Lecture 6: Convex sets and convex functions

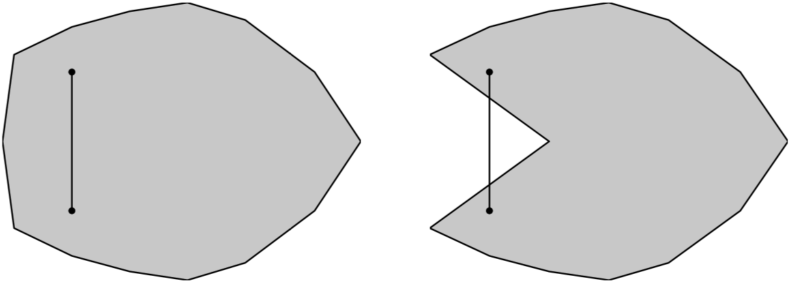
School of Mathematical Sciences, Xiamen University

1. 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).



1.1 Basic properties of convex sets

- If $\alpha_i \in \mathbb{R}$ and all $\mathcal{C}_i$, $i = 1 : m$, 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 (possibly infinite), 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. Let $\mathcal{S} \subseteq \mathbb{R}^n$. Then

$$\operatorname{conv}(\mathcal{S}) = \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{x}_i : \mathbf{x}_i \in \mathcal{S}, \lambda_i \geq 0, \sum_{i=1}^{k} \lambda_i = 1, k \in \mathbb{N} \right\}$$

is the "smallest" convex set containing $\mathcal{S}$.

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,

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

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

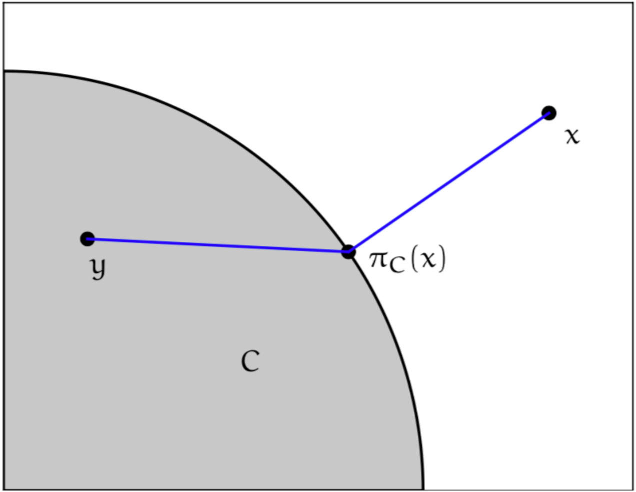
$$\mathbf{z} = \pi_{\mathcal{C}}(\mathbf{x}),$$

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*

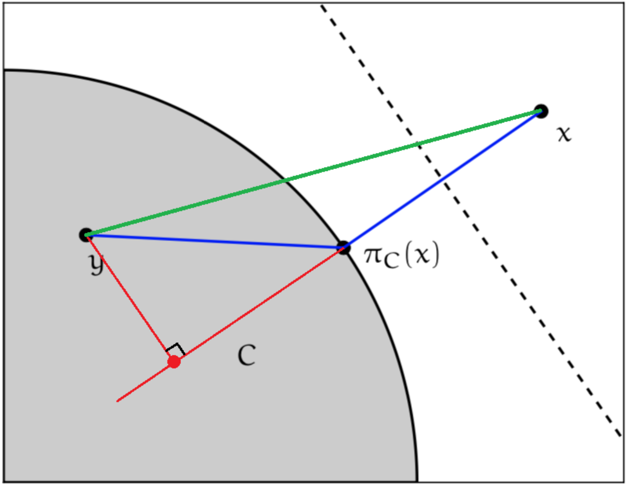
$$\mathbf{v} = \mathbf{x} - \pi_{\mathcal{C}}(\mathbf{x})$$

