Lecture 8: Preliminaries IV. Convex Analysis

School of Mathematical Sciences, Xiamen University

1. Notation

- $\mathbb{R}^n$: n-dimensional real Euclidean space with inner product $\langle \cdot, \cdot \rangle$:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

- Givens sets $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{A} \subseteq \mathcal{B}$ denotes that $\mathcal{A}$ is a subset (possibly equal to) $\mathcal{B}$, and $\mathcal{A} \subset \mathcal{B}$ means that $\mathcal{A}$ is a strict subset of $\mathcal{B}$. $\operatorname{int} \mathcal{A}$ and $\operatorname{cl} \mathcal{A}$ denote the interior and the closure of $\mathcal{A}$, respectively.
- Given a norm $\|\cdot\|$, its dual norm $\|\cdot\|_*$ is defined as

$$\|\mathbf{z}\|_* := \sup\{\langle \mathbf{z}, \mathbf{x} \rangle \mid \|\mathbf{x}\| \leq 1\}.$$

  We have

$$\|\mathbf{x}\| = \sup\{\langle \mathbf{z}, \mathbf{x} \rangle \mid \|\mathbf{z}\|_* \leq 1\}.$$

- ℓ_p norm $(1 \leq p \leq \infty)$:

$$\|\mathbf{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}, \quad p \in [1, \infty), \quad \|\mathbf{x}\|_\infty = \max_j |x_j|.$$

- Hölder's inequality:

  For $p, q \in [1, \infty]$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$,

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

  Moreover, $\|\cdot\|_p$ and $\|\cdot\|_q$ are a pair of dual norms.

- Generalized Cauchy–Schwarz inequality:

  For any pair of dual norms $\|\cdot\|$ and $\|\cdot\|_*$,

$$\langle \mathbf{x}, \mathbf{y} \rangle \leq \|\mathbf{x}\| \|\mathbf{y}\|_*$$

- Fenchel–Young inequality:

  For any pair of dual norms $\|\cdot\|$, $\|\cdot\|_*$ and any $\eta > 0$,

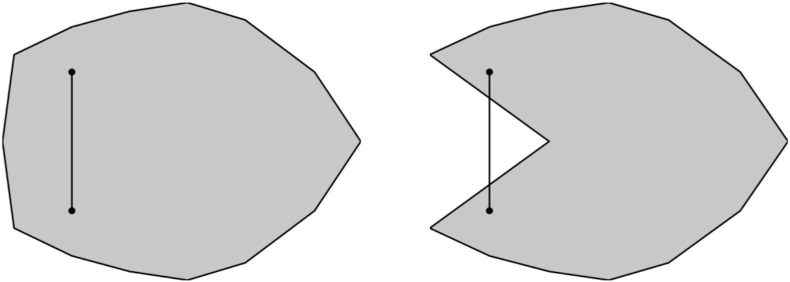
$$\langle \mathbf{x}, \mathbf{y} \rangle \leq \frac{\eta}{2} \|\mathbf{x}\|^2 + \frac{1}{2\eta} \|\mathbf{y}\|_*^2.$$

2. Convex sets

- A set $\mathcal{C} \in \mathbb{R}^n$ is a *convex set* if the straight line segment connecting any two points in $\mathcal{C}$ lies entirely inside $\mathcal{C}$. Formally,

$$\forall\, \mathbf{x}, \mathbf{y} \in \mathcal{C},\ \alpha \in [0,1] : \quad \alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in \mathcal{C}.$$

Example: A convex set (left) and a non-convex set (right).



2.1 Basic properties of convex sets

- If $\alpha \in \mathbb{R}$ and $\mathcal{C}$ is convex, then

$$\alpha\mathcal{C} := \{\alpha\mathbf{x} : \mathbf{x} \in \mathcal{C}\}$$

  is convex.

- If $\alpha_i \in \mathbb{R}$ and all $\mathcal{C}_i$ are convex, then

$$\mathcal{C} = \sum_{i=1}^{m} \alpha_i \mathcal{C}_i := \left\{ \sum_{i=1}^{m} \alpha_i \mathbf{x}_i : \mathbf{x}_i \in \mathcal{C}_i \right\}$$

  is convex.

