indicator function δ_S. Then for any $x \in S$, we have that

$$\partial \delta_S(x) = N_S(x).$$

- Example 4 (subdifferential of the indicator functions of the unit ball). Let

$$S = \mathbb{B}[0,1] = \{x \in \mathbb{R}^n : \|x\| \leq 1\}.$$

Then

$$\partial\delta_{\mathbb{B}[0,1]} = N_{\mathbb{B}[0,1]}(x) = \begin{cases} \{y \in \mathbb{R}^n : \|y\|_* \leq \langle y, x\rangle\}, & \|x\| \leq 1, \\ \varnothing, & \|x\| > 1. \end{cases}$$

- Example 5 (subgradient of the dual function). Consider the minimization problem

$$\min \{f(x) : g(x) \leq 0, x \in X\},$$

where $\varnothing \neq X \subseteq \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ is vector-valued. The Lagrangian dual objective function is given by

$$q(\lambda) = \min_{x \in X} \left\{ L(x; \lambda) \equiv f(x) + \lambda^T g(x) \right\}.$$

The effective domain of $-q$ is given by

$$\mathrm{dom}(-q) = \left\{ \lambda \in \mathbb{R}^m_+ : q(\lambda) > -\infty \right\}.$$

No matter whether the primal problem is convex or not, the the dual problem

$$\max_{\lambda \in \mathbb{R}^m} \{q(\lambda) \,:\, \lambda \in \mathsf{dom}(-q)\}$$

is always convex.
Let $\lambda_0 \in \mathsf{dom}(-q)$ and assume that the minimum in the minimization problem defining $q(\lambda_0)$,

$$q(\lambda_0) = \min_{x \in X} \left\{ f(x) + \lambda_0^T g(x) \right\},$$

is attained at $x_0 \in X$.
We seek to find a subgradient of the convex function $-q$ at λ_0:

$$-g(x_0) \in \partial(-q)(\lambda_0).$$

- Example 6 (subgradient of the maximum eigenvalue function). Consider the function $f : \mathbb{S}^n \to \mathbb{R}$ given by $f(X) = \lambda_{\mathsf{max}}(X)$ (where $\mathbb{S}^n$ is the set of all $n \times n$ symmetric matrices). Let $X \in \mathbb{S}^n$ and let v be a normalized eigenvector of X $(\|v\|_2 = 1)$ associated with the maximum eigenvalue of X. Then

$$vv^T \in \partial f(X).$$

2. Properties of the Subdifferential Set

Recall that the subdifferential of f at x is denoted by

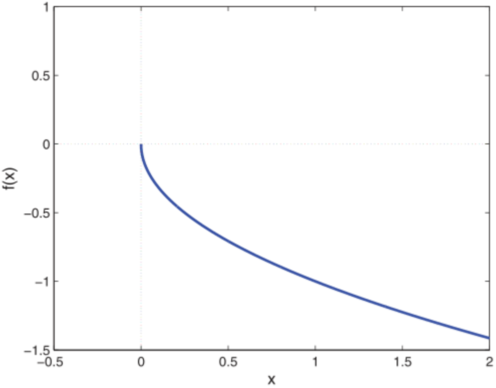
$$\partial f(x) \equiv \{g \in \mathbb{R}^n \ : \ f(y) \geq f(x) + \langle g, y - x \rangle \text{ for all } y \in \mathbb{R}^n\}.$$

- **Theorem (closedness and convexity of the subdifferential set).** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper. Then the set $\partial f(x)$ is closed and convex for any $x \in \mathbf{R}^n$.

- **Definition (subdifferentiability).** A proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called subdifferentiable at $x \in \mathrm{dom}\,(f)$ if $\partial f(x) \neq \varnothing$.
  The collection of points of subdifferentiability is denoted by $\mathrm{dom}\,(\partial f)$:
$$\mathrm{dom}(\partial f) = \{x \in \mathbb{R}^n : \partial f(x) \neq \varnothing\}.$$
- **Lemma (nonemptiness of subdifferential sets $\Rightarrow$ convexity).** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and assume that $\mathrm{dom}(f)$ is convex. Suppose that for any $x \in \mathrm{dom}(f)$, the set $\partial f(x)$ is nonempty. Then f is convex.