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

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

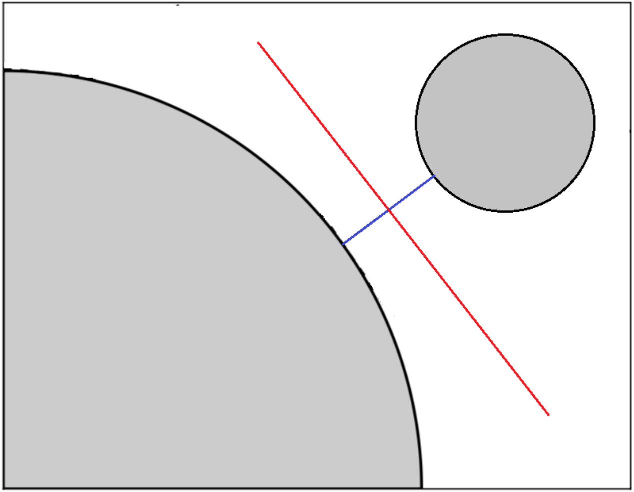
is called a separating hyperplane for nonempty sets $\mathcal{S}_1$ and $\mathcal{S}_2$.

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

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

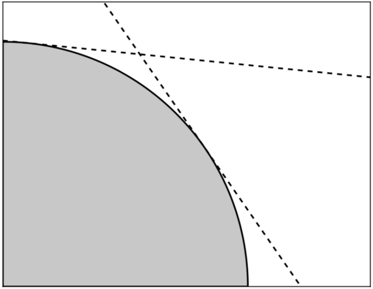
  is called a supporting hyperplane supporting $\mathcal{S}$ at $\mathbf{x}$.

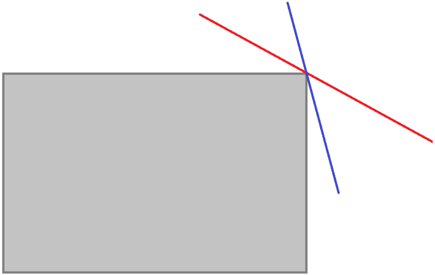
Theorem 5 (Supporting hyperplane theorem)

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 $\mathbf{x}$.

2. Convex functions

- A function $f : \mathcal{C} \to \mathbb{R}$ defined on a convex set $\mathcal{C} \subseteq \mathbb{R}^n$ is called *convex* (or *convex over* $\mathcal{C}$) if for any $\mathbf{x}, \mathbf{y} \in \mathcal{C}$, $\lambda \in [0, 1]$,

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

  It is called *strictly convex* if for any $\mathbf{x} \neq \mathbf{y}$, $\lambda \in (0, 1)$,

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

  Examples of convex functions: afines functions, norms.

- Jensen's inequality.

  Let $f : \mathcal{C} \to \mathbb{R}$ be a convex function defined on the convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Then for any $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathcal{C}$ and $\lambda_i \geq 0$, $\sum_{i=1}^k \lambda_i = 1$, the following inequality holds:

  $$f\left(\sum_{i=1}^k \lambda_i \mathbf{x}_i\right) \leq \sum_{i=1}^k \lambda_i f(\mathbf{x}_i).$$

2.1 Characterizations of convex functions

Theorem 7 (the gradient inequality)

Let $f : \mathcal{C} \to \mathbb{R}$ be a continuously differentiable function defined on a nonempty convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Then f is convex over $\mathcal{C}$ if and only if

$$f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \le f(\mathbf{y}) \quad \textit{for any} \quad \mathbf{x}, \mathbf{y} \in \mathcal{C},$$

and f is strictly convex over $\mathcal{C}$ if and only if

$$f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) < f(\mathbf{y}) \ \textit{for any} \ \mathbf{x}, \mathbf{y} \in \mathcal{C} \ \textit{satisfying} \ \mathbf{x} \neq \mathbf{y}.$$

Theorem 8 (monotonicity of the gradient)

Suppose that f is a continuously differentiable function over a nonempty convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Then f is convex over $\mathcal{C}$ if and only if

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^\top (\mathbf{x} - \mathbf{y}) \ge 0 \quad \textit{for any} \quad \mathbf{x}, \mathbf{y} \in \mathcal{C}.$$

Proposition 9 (optimality conditions)

Let f be a continuously differentiable function which is convex over a convex set $\mathcal{C} \subseteq \mathbb{R}^n$.