- If all $\mathcal{C}_i$, $i = 1 : m$, are convex. Then the Cartesian product

$$\mathcal{C}_1 \times \mathcal{C}_2 \times \cdots \times \mathcal{C}_m := \{(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m) : \mathbf{x}_i \in \mathcal{C}_i\}$$

  is convex.

- Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a convex set and let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times m}$. Then the sets

$$\mathbf{A}(\mathcal{C}) := \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathcal{C}\}, \quad \mathbf{B}^{-1}(\mathcal{C}) := \{\mathbf{y} \in \mathbb{R}^m : \mathbf{B}\mathbf{y} \in \mathcal{C}\}$$

  are both convex.
- If $\mathcal{C}_\alpha$ are convex sets for each $\alpha \in \mathcal{A}$, where $\mathcal{A}$ is an arbitrary index set, then the intersection

$$\mathcal{C} = \bigcap_{\alpha\in\mathcal{A}} \mathcal{C}_\alpha$$

  is convex.
- The convex hull of a set of points $\mathbf{x}_1, \cdots, \mathbf{x}_m \in \mathbb{R}^n$, defined by

$$\operatorname{conv}\{\mathbf{x}_1, \cdots, \mathbf{x}_m\} := \left\{ \sum_{i=1}^m \lambda_i \mathbf{x}_i : \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1 \right\},$$

  is convex.

Theorem 1 (Projection onto closed convex sets)

Let $\mathcal{C}$ be a closed convex set and $\mathbf{x} \in \mathbb{R}^n$. Then there is a unique point $\pi_{\mathcal{C}}(\mathbf{x})$, called the projection of $\mathbf{x}$ onto $\mathcal{C}$, such that

$$\|\mathbf{x} - \pi_{\mathcal{C}}(\mathbf{x})\|_2 = \inf_{\mathbf{y} \in \mathcal{C}} \|\mathbf{y} - \mathbf{x}\|_2,$$

that is,

$$\pi_{\mathcal{C}}(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{y} \in \mathcal{C}} \|\mathbf{y} - \mathbf{x}\|_2^2.$$

A point $\mathbf{z}$ is the projection of $\mathbf{x}$ onto $\mathcal{C}$, i.e.,

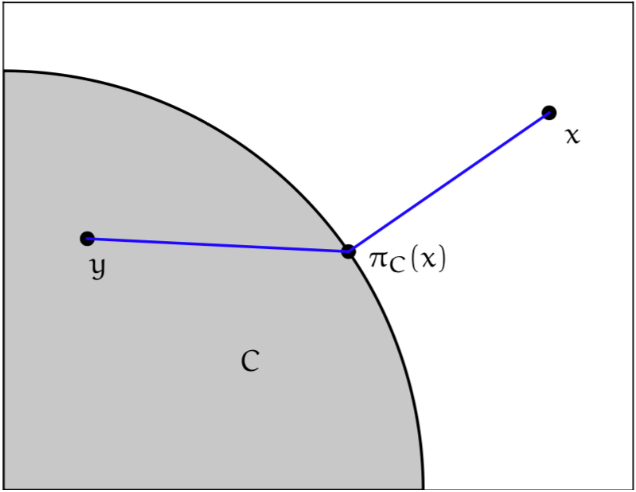
$$\mathbf{z} = \pi_{\mathcal{C}}(\mathbf{x}),$$

if and only if

$$\langle \mathbf{x} - \mathbf{z}, \mathbf{y} - \mathbf{z} \rangle \leq 0,$$

for all $\mathbf{y} \in \mathcal{C}$.

- Projection of the point $\mathbf{x}$ onto the set $\mathcal{C}$ (with projection $\pi_{\mathcal{C}}(\mathbf{x})$), exhibiting $\langle \mathbf{x} - \pi_{\mathcal{C}}(\mathbf{x}), \mathbf{y} - \pi_{\mathcal{C}}(\mathbf{x}) \rangle \leq 0$.