- **Example (A convex function, which is not subdifferentiable at one of the points in its domain).** Consider the convex function $f : \mathbb{R} \to \overline{\mathbb{R}}$ defined by

$$f(x) = \begin{cases} -\sqrt{x}, & x \geq 0, \\ \infty, & \text{else.} \end{cases}$$



It is not subdifferentiable at $x = 0$

- **Theorem (supporting hyperplane theorem).** Let $\emptyset \neq C \subseteq \mathbb{R}^n$ be convex, and let $y \notin \text{int}(C)$. Then there exists $0 \neq p \in \mathbb{R}^n$ such that

$$\langle p, x \rangle \leq \langle p, y \rangle \text{ for any } x \in C.$$

- **Theorem (nonemptiness and boundedness of the subdifferential set at interior points of the domain).** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex, and assume that $\tilde{x} \in \operatorname{int}(\operatorname{dom}(f))$. Then $\partial f(\tilde{x})$ is nonempty and bounded.
- **Corollary (subdifferentiability of real-valued convex functions).** Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then f is subdifferentiable over $\mathbb{R}^n$.

- **Theorem (boundedness of subgradients over compact sets).** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex, and assume that $X \subseteq \mathrm{int}\,(\mathrm{dom}(f))$ is nonempty and compact. Then $Y = \bigcup_{x \in X} \partial f(x)$ is nonempty and bounded.

Recall that the relative interior of a convex set $S \subseteq \mathbb{R}^n$ is denoted by

$$\mathrm{ri}(S) = \{x \in \mathrm{aff}(S) : \exists V \text{ be some neighborhood of } x \text{ s.t. } V \cap \mathrm{aff}(S) \subseteq S\}.$$

- **Theorem (nonemptiness of the relative interior).**
  Let $\varnothing \neq C \subseteq \mathbb{R}^n$ be convex. Then $\mathrm{ri}(C)$ is nonempty.

- **Theorem (nonemptiness of the subdifferential set at relative interior points).** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex, and let $\tilde{x} \in \text{ri}\,(\text{dom}(f))$. Then $\partial f(\tilde{x})$ is nonempty.
- **Corollary.** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex. Then there exists $x \in \text{dom}(f)$ for which $\partial f(x)$ is nonempty.

- **Theorem (unboundedness of the subdifferential set when $\mathbf{dim}\,(\mathbf{dom}(f)) < n$).** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex. Suppose that $\dim\,(\mathrm{dom}(f)) < n$ and let $x \in \mathrm{dom}(f)$. If $\partial f(x) \neq \varnothing$, then $\partial f(x)$ is unbounded.

3.1 Definition and Basic Properties

- **Definition.** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper. The directional derivative of f at $x \in \operatorname{int}(\operatorname{dom}(f))$ in a given direction $d \in \mathbb{R}^n$, if it exists, is defined by
$$f'(x; d) \equiv \lim_{\alpha \to 0^+} \frac{f(x + \alpha d) - f(x)}{\alpha}.$$
- **Theorem.** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex and let $x \in \operatorname{int}(\operatorname{dom}(f))$. Then for any $d \in \mathbb{R}^n$, the directional derivative $f'(x; d)$ exists.

- **Lemma (convexity and homogeneity of $d \mapsto f'(x;d)$).** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex and let $x \in \operatorname{int}(\operatorname{dom}(f))$. Then
  (a). the function $d \mapsto f'(x;d)$ is convex;
  (b). for any $\lambda \geq 0$ and $d \in \mathbb{R}^n$, it holds that $f'(x;\lambda d) = \lambda f'(x;d)$.
- **Lemma.** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex, and let $x \in \operatorname{int}(\operatorname{dom}(f))$. Then

$$f(y) \geq f(x) + f'(x;y-x) \text{ for all } y \in \operatorname{dom}(f).$$

- **Theorem (directional derivative of maximum of functions).** Suppose that

$$f(x) = \max\{f_1(x), f_2(x), \cdots, f_m(x)\},$$

where $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \overline{\mathbb{R}}$ are proper. Let $x \in \bigcap_{i=1}^m \operatorname{int}(\operatorname{dom}(f))$ and $d \in \mathbb{R}^n$. Assume that ${f_i}'(x; d)$ exist for any $i \in \{1, 2, \cdots, m\}$. Then

$$f'(x; d) = \max_{i \in I(x)} {f_i}'(x; d),$$

where $I(x) = \{i : f_i(x) = f(x)\}$.