(1) *Suppose that $\nabla f(\mathbf{x}_\star) = \mathbf{0}$ for some $\mathbf{x}_\star \in \mathcal{C}$. Then $\mathbf{x}_\star$ is a global minimizer of f over $\mathcal{C}$.*

(2) *If $\mathcal{C} = \mathbb{R}^n$, then $\nabla f(\mathbf{x}_\star) = \mathbf{0}$ if and only if $\mathbf{x}_\star$ is a global minimizer of f over $\mathbb{R}^n$.*

Theorem 10 (second order characterization of convex functions)

Let f be a twice continuously differentiable function over a nonempty convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Then

(1) *If $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for any $\mathbf{x} \in \mathcal{C}$, then f is convex over $\mathcal{C}$.*

(2) *If $\nabla^2 f(\mathbf{x}) \succ \mathbf{0}$ for any $\mathbf{x} \in \mathcal{C}$, then f is strictly convex over $\mathcal{C}$.*

(3) *If $\mathcal{C}$ is open, then f is convex over $\mathcal{C}$ if and only if $\nabla^2 f(\mathbf{x}) \succeq \mathbf{0}$ for any $\mathbf{x} \in \mathcal{C}$.*

2.2 Operations preserving convexity

Theorem 11 (nonnegative scalar multiplication and summation)

(1) *Let f be a convex function defined over a convex set $\mathcal{C} \subseteq \mathbb{R}^n$ and let $\alpha \geq 0$. Then αf is a convex function over $\mathcal{C}$.*

(2) *Let $f_1, f_2, \ldots, f_p$ be convex functions over a convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Then the sum function $f_1 + f_2 + \cdots + f_p$ is convex over $\mathcal{C}$.*

Theorem 12 (affine change of variables)

Let $f : \mathcal{C} \to \mathbb{R}$ 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 $\mathbf{b} \in \mathbb{R}^n$. Then the function g defined by

$$g(\mathbf{y}) := f(\mathbf{A}\mathbf{y} + \mathbf{b})$$

is convex over the convex set

$$\mathcal{D} = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{A}\mathbf{y} + \mathbf{b} \in \mathcal{C}\}.$$

Theorem 13 (composition with a nondecreasing convex function)

Let $f : \mathcal{C} \to \mathbb{R}$ *be a convex function over the convex set* $\mathcal{C} \subseteq \mathbb{R}^n$*. 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*

$$h(\mathbf{x}) := g(f(\mathbf{x})), \quad \mathbf{x} \in \mathcal{C},$$

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

Theorem 14 (pointwise maximum of convex functions)

Let $f_1, \ldots, f_p : \mathcal{C} \to \mathbb{R}$ *be* p *convex functions over the convex set* $\mathcal{C} \subseteq \mathbb{R}^n$*. Then the maximum function*

$$f(\mathbf{x}) := \max_{i=1,\ldots,p} f_i(\mathbf{x})$$

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

Theorem 15 (partial minimization)

Let $f : \mathcal{C} \times \mathcal{D} \to \mathbb{R}$ be a convex function defined over the set $\mathcal{C} \times \mathcal{D}$, where $\mathcal{C} \subseteq \mathbb{R}^m$ and $\mathcal{D} \subseteq \mathbb{R}^n$ 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 minimal value (maybe not attained) in the above definition is finite. Then g is convex over $\mathcal{C}$.

- Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a nonempty convex set and $\|\cdot\|$ an arbitrary norm. The distance function defined by

$$d(\mathbf{x}, \mathcal{C}) := \min_{\mathbf{y} \in \mathcal{C}} \|\mathbf{x} - \mathbf{y}\|$$

  is convex since the function $f(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ is convex over $\mathbb{R}^n \times \mathcal{C}$.