Corollary 2 (Nonexpansiveness)

Projections onto closed convex sets are nonexpansive, in particular,

$$\|\pi_{\mathcal{C}}(\mathbf{x}) - \mathbf{y}\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_2$$

for any $\mathbf{x} \in \mathbb{R}^n$ *and* $\mathbf{y} \in \mathcal{C}$*.*

Theorem 3 (Strict separation of points)

Let $\mathcal{C}$ *be a closed convex set. For any* $\mathbf{x} \notin \mathcal{C}$*, the vector*

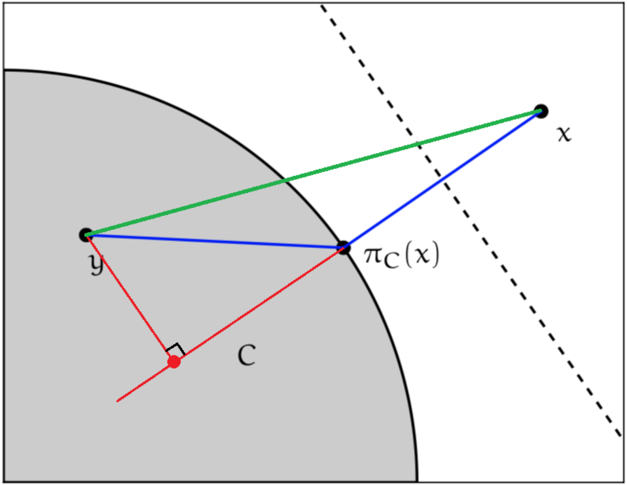
$$\mathbf{v} = \mathbf{x} - \pi_{\mathcal{C}}(\mathbf{x})$$

satisfies

$$\langle \mathbf{v}, \mathbf{x} \rangle \geq \sup_{\mathbf{y} \in \mathcal{C}} \langle \mathbf{v}, \mathbf{y} \rangle + \|\mathbf{v}\|_2^2 > \sup_{\mathbf{y} \in \mathcal{C}} \langle \mathbf{v}, \mathbf{y} \rangle.$$

This means the strict separation of the point $\mathbf{x} \notin \mathcal{C}$ *from the closed convex set* $\mathcal{C}$*.*

- Strict separation of $\mathbf{x}$ from $\mathcal{C}$ by the vector $\mathbf{v} = \mathbf{x} - \pi_{\mathcal{C}}(\mathbf{x})$.



- For nonempty sets $\mathcal{S}_1$ and $\mathcal{S}_2$ satisfying $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$, if there exist vector $\mathbf{v} \neq \mathbf{0}$ and scalar b such that

$$\langle \mathbf{v}, \mathbf{x} \rangle \geq b \quad \text{for all} \quad \mathbf{x} \in \mathcal{S}_1,$$

and

$$\langle \mathbf{v}, \mathbf{x} \rangle \leq b \quad \text{for all} \quad \mathbf{x} \in \mathcal{S}_2,$$

then

$$\{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{v}, \mathbf{x} \rangle = b\}$$

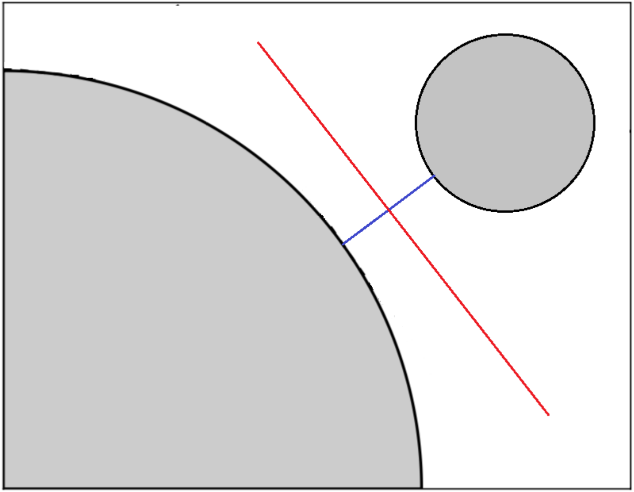
is called a separating hyperplane for nonempty sets $\mathcal{S}_1$ and $\mathcal{S}_2$.