- **Corollary (directional derivative of maximum of functions -convex case).**
  Suppose that

$$f(x) = \max\{f_1(x), f_2(x), \cdots, f_m(x)\},$$

where $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \overline{\mathbb{R}}$ are proper convex. Let $x \in \bigcap_{i=1}^{m} \operatorname{int}(\operatorname{dom}(f))$ and $d \in \mathbb{R}^n$. Then

$$f'(x; d) = \max_{i \in I(x)} f_i'(x; d),$$

where $I(x) = \{i : f_i(x) = f(x)\}$.

- **Theorem (max formula).** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex. Then for any $x \in \text{int}\,(\text{dom}(f))$ and $d \in \mathbb{R}^n$,

$$f'(x; d) = \max\left\{\langle g, d\rangle \ : \ g \in \partial f(x)\right\}.$$

- **Remark.** The max formula can also be rewritten using the support function notation as follows:

$$f'(x; d) = \sigma_{\partial f(x)}(d).$$

- **Definition (differentiability).** The function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be differentiable at $x \in \text{int}(\text{dom}(f))$ if there exists $g \in \mathbb{R}^n$ such that

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - \langle g, h \rangle}{\|h\|} = 0.$$

  In this case it's obvious that g is uniquely defined. The unique vector g is called the gradient of f at x and is denoted by $\nabla f(x)$.

- **Theorem (directional derivatives at points of differentiability).** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, and suppose that f is differentiable at $x \in \text{int}\,(\text{dom}(f))$. Then for any $d \in \mathbb{R}^n$

$$f'(x; d) = \langle \nabla f(x), d \rangle.$$

- **Example 1 (directional derivative of maximum of differentiable function).** Consider the function

$$f(x) = \max\left\{f_1(x), f_2(x), \cdots, f_m(x)\right\},$$

where $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \overline{\mathbb{R}}$ are proper and differentiable at a given point $x \in \bigcap_{i=1}^m \mathrm{int}\,(\mathrm{dom}(f))$. Then for any $d \in \mathbb{R}^n$, $f'(x; d)$ exists and

$$f'(x; d) = \max_{i \in I(x)} \langle \nabla f_i(x), d \rangle,$$

where $I(x) = \{i \,:\, f_i(x) = f(x)\}$.

- **Example 2 (gradient of $\frac{1}{2}d_C^2(\cdot)$).** Let $\Omega \subseteq \mathbb{R}^n$, then P_Ω is the so-called orthogonal projection mapping defined by

$$P_\Omega(x) \equiv \arg\min_{y\in\Omega} \|y - x\| .$$

It is well known that P_Ω is well-defined when Ω is nonempty closed and convex.
Let $C \subseteq \mathbb{R}^n$ be nonempty closed and convex. Consider the function $\varphi_C : \mathbb{R}^n \to \mathbb{R}$ given by

$$\varphi_C(x) \equiv \frac{1}{2}d_C^2(x) = \frac{1}{2}\|x - P_C(x)\|^2 .$$

Then for any $x \in \mathbb{R}^n$,

$$\nabla\varphi_C(x) = x - P_C(x).$$

- **Theorem (the subdifferential at points of differentiability).** Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex, and let $x \in \operatorname{int}(\operatorname{dom}(f))$. If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$. Conversely, if f has a unique subgradient at x, then it is differentiable at x and $\partial f(x) = \{\nabla f(x)\}$.

- **Example 3 (subdifferential of the l_2-norm).** Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \|x\|_2$. Then

$$\partial f(x) = \begin{cases} \left\{ \frac{x}{\|x\|_2} \right\}, & x \neq 0, \\ \mathbb{B}_{\|\cdot\|_2}[0,1], & x = 0. \end{cases}$$

Theorem. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and let $\alpha > 0$. Then for any $x \in \mathrm{dom}(f)$

$$\partial(\alpha f)(x) = \alpha \partial f(x).$$

4. Computing Subgradients
4.2 Summation

Theorem. Let $f_1, f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex, and let $x \in \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2)$.