2.3 Level sets of convex functions

- Let $f : \mathcal{S} \to \mathbb{R}$ be a function defined over a set $\mathcal{S} \subseteq \mathbb{R}^n$. Then the *level set* of f with level $\alpha \in \mathbb{R}$ is given by

$$\mathrm{Lev}(f, \alpha) = \{\mathbf{x} \in \mathcal{S} : f(\mathbf{x}) \leq \alpha\}.$$

Theorem 16 (level sets of convex functions are convex)

Let $f : \mathcal{C} \to \mathbb{R}$ be a convex function defined over a convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Then for any $\alpha \in \mathbb{R}$ the level set $\mathrm{Lev}(f, \alpha)$ is convex.

- A function $f : \mathcal{C} \to \mathbb{R}$ defined over the convex set $\mathcal{C} \subseteq \mathbb{R}^n$ is called *quasi-convex* if for any $\alpha \in \mathbb{R}$ the set $\mathrm{Lev}(f, \alpha)$ is convex.
- Quasi-convex functions may be nonconvex.

  For example, $f(x) = \sqrt{|x|}$ with level sets

$$\mathrm{Lev}(f, \alpha) = \begin{cases} [-\alpha^2, \alpha^2], & \alpha \geq 0, \\ \emptyset, & \alpha < 0. \end{cases}$$

2.4 Continuity and differentiability of convex functions

- Convex functions are always continuous at interior points of their domain. Thus, for example, functions which are convex over $\mathbb{R}^n$ are always continuous. A stronger result is given below.

Theorem 17 (local Lipschitz continuity at interior points)

Let $f : \mathcal{C} \to \mathbb{R}$ be a convex function defined over a convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Let $\mathbf{x}_0 \in \operatorname{int}(\mathcal{C})$. Then there exist $\varepsilon > 0$ and $L > 0$ such that $\mathcal{B}[\mathbf{x}_0, \varepsilon] \subseteq \mathcal{C}$ and

$$|f(\mathbf{x}) - f(\mathbf{x}_0)| \leq L\|\mathbf{x} - \mathbf{x}_0\|$$

for all $\mathbf{x} \in \mathcal{B}[\mathbf{x}_0, \varepsilon]$.

Theorem 18 (existence of directional derivatives at interior points)

Let $f : \mathcal{C} \to \mathbb{R}$ be a convex function defined over a convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Let $\mathbf{x} \in \operatorname{int}(\mathcal{C})$. Then for any $\mathbf{d} \neq \mathbf{0}$, the directional derivative $f'(\mathbf{x}; \mathbf{d})$ exists.

2.5 Maxima of convex functions

Theorem 19

Let $f : \mathcal{C} \to \mathbb{R}$ be a convex function which is not constant over the convex set $\mathcal{C}$. Then f does not attain a maximum at a point in $\operatorname{int}(\mathcal{C})$.

- Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a convex set. A point $\mathbf{x} \in \mathcal{C}$ is called an *extreme point* of $\mathcal{C}$ if there do not exist $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}, \mathbf{x}_1 \neq \mathbf{x}_2$, and $\lambda \in (0, 1)$ such that $\mathbf{x} = \lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$. The set of extreme points is denoted by $\operatorname{ext}(\mathcal{C})$.

Theorem 20 (Krein–Milman)

Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a compact convex set. Then $\mathcal{C} = \operatorname{conv}(\operatorname{ext}(\mathcal{C}))$.

Theorem 21

Let $f : \mathcal{C} \to \mathbb{R}$ be a convex and continuous function over the nonempty convex and compact set $\mathcal{C} \subseteq \mathbb{R}^n$. Then there exists at least one maximizer of f over $\mathcal{C}$ that is an extreme point of $\mathcal{C}$.

2.6 Convexity and inequalities

- The arithmetic geometric mean inequality

  For any $x_1, \ldots, x_n \geq 0$ and $\boldsymbol{\lambda} \in \Delta_n$ the following inequality holds:

$$\sum_{i=1}^{n} \lambda_i x_i \geq \prod_{i=1}^{n} x_i^{\lambda_i}.$$

- Young's inequality

  For any $s, t \geq 0$ and $p, q > 1$ satisfying $1/p + 1/q = 1$ it holds that

$$st \leq s^p/p + t^q/q.$$

- Hölder's inequality

  For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $p, q \in [1, \infty]$ satisfying $1/p + 1/q = 1$, it holds that

$$|\mathbf{x}^\top \mathbf{y}| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q.$$

- Minkowski's inequality

  Let $p \geq 1$. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$.