Theorem 4 (Strict separation of closed convex sets)

Let $\mathcal{C}_1, \mathcal{C}_2$ be closed convex sets, with $\mathcal{C}_2$ compact and $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$. Then there is a vector $\mathbf{v}$ such that

$$\inf_{\mathbf{x} \in \mathcal{C}_1} \langle \mathbf{v}, \mathbf{x} \rangle > \sup_{\mathbf{x} \in \mathcal{C}_2} \langle \mathbf{v}, \mathbf{x} \rangle.$$

- Strict separation of closed convex sets.



- For a set $\mathcal{S}$ and a boundary point $\mathbf{x}$, i.e.,

$$\mathbf{x} \in \mathrm{bd}\mathcal{S} := \mathrm{cl}\mathcal{S} \setminus \mathrm{int}\mathcal{S},$$

if vector $\mathbf{v} \neq \mathbf{0}$ satisfies

$$\langle \mathbf{v}, \mathbf{x} \rangle \geq \langle \mathbf{v}, \mathbf{y} \rangle \quad \text{for all} \quad \mathbf{y} \in \mathcal{S},$$

then

$$\{\mathbf{z} \in \mathbb{R}^n \mid \mathbf{v}^\top (\mathbf{z} - \mathbf{x}) = 0\}$$

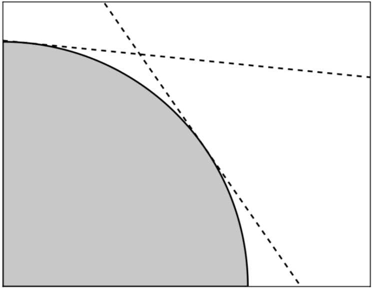
is called a supporting hyperplane supporting $\mathcal{S}$ at $\mathbf{x}$.

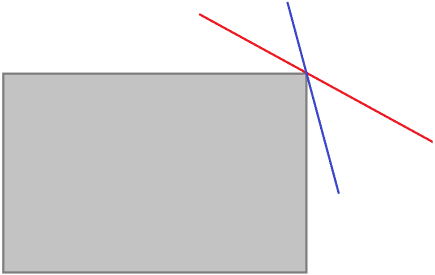
Theorem 5 (Supporting hyperplane theorem)

For convex set $\mathcal{C}$ and any $\mathbf{x} \in \mathrm{bd}\mathcal{C}$, theres exists a supporting hyperplane supporting $\mathcal{C}$ at $\mathbf{x}$, i.e., $\exists\ \mathbf{v} \neq \mathbf{0}$ satisfying

$$\langle \mathbf{v}, \mathbf{x} \rangle \geq \langle \mathbf{v}, \mathbf{y} \rangle \quad \textit{for all} \quad \mathbf{y} \in \mathcal{C}.$$

- Supporting hyperplanes to a convex set. (unique?)





Theorem 6 (Halfspace intersections)

Let $\mathcal{C} \subset \mathbb{R}^n$ be a closed convex set. Then $\mathcal{C}$ is the intersection of all the halfspaces containing it. Moreover, $\mathcal{C} = \bigcap_{\mathbf{x} \in \mathrm{bd}\,\mathcal{C}} \mathcal{H}_{\mathbf{x}}$, where $\mathcal{H}_{\mathbf{x}}$ denotes the intersection of the halfspaces contained in the hyperplanes supporting $\mathcal{C}$ at $\mathbf{x}$.

3. Convex functions

- The *epigraph* of a function f is defined as

$$\operatorname{epi} f := \{(\mathbf{x}, t) : f(\mathbf{x}) \leq t\}.$$

- A function f is called *closed* if its epigraph is a closed set.

  If $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is continuous over its domain and $\operatorname{dom}(f)$ is closed. Then f is closed.

- A function $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$ is a *convex* function if its domain $\operatorname{dom}(f)$ is convex and for all $\mathbf{x}, \mathbf{y} \in \operatorname{dom}(f)$, $\alpha \in [0, 1]$,

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}).$$

  (*strictly convex* means $<$)

  A function is convex if and only if its epigraph is a convex set.

Lemma 7 (Convexity + compactness $\Rightarrow$ boundedness)

Let f be convex and defined on the ℓ_1 ball in n dimensions:

$$\mathcal{B}_1 = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 \leq 1\}.$$

Then there exist $-\infty < m \leq M < \infty$ such that

$$m \leq f(\mathbf{x}) \leq M, \quad \forall\, \mathbf{x} \in \mathcal{B}_1.$$

More general, convex f on a compact domain ($\subseteq \mathrm{dom}(f)$) is bounded.