(a). The following inclusion holds:

$$\partial f_1(x) + \partial f_2(x) \subseteq \partial(f_1 + f_2)(x).$$

(b). If $x \in \mathrm{int}\,(\mathrm{dom}(f_1)) \cap \mathrm{int}\,(\mathrm{dom}(f_1))$, then

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x).$$

Corollary. Let $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex, and let $x \in \bigcap_{i=1}^m \mathrm{dom}(f_i)$.

(a). **(weak sum rule of subdifferential calculus)** The following inclusion holds:

$$\sum_{i=1}^m \partial f_i(x) \subseteq \partial \left(\sum_{i=1}^m f_i \right)(x).$$

(b). If $x \in \bigcap_{i=1}^m \mathrm{int}\,(\mathrm{dom}(f_i))$, then

$$\partial \left(\sum_{i=1}^m f_i \right)(x) = \sum_{i=1}^m \partial f_i(x).$$

Corollary. Let $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \mathbb{R}$ be real-valued convex. Then for any $x \in \mathbf{R}^n$

$$\partial \left(\sum_{i=1}^m f_i \right)(x) = \sum_{i=1}^m \partial f_i(x).$$

Theorem (sum rule of subdifferential calculus). Let $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex, and assume that $\bigcap_{i=1}^m \operatorname{ri}(\operatorname{dom}(f_i)) \neq \varnothing$. Then for any $x \in \mathbb{R}^n$

$$\partial \left(\sum_{i=1}^m f_i \right)(x) = \sum_{i=1}^m \partial f_i(x).$$

Example 1 (subdifferential set of the l_1-norm function). Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = \|x\|_1$.

(a). **(strong result).** Let

$$I_{\neq}(x) = \{i : x_i \neq 0\}, I_0(x) = \{i : x_i = 0\}.$$

Then

$$\partial f(x) = \{z \in \mathbb{R}^n : z_i = \mathrm{sgn}(x_i), i \in I_{\neq}(x), |z_j| \leq 1, j \in I_0(x)\}.$$

(b). **(weak result).** We have

$$\mathrm{sgn}(x) \in \partial f(x).$$

Theorem. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex and $\mathcal{A} : \mathbb{R}^p \to \mathbb{R}^n$ be a linear transformation. Let $h(x) = f(\mathcal{A}(x) + b)$ with $b \in \mathbb{R}^n$. Assume that h is proper.

(a). **(weak affine transformation rule of subdifferential calculus).**
For any $x \in \mathrm{dom}(h)$,

$$\mathcal{A}^T(\partial f(\mathcal{A}(x) + b)) \subseteq \partial h(x).$$

(b). **(affine transformation of subdifferential calculus).**
If $x \in \mathrm{int}(\mathrm{dom}(h))$ and $\mathcal{A}(x) + b \in \mathrm{int}(\mathrm{dom}(f))$, then

$$\partial h(x) = \mathcal{A}^T(\partial f(\mathcal{A}(x) + b)).$$

Example 2 (subdifferential of $\|Ax+b\|_1$). Let $f:\mathbb{R}^n\to\mathbb{R}$ be the function given by $f(x)=\|Ax+b\|_1$, where $A=(a_1,\cdots,a_m)^T\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$. We conclude that

$$\partial f(x)=\sum_{i\in I_{\neq}(x)}\operatorname{sgn}\left(a_i^T x+b_i\right)a_i+\sum_{i\in I_0(x)}\left[-a_1,a_i\right],$$

where

$$\begin{aligned} I_{\neq}(x) &= \left\{i\,:\,a_i^T x+b_i\neq 0\right\} \\ I_0(x) &= \left\{i\,:\,a_i^T x+b_i=0\right\}. \end{aligned}$$

Example 3 (subdifferential of $\|Ax+b\|_2$). Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be the function given by $f(x) = \|Ax+b\|_2$, where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. We conclude that

$$\partial f(x) = \begin{cases} \frac{A^T(Ax+b)}{\|Ax+b\|_2}, & Ax+b \neq 0, \\ A^T \mathbb{B}_{\|\cdot\|_2}[0,1], & Ax+b = 0. \end{cases}$$