2.7 Extended real-valued function

- The *effective domain* of an *extended real-valued function* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$\mathrm{dom}(f) := \{\mathbf{x} \mid f(\mathbf{x}) < +\infty\}.$$

- An extended real-valued function is called *proper* if there exists at least one $\mathbf{x} \in \mathbb{R}^n$ such that $f(\mathbf{x}) < +\infty$, meaning that $\mathrm{dom}(f) \neq \emptyset$.
- An extended real-valued function f is convex if $\forall\ \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha \in [0, 1]$ the following inequality holds:

$$f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) \leqslant (1-\alpha) f(\mathbf{x}) + \alpha f(\mathbf{y}),$$

  where we use the arithmetic with $+\infty$:

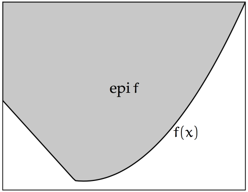
$$a + (+\infty) = +\infty\ (a \in \mathbb{R}), \quad b \cdot (+\infty) = +\infty\ (b > 0),$$

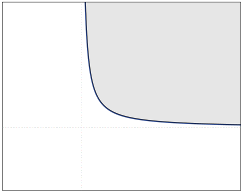
  and

$$0 \cdot (+\infty) = 0.$$

- The definition of convexity of extended real-valued functions is equivalent to saying that $\mathrm{dom}(f)$ is a convex set and that the restriction of f to its effective domain $\mathrm{dom}(f)$ is a convex function.
- The *epigraph* of $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\mathrm{epi}(f) = \{(\mathbf{x}, y) : f(\mathbf{x}) \le y, \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R}\}.$$





An extended real-valued function f convex "$\Leftrightarrow$" $\mathrm{epi}(f)$ convex.

Theorem 22

Let $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be an extended real-valued convex function for any $i \in \mathcal{I}$ ($\mathcal{I}$ being an arbitrary index set). Then $f(\mathbf{x}) = \max_{i \in \mathcal{I}} f_i(\mathbf{x})$ is an extended real-valued convex function.

- If there exists a value $\gamma > 0$ such that

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leqslant \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{y}) - \frac{\gamma}{2}\lambda(1-\lambda)\|\mathbf{x}-\mathbf{y}\|_2^2$$

  for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in [0,1]$, we say that f is *strongly convex with modulus of convexity* γ. Exercise: If f is strongly convex with modulus of convexity γ, then $f(\mathbf{x}) - \frac{\gamma}{2}\|\mathbf{x}\|_2^2$ is convex.

- For differentiable f, *equivalent* definition of *strongly convex with modulus of convexity* γ: $\forall$ $\mathbf{x}, \mathbf{y}$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x}) + \frac{\gamma}{2}\|\mathbf{y}-\mathbf{x}\|_2^2.$$

  Lipschitz continuously differentiable with constant L: $\forall$ $\mathbf{x}, \mathbf{y}$,

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top(\mathbf{y}-\mathbf{x}) + \frac{L}{2}\|\mathbf{y}-\mathbf{x}\|_2^2.$$

  If further f is twice continuously differentiable, then for all $\mathbf{x}$, the above inequalities is equivalent to $\gamma\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq L\mathbf{I}$.

Theorem 23

Let f be differentiable and strongly convex with modulus of convexity $\gamma > 0$. Then the minimizer $\mathbf{x}_\star$ of f exists and is unique.

Proof. (i) Compactness of level set: Show that for any point $\mathbf{x}^0$, the level set

$$\{\mathbf{x} \mid f(\mathbf{x}) \leq f(\mathbf{x}^0)\}$$

is closed and bounded, and hence compact.

(ii) Existence: Since f is continuous, it attains its minimum on the compact level set, which is also the solution of $\min_{\mathbf{x}} f(\mathbf{x})$.