Theorem 8 (Convexity + compactness $\Rightarrow$ L-continuity)

Let f be convex and defined on a convex set $\mathcal{C}$ with non-empty interior. Let $\mathcal{B} \subseteq \mathrm{int}\mathcal{C}$ be compact. Then there is a constant L such that

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\|$$

on $\mathcal{B}$, that is, f is L-Lipschitz continuous on $\mathcal{B}$.

- Definition: The *directional derivative* of a function f at a point $\mathbf{x}$ in the direction $\mathbf{d}$ is

$$f'(\mathbf{x};\mathbf{d}) := \lim_{\alpha \to 0^+} \frac{f(\mathbf{x}+\alpha\mathbf{d}) - f(\mathbf{x})}{\alpha}.$$

Theorem 9 (Convexity ⇒ existence of directional derivative)

For convex f, at any point $\mathbf{x} \in \operatorname{int}\operatorname{dom}(f)$ and for any $\mathbf{d} \in \mathbb{R}^n$, the directional derivative $f'(\mathbf{x};\mathbf{d})$ exists. The function $g_{\mathbf{x}}(\mathbf{d}) = f'(\mathbf{x};\mathbf{d})$ is convex and satisfies for any $\lambda \geq 0$,

$$g_{\mathbf{x}}(\lambda\mathbf{d}) = f'(\mathbf{x},\lambda\mathbf{d}) = \lambda f'(\mathbf{x},\mathbf{d}) = \lambda g_{\mathbf{x}}(\mathbf{d}).$$

Moreover, there exists a constant $L < \infty$ such that

$$|g_{\mathbf{x}}(\mathbf{d})| = |f'(\mathbf{x};\mathbf{d})| \leq L\|\mathbf{d}\|$$

for any $\mathbf{d} \in \mathbb{R}^n$.

3.1 Operations preserving convexity

- Summation and multiplication by nonnegative scalars.

  Let $\{f_i\}_{i=1}^m$ be convex functions defined over a convex set $\mathcal{C}$, and let $\{\alpha_i \geq 0\}_{i=1}^m$. Then $\displaystyle\sum_{i=1}^m \alpha_i f_i$ is convex over $\mathcal{C}$.

- Composition of a convex function with an affine transformation.

  Let f be a convex function defined on a convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Let $\mathbf{A} \in \mathbb{R}^{n\times m}$ and $\mathbf{b} \in \mathbb{R}^n$. The $g(\mathbf{y}) = f(\mathbf{A}\mathbf{y} + \mathbf{b})$ is convex over the convex set $\mathcal{D} = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{A}\mathbf{y} + \mathbf{b} \in \mathcal{C}\}$.

- Composition of a nondecreasing convex function with a convex function. Example: $h(\mathbf{x}) = (\|\mathbf{x}\|_2^2 + 1)^2$.

  Let $f : \mathcal{C} \to \mathbb{R}$ be a convex function over the convex set $\mathcal{C}$. Let $g : \mathcal{I} \to \mathbb{R}$ be a one-dimensional nondecreasing convex function over the interval $\mathcal{I} \subseteq \mathbb{R}$. Assume that the image of $\mathcal{C}$ under f is contained in $\mathcal{I}$: $f(\mathcal{C}) \subseteq \mathcal{I}$. Then the composition of g with f defined by $h(\mathbf{x}) = g(f(\mathbf{x}))$ is a convex function over $\mathcal{C}$.

- Pointwise maximum of convex functions.

  Let $f_1, \cdots, f_m : \mathcal{C} \to \mathbb{R}$ be m convex functions over the convex set $\mathcal{C}$. Then the maximum function

$$f(\mathbf{x}) = \max_i f_i(\mathbf{x})$$

  is a convex function over $\mathcal{C}$.

  Examples: (1) $f(\mathbf{x}) = \max\{x_1, x_2, \cdots, x_n\}$, (2) the sum of the k largest values:
  $h_k(\mathbf{x}) = \max\{x_{i_1} + \cdots + x_{i_k} : i_1, \cdots, i_k \in [n] \text{ are different}\}$.