Theorem. Suppose that f is continuous on $[a, b]$ $(a < b)$ and that $f'_+(a)$ exists. Let g be a function defined on an open interval I which contains the range of f, and assume that g is differentiable at $f(a)$. Then the function

$$h(t) = g\left(f(t)\right) \quad (a \leq t \leq b)$$

is right differentiable at $t = a$ and

$$h'_+(a) = g'\left(f(a)\right) f'_+(a).$$

Theorem (chain rule of subdifferential calculus). Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and $g : \mathbb{R} \to \mathbb{R}$ be nondecreasing convex. Let $x \in \mathbb{R}^n$, and suppose that g is differentiable at the point $f(x)$. Let $h = g \circ f$. Then

$$\partial h(x) = g'\,(f(x))\,\partial f(x).$$

Example 4 (subdifferential of $\|\cdot\|_1^2$). Consider the function $h : \mathbb{R}^n \to \mathbb{R}$ given by $h(x) = \|x\|_1^2$. Then

$$\partial h(x) = 2\|x\|_1 \{z \in \mathbb{R}^n \,:\, z_i = \mathsf{sgn}(x_i), i \in I_{\neq}(x), |z+_j| \leq 1, j \in I_0(x)\},$$

where $I_{\neq}(x) = \{i \,:\, x_i \neq 0\}$, $I_0(x) = \{i \,:\, x_i = 0\}$.

Example 5 (subdifferential of $d_C(\cdot)$). Let $C \subseteq \mathbb{R}^n$ be nonempty closed and convex. Then we have

$$\partial d_C(x) = \begin{cases} N_C(x) \cap \mathbb{B}[0,1] & \text{if } x \in C. \\ \left\{ \frac{x - P_C(x)}{d_C(x)} \right\} & \text{otherwise }. \end{cases}$$

Theorem (max rule of subdifferential calculus). Let $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex, and define

$$f(x) = \max\{f_1(x), f_2(x), \cdots, f_m(x)\}.$$

Let $x \in \cap_{i=1}^m \text{int}(\text{dom}(f_i))$. Then

$$\partial f(x) = \text{co}\left(\bigcup_{i \in I(x)} \partial f_i(x)\right),$$

where $I(x) = \{i = 1, 2, \cdots, m : f_i(x) = f(x)\}$.

Example 6 (subdifferential of the max function). Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \max\{x_1, x_2, \cdots, x_n\}$. Denote

$$I(x) = \{i \,:\, f(x) = x_i\}.$$

Then

$$\partial f(x) = \left\{ \sum_{i \in I(x)} \lambda_i e_i \,:\, \sum_{i \in I(x)} \lambda_i = 1, \lambda_j \geq 0, j \in I(x). \right\}$$

Example 7 (subdifferential of the l_∞-norm). Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \|x\|_\infty$. Denote

$$I(x) = \{i \,:\, |x_i| = \|x\|_\infty\}.$$

Then

$$\partial f(x) = \begin{cases} \mathbb{B}_{\|\cdot\|_1}[0,1], & x = \\ \left\{\sum_{i\in I(x)} \lambda_i \mathsf{sgn}(x_i) e_i \,:\, \sum_{i\in I(x)} \lambda_i = 1, \lambda_j \geq 0, j \in I(x)\right\}, & x \neq \end{cases}$$

Example 8 (subdifferential of piecewise linear functions). Consider the piecewise linear function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(x) = \max_{i=1,2,\cdots,m} \left\{ a_i^T x + b_i \right\},$$

where $a_i \in \mathbb{R}^n, b_i \in \mathbb{R}, i = 1, 2, \cdots, m$. Then

$$\partial f(x) = \left\{ \sum_{i \in I(x)} \lambda_i a_i \ : \ \sum_{i \in I(x)} \lambda_i = 1, \lambda_j \geq 0, j \in I(x) \right\},$$

where $I(x) = \left\{ i \ : \ f(x) = a_i^T x + b_i \right\}$.

Theorem (weak maximum rule of subdifferential calculus). Let I be an arbitrary set, and suppose that any $i \in I$ is associated with a proper convex function $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$. Let

$$f(x) = \max_{i \in I} f_i(x).$$

Then for any $x \in \mathrm{dom}(f)$

$$\mathrm{co}\left(\bigcup_{i \in I(x)} \partial f_i(x)\right) \subseteq \partial f(x),$$

where $I(x) = \{i \in I \,:\, f(x) = f_i(x)\}$.