(iii) Uniqueness: Suppose for contradiction that the minimizer is not unique, so that we have two points $\mathbf{x}_\star^1$ and $\mathbf{x}_\star^2$ that minimize f. By using the strongly convex property, we can prove

$$f\left(\frac{\mathbf{x}_\star^1 + \mathbf{x}_\star^2}{2}\right) < f(\mathbf{x}_\star^1) = f(\mathbf{x}_\star^2).$$

This is a contradiction. □

3. Subgradient and subdifferential

- Definition: A vector $\mathbf{v} \in \mathbb{R}^n$ is a *subgradient* of $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ at a point $\mathbf{x}$ if for all $\mathbf{y} \in \mathbb{R}^n$, it holds

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{v}^\top(\mathbf{y} - \mathbf{x}).$$

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

Lemma 24 (Monotonicity of subdifferentials)

For all $\mathbf{a} \in \partial f(\mathbf{x})$ *and* $\mathbf{b} \in \partial f(\mathbf{y})$, *we have* $(\mathbf{a} - \mathbf{b})^\top(\mathbf{x} - \mathbf{y}) \geq 0$.

Proof. By the definition of subgradient, we have

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{a}^\top(\mathbf{y} - \mathbf{x}) \quad \text{and} \quad f(\mathbf{x}) \geq f(\mathbf{y}) + \mathbf{b}^\top(\mathbf{x} - \mathbf{y}).$$

Adding these two inequalities yields the statement. □

Theorem 25 (Fermat's lemma: generalization in convex functions)

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. Then the point $\mathbf{x}_\star$ is a minimizer of $f(\mathbf{x})$ if and only if

$$\mathbf{0} \in \partial f(\mathbf{x}_\star).$$

Theorem 26

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper convex, and let $\mathbf{x} \in \mathrm{int}(\mathrm{dom}(f))$.

(i) *If f is differentiable at $\mathbf{x}$, then $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$.*

(ii) *If $\partial f(\mathbf{x})$ is a singleton (a set containing a single vector), then f is differentiable at $\mathbf{x}$ with gradient equal to the unique subgradient.*

- Example: If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and convex, then

$$\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\} \quad \text{and} \quad f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x})$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

- 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 27 (Nonemptiness, closedness, convexity, boundedness of subdifferential at interior points of dom(f) of convex f)

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

Theorem 28 (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 29 (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 30 (Equivalent characterization of subdifferential)

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

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

Theorem 31 (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 32 (Subgradient bounded by Lipschitz constant)

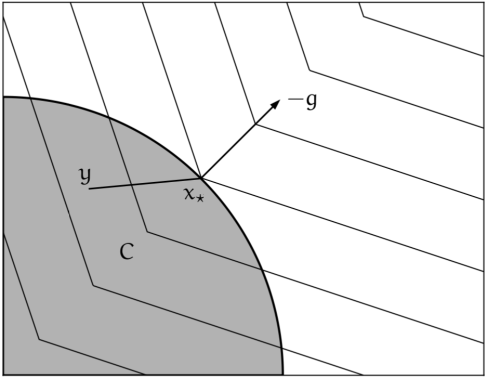
Suppose that convex function f *is* L*-Lipschitz continuous with respect to the norm* $\|\cdot\|$ *over a set* $\mathcal{C}$*, where* $\mathcal{C} \subset \operatorname{int}\operatorname{dom}(f)$*. Then*

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

Theorem 33 (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: $\mathbf{x}_\star \in \operatorname{bd}\mathcal{C}$

$-\mathbf{g}$: supporting hyperplane

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

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

3.1 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 $f(\mathbf{x}) = \max\limits_{1 \le i \le m} f_i(\mathbf{x})$.

  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,

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

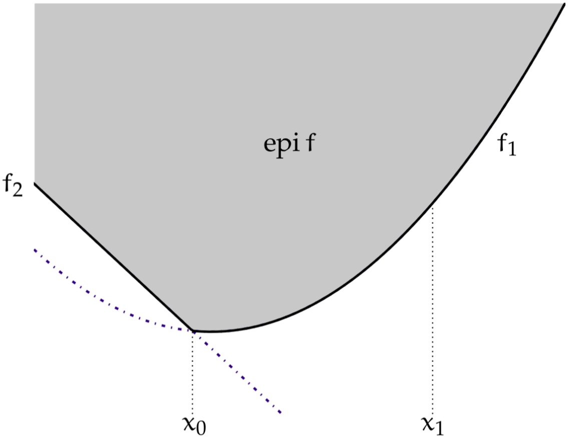
  If there is only a single unique active function f_i, then

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

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 $\mathbf{x}$.