- Partial minimization of a convex function.

  Let $f : \mathcal{C} \times \mathcal{D} \to \mathbb{R}$ be a convex function defined over the set $\mathcal{C} \times \mathcal{D}$ where $\mathcal{C}$ and $\mathcal{D}$ are convex sets. Let

$$g(\mathbf{x}) = \min_{\mathbf{y} \in \mathcal{D}} f(\mathbf{x}, \mathbf{y}), \quad \mathbf{x} \in \mathcal{C},$$

  where we assume that the minimum in the above definition is finite. Then g is convex over $\mathcal{C}$.

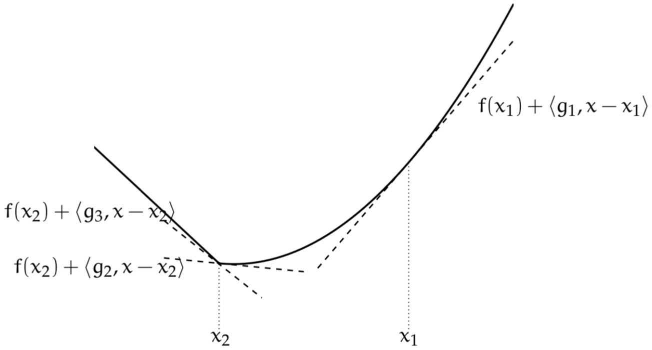
4. Subgradient and subdifferential

- Definition: A vector $\mathbf{g} \in \mathbb{R}^n$ is a *subgradient* of f at a point $\mathbf{x}$ if

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

The *subdifferential*, denoted $\partial f(\mathbf{x})$, is the set of all subgradients of f at $\mathbf{x}$.

Example: $\mathbf{g}_1 = \nabla f(\mathbf{x}_1)$, $\mathbf{g}_2, \mathbf{g}_3 \in \partial f(\mathbf{x}_2)$



- Examples: Let $\|\cdot\|$ be a norm. Then

$$\partial\|\mathbf{x}\| = \begin{cases} \{\mathbf{g} \in \mathbb{R}^n : \|\mathbf{g}\|_* = 1, \langle \mathbf{g}, \mathbf{x} \rangle = \|\mathbf{x}\|\} & \text{if } \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{g} \in \mathbb{R}^n : \|\mathbf{g}\|_* \leq 1\} & \text{if } \mathbf{x} = \mathbf{0}. \end{cases}$$

For $\|\mathbf{x}\|_2$, we have $\partial\|\mathbf{x}\|_2 = \begin{cases} \{\mathbf{x}/\|\mathbf{x}\|_2\} & \text{if } \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{g} \in \mathbb{R}^n : \|\mathbf{g}\|_2 \leq 1\} & \text{if } \mathbf{x} = \mathbf{0}. \end{cases}$

For case $n = 1$, we have $\partial|x| = \begin{cases} \{-1\} & \text{if } x < 0, \\ [-1, 1] & \text{if } x = 0, \\ \{1\} & \text{if } x > 0. \end{cases}$

Theorem 10 (Nonemptiness, closedness, convexity, boundedness of subdifferential at interior points of $\mathrm{dom}(f)$ of convex f)

Suppose f is convex. Let $\mathbf{x} \in \mathrm{int}\,\mathrm{dom}(f)$. Then $\partial f(\mathbf{x})$ is nonempty, closed, convex, and bounded.

Theorem 11 (Nonemptiness of subdifferential $\Rightarrow$ convexity)

Let $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$ be proper and assume that $\mathrm{dom}(f)$ is convex. Suppose that for any $\mathbf{x} \in \mathrm{dom}(f)$, the set $\partial f(\mathbf{x})$ is nonempty. Then f is convex.