Consider the minimization problem

$$\min_{x \in X} \{f(x) \,:\, g(x) \le 0, Ax + b = 0\}, \tag{1}$$

where $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $g = (g_1, g_2, \cdots, g_m)^T : \mathbb{R}^n \to \overline{\mathbb{R}}^m$, $\varnothing \neq X \subseteq \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$, and $b \in \mathbb{R}^p$.
The value function associated with Problem (1) is the function $v : \mathbb{R}^m \times \mathbb{R}^p \to [-\infty, \infty]$ given by

$$v(u, t) = \min_{x \in X} \{f(x) \,:\, g(x) \le u, Ax + b = t\}. \tag{2}$$

The feasible set of the minimization problem in (2) will be denoted by

$$C(u, t) = \{x \in X \,:\, g(x) \le u, Ax + b = t\}.$$

The value function can be rewritten as $v(u, t) = \min\{f(x) \,:\, x \in C(u, t))\}$. By convention $v(u, t) = \infty$ if $C(u, t) = \varnothing$.

Lemma. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $g : \mathbb{R}^n \to \overline{\mathbb{R}}^m$, $\varnothing \neq X \subseteq \mathbb{R}^n$, $A \in \mathbb{R}^{p \times n}$, and $b \in \mathbb{R}^p$. Let v be the value function of Problem (1).

(a). **(monotonicity of the value function).** Then

$$v(u,t) \geq v(w,t) \text{ for any } u, w \in \mathbb{R}^m, t \in \mathbb{R}^p \text{ satisfying } u \leq w.$$

(b). **(convexity of the value function).** Moreover, let $f, g_1, g_2, \cdots, g_m$ be convex functions and X be convex set. Suppose that the value function v is proper. Then v is convex over $\mathbb{R}^m \times \mathbb{R}^p$.

Theorem (Lipschitz continuity and boundedness of the subdifferential sets). Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Suppose that $X \subseteq \text{int}\,(\text{dom}(f))$. Consider the following two claims:

(i). $|f(x) - f(y)| \leq L\,\|x - y\|$ for any $x, y \in X$.

(ii). $\|g\|_* \leq L$ for any $g \in \partial f(x)$, $x \in X$.

Then

(a). the implication $(i) \Rightarrow (ii)$ holds;

(b). if X is open, then (i) holds if and only if (ii) holds.

Corollary (Lipschitz continuity of convex functions over compact domains).
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Suppose that $X \subseteq \text{int}\,(\text{dom}(f))$ is compact. Then there exists $L > 0$ for which

$$|f(x) - f(y)| \leq L \|x - y\| \text{ for any } x, y \in X.$$

Theorem (Fermat's optimality condition). Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex. Then

$$x^* \in \arg\min \{f(x) \,:\, x \in \mathbb{R}^n\}$$

if and only if $0 \in \partial f(x^*)$.

Example 1 (minimizing piecewise linear functions). Consider the problem

$$\min_{x\in\mathbb{R}^n}\left[f(x)\equiv\max_{i=1,2,\cdots,m}\left\{a_i^Tx+b_i\right\}\right],$$

where $a_i\in\mathbb{R}^n$, $b_i\in\mathbb{R}$, $i=1,2,\cdots,m$. Denote

$$I(x)=\left\{i\,:\,f(x)=a_i^Tx+b_i\right\}.$$

Then x^* is an optimal solution if and only if

$$\exists\lambda\in\Delta_m \text{ s.t. } A^T\lambda=0 \text{ and } \lambda_j\left(a_j^Tx^*+b_j-f(x^*)\right), j=1,2,\cdots,m.$$

Theorem (necessary and sufficient optimality conditions for convex constrained optimization).
Let $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex, and let $C \subseteq \mathbb{R}^n$ be convex for which $\operatorname{ri}(\operatorname{dom}(f)) \cap \operatorname{ri}(C) \neq \varnothing$. Then $x^* \in C$ is an optimal solution of

$$\min\{f(x) : x \in C\}$$

if and only if

$$\text{there exists } g \in \partial f(x^*) \text{ for which } -g \in N_C(x^*).$$

Corollary (necessary and sufficient optimality conditions for convex constrained optimization-second version).
Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex, and let $C \subseteq \mathbb{R}^n$ be convex for which $\mathrm{ri}\,(\mathrm{dom}(f)) \cap \mathrm{ri}\,(C) \neq \varnothing$. Then $x^* \in C$ is an optimal solution of

$$\min\{f(x) \,:\, x \in C\}$$

if and only if

there exists $g \in \partial f(x^*)$ for which $\langle g, x - x^* \rangle \geq 0$ for any $x \in C$.

Theorem (optimality conditions for the composite problem). Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, and let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex such that $\text{dom}(g) \subseteq \text{int}\,(\text{dom}(f))$. Consider the problem

$$(P) \quad \min_{x \in \mathbb{R}^n} f(x) + g(x).$$

(a). **(necessary condition).** If $x^* \in \text{dom}(g)$ is a local optimal solution of (P) and f is differentiable at x^*, then

$$-\nabla f(x^*) \in \partial g(x^*).$$

(b). **(necessary and sufficient condition for convex problem).** Suppose that f is convex. If f is differentiable at $x^* \in \text{dom}(g)$, then x^* is a global optimal solution of (P) if and only if

$$-\nabla f(x^*) \in \partial g(x^*).$$

Definition (stationary). Let Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, and let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper convex such that $\mathrm{dom}(g) \subseteq \mathrm{int}\,(\mathrm{dom}(f))$. Consider the problem

$$(P) \quad \min_{x \in \mathbb{R}^n} f(x) + g(x).$$

A point x^* in which f is differentiable is called a stationary point of (P) if

$$-\nabla f(x^*) \in \partial g(x^*).$$

Example 2 (convex constrained nonconvex programming). When $g = \delta_C$ for a nonempty convex set $C \subseteq \mathbb{R}^n$, Problem (P) becomes

$$\min \{ f(x) \, : \, x \in C \} \, .$$

A point $x^* \in C$ in which f is differentiable is a stationary point if and only if

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0 \text{ for any } x \in C.$$

Example 3. Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x) + \lambda \|x\|_1 ,$$

where $\lambda \geq 0$. A point $x^* \in \operatorname{int}(\operatorname{dom}(f))$ in which f is differentiable is a stationary point if and only if

$$\frac{\partial f(x^*)}{\partial x_i} \begin{cases} = -\lambda, & x_i^* > 0, \\ = \lambda, & x_i^* < 0, \\ \in [-\lambda, \lambda], & x_i^* = 0. \end{cases}$$

Lemma. Let $f, g_1, g_2, \cdots, g_m : \mathbb{R}^n \to \mathbb{R}$ be real-valued. Consider the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \cdots, m. \end{aligned}$$

Assume that the minimum value of the above problem is finite and equal to $\bar{f}$. Define the function

$$F(x) \equiv \max\left\{f(x) - \bar{f}, g_1(x), g_2(x), \cdots, g_m(x)\right\}.$$

Then the optimal set of the above inequalities constrained problem is the same as the set of minimizers of F.

Theorem (Fritz-John necessary optimality conditions). Consider the minimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \cdots, m. \end{aligned}$$

where $f, g_1, g_2, \cdots, g_m : \mathbb{R}^n \to \mathbb{R}$ be real-valued convex. Let x^* be an optimal solution. Then there exist $\lambda_0, \lambda_1, \cdots, \lambda_m \geq 0$, not all zeros, for which

$$\begin{aligned} 0 \in \lambda_0 \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*) \\ \lambda_i g_i(x^*) = 0, \quad i = 1, 2, \cdots, m. \end{aligned}$$

Theorem (KKT conditions). Consider the minimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, 2, \cdots, m. \end{aligned}$$

where $f, g_1, g_2, \cdots, g_m : \mathbb{R}^n \to \mathbb{R}$ be real-valued convex.

(a). Let x^* be an optimal solution, and assume that Slater's condition

there exists $\bar{x} \in \mathbb{R}^n$ for which $g_i(\bar{x}) < 0, \quad i = 1, 2, \cdots, m.$

is satisfied. Then there exist $\lambda_1, \cdots, \lambda_m \geq 0$ for which

$$0 \in \partial f(x^*) + \sum_{i=1}^{m} \lambda_i \partial g_i(x^*)$$
$$\lambda_i g_i(x^*) = 0, \quad i = 1, 2, \cdots, m.$$

(b). If x^* satisfies both above conditions for some $\lambda_1, \cdots, \lambda_m \geq 0$, then it is an optimal solution.