Theorem 12 (First-order characterizations of strong convexity)

Let $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$ be proper closed and convex. Then for a given $\gamma > 0$, the following three claims are equivalent:

(i) *f is γ-strongly convex.*

(ii) *For any $\mathbf{x}$ satisfying $\partial f(\mathbf{x}) \neq \emptyset$, $\mathbf{y} \in \mathrm{dom}(f)$ and $\mathbf{g} \in \partial f(\mathbf{x})$,*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{y} - \mathbf{x} \rangle + \frac{\gamma}{2}\|\mathbf{y} - \mathbf{x}\|^2.$$

(iii) *For any $\mathbf{x}$ and $\mathbf{y}$ satisfying $\partial f(\mathbf{x}) \neq \emptyset$, $\partial f(\mathbf{y}) \neq \emptyset$, and $\mathbf{g}_{\mathbf{x}} \in \partial f(\mathbf{x})$, $\mathbf{g}_{\mathbf{y}} \in \partial f(\mathbf{y})$,*

$$\langle \mathbf{g}_{\mathbf{x}} - \mathbf{g}_{\mathbf{y}}, \mathbf{x} - \mathbf{y} \rangle \geq \gamma \|\mathbf{x} - \mathbf{y}\|^2.$$

Theorem 13 (Equivalent characterization of subdifferential)

An equivalent characterization of the subdifferential $\partial f(\mathbf{x})$ of convex f at $\mathbf{x}$ is

$$\partial f(\mathbf{x}) = \{\mathbf{g} : \langle \mathbf{g}, \mathbf{d} \rangle \le f'(\mathbf{x}; \mathbf{d}) \ \forall \ \mathbf{d} \in \mathbb{R}^n\}.$$

Theorem 14 (Max formula of directional derivative)

Suppose f is closed convex and $\partial f(\mathbf{x}) \neq \emptyset$. Then, for all $\mathbf{d} \in \mathbb{R}^n$,

$$f'(\mathbf{x}; \mathbf{d}) = \sup_{\mathbf{g} \in \partial f(\mathbf{x})} \langle \mathbf{g}, \mathbf{d} \rangle.$$

Theorem 15 (Subgradient bounded by Lipschitz constant)

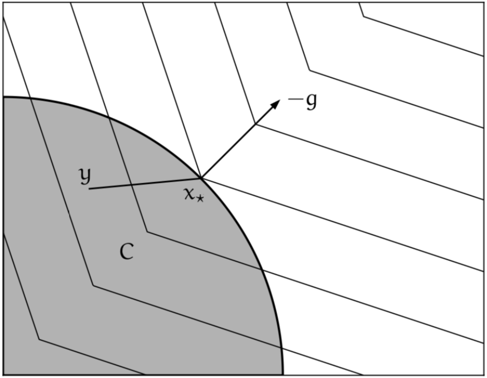
Suppose that convex function f is L-Lipschitz continuous with respect to the norm $\| \cdot \|$ over a set $\mathcal{C}$, where $\mathcal{C} \subset \operatorname{int} \operatorname{dom}(f)$. Then

$$\sup\{\|\mathbf{g}\|_* : \mathbf{g} \in \partial f(\mathbf{x}), \mathbf{x} \in \mathcal{C}\} \le L,$$

Theorem 16 (Minimizer of convex function over convex set)

Let f be convex. The point $\mathbf{x}_\star \in \operatorname{int}\operatorname{dom}(f)$ minimizes f over a closed convex set $\mathcal{C}$ if and only if there exists a subgradient $\mathbf{g} \in \partial f(\mathbf{x}_\star)$ such that

$$\langle \mathbf{g}, \mathbf{y} - \mathbf{x}_\star \rangle \geq 0 \quad \textit{for all} \quad \mathbf{y} \in \mathcal{C}.$$



The point $\mathbf{x}_\star$ minimizes f over $\mathcal{C}$

(the shown level curves)

Active case: $\mathbf{x}_\star \in \mathrm{bd}\mathcal{C}$

$-\mathbf{g}$: supporting hyperplane

Inactive case: $\mathbf{x}_\star \in \mathrm{int}\mathcal{C}$

$\mathbf{g} = \mathbf{0} \Rightarrow \mathbf{0} \in \partial f(\mathbf{x}_\star)$

5. Calculus rules with subgradients

- Scaling.

  If $h(\mathbf{x}) = \alpha f(\mathbf{x})$ for some $\alpha \geq 0$, then $\partial h(\mathbf{x}) = \alpha \partial f(\mathbf{x})$.

- Finite sums.

  Suppose that f_i, $i = 1 : m$ are convex functions and let $f = \sum_{i=1}^{m} f_i$.

  If $\mathbf{x} \in \operatorname{int}\operatorname{dom}(f_i)$, $i = 1 : m$, then $\partial f(\mathbf{x}) = \sum_{i=1}^{m} \partial f_i(\mathbf{x})$.

  Exercise: $\mathbf{x} \in \mathbb{R}^m$, $\|\mathbf{x}\|_1 = \sum_{i=1}^{m} f_i(\mathbf{x})$, $f_i(\mathbf{x}) = |x_i|$. $\partial \|\mathbf{x}\|_1 =?$

- Affine transformations.

  Let $f : \mathbb{R}^m \mapsto \mathbb{R}$ be convex and $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Then $h : \mathbb{R}^n \mapsto \mathbb{R}$ defined by $h(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$ is convex and has subdifferential

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

  Exercises: (1) *proof*? (2) $\partial \|\mathbf{A}\mathbf{x} + \mathbf{b}\|_1 =?$ (3) $\partial \|\mathbf{A}\mathbf{x} + \mathbf{b}\|_2 =?$

- Maximum of a finite collection of convex functions.

  Let f_i, $i = 1 : m$, be convex functions, and $f(\mathbf{x}) = \max_{1 \le i \le m} f_i(\mathbf{x})$.

  Then we have

$$\operatorname{epi} f = \bigcap_{1 \le i \le m} \operatorname{epi} f_i,$$

  which is convex, and therefore f is convex.

  If $\mathbf{x} \in \operatorname{int} \operatorname{dom}(f_i)$, $i = 1 : m$, then the subdifferential $\partial f(\mathbf{x})$ is the convex hull of the subgradients of active functions (those attaining the maximum) at $\mathbf{x}$, that is,

$$\partial f(\mathbf{x}) = \operatorname{conv}\{\partial f_i(\mathbf{x}) : f_i(\mathbf{x}) = f(\mathbf{x})\}.$$

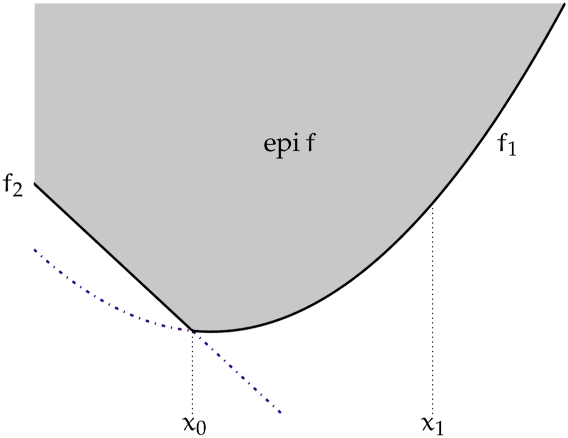
  If there is only a single unique active function f_i, then

$$\partial f(\mathbf{x}) = \partial f_i(\mathbf{x}).$$

Exercise: Let $f(x) = \max\{f_1(x), f_2(x)\}$, where

$$f_1(x) = x^2, \qquad f_2(x) = -2x - 1/5.$$

For $x_0 = -1 + \sqrt{4/5}$, $\quad \partial f(x_0) = ?$



Exercise: $\mathbf{x} \in \mathbb{R}^m$, $\|\mathbf{x}\|_\infty = \max_{1 \le i \le m} f_i(\mathbf{x})$, $f_i(\mathbf{x}) = |x_i|$. $\partial \|\mathbf{x}\|_\infty =?$

- Supremum of an infinite collection of convex functions.

  Consider

$$f(\mathbf{x}) = \sup_{\alpha \in \mathcal{A}} f_\alpha(\mathbf{x}),$$

  where $\mathcal{A}$ is an arbitrary index set and f_α is convex for each α.

  If the supremum is attained, then

$$\partial f(\mathbf{x}) \supseteq \operatorname{conv}\{\partial f_\alpha(\mathbf{x}) : f_\alpha(\mathbf{x}) = f(\mathbf{x})\}.$$

  If the supremum is not attained, the function f may not be subdifferentiable at $\mathbf{x}$.