4. Proximal operator

- For a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the *proximal operator* of f is

$$\operatorname{prox}_f(\mathbf{x}) := \operatorname*{argmin}_{\mathbf{u}} \left\{ f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 \right\}.$$

- A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called *closed* if $\operatorname{epi}(f)$ is closed.

- For a closed proper convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, from the optimality condition (see Theorem 25), we have

$$\mathbf{0} \in \partial f(\operatorname{prox}_f(\mathbf{x})) + (\operatorname{prox}_f(\mathbf{x}) - \mathbf{x}).$$

- For a closed proper convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the point $\mathbf{x}_\star$ is a minimizer of f if and only if

$$\mathbf{x}_\star = \operatorname{prox}_f(\mathbf{x}_\star).$$

Lemma 34 (Nonexpansivity of proximal operator)

For a closed proper convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, *we have for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\|\mathrm{prox}_f(\mathbf{x}) - \mathrm{prox}_f(\mathbf{y})\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_2.$$

Proof. From the optimality conditions at two points $\mathbf{x}$ and $\mathbf{y}$, we have

$$\mathbf{x} - \mathrm{prox}_f(\mathbf{x}) \in \partial f(\mathrm{prox}_f(\mathbf{x})) \quad \text{and} \quad \mathbf{y} - \mathrm{prox}_f(\mathbf{y}) \in \partial f(\mathrm{prox}_f(\mathbf{y})).$$

By applying monotonicity (see Lemma 24), we have

$$((\mathbf{x} - \mathrm{prox}_f(\mathbf{x})) - (\mathbf{y} - \mathrm{prox}_f(\mathbf{y})))^\top (\mathrm{prox}_f(\mathbf{x}) - \mathrm{prox}_f(\mathbf{y})) \geq 0.$$

Rearranging this and applying the Cauchy–Schwartz inequality yields

$$\begin{aligned} \|\mathrm{prox}_f(\mathbf{x}) - \mathrm{prox}_f(\mathbf{y})\|_2^2 &\leq (\mathbf{x} - \mathbf{y})^\top (\mathrm{prox}_f(\mathbf{x}) - \mathrm{prox}_f(\mathbf{y})) \\ &\leq \|\mathbf{x} - \mathbf{y}\|_2 \|\mathrm{prox}_f(\mathbf{x}) - \mathrm{prox}_f(\mathbf{y})\|_2. \qquad \square \end{aligned}$$

- Examples of several proximal operators

(1) Constant function $f(\mathbf{x}) = c$:

$$\text{prox}_f(\mathbf{x}) = \mathbf{x}.$$

(2) $f(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$ with $\lambda > 0$:

$$\begin{aligned}\left[\text{prox}_{\lambda\|\cdot\|_1}(\mathbf{x})\right]_i &= \operatorname*{argmin}_{u\in\mathbb{R}} \left\{\lambda|u| + \frac{1}{2}(u - x_i)^2\right\} \\ &= \begin{cases} x_i - \lambda & \text{if } x_i > \lambda, \\ 0 & \text{if } x_i \in [-\lambda, \lambda], \\ x_i + \lambda & \text{if } x_i < -\lambda, \end{cases}\end{aligned}$$

which is known as *soft-thresholding*.

(3) Let $\|\mathbf{x}\|_0$ denote the number of nonzero components of $\mathbf{x}$. For $f(\mathbf{x}) = \lambda\|\mathbf{x}\|_0$ with $\lambda > 0$:

$$\left[\mathrm{prox}_{\lambda\|\cdot\|_0}(\mathbf{x})\right]_i = \begin{cases} x_i & \text{if } |x_i| > \sqrt{2\lambda}, \\ \{0, x_i\} & \text{if } |x_i| = \sqrt{2\lambda}, \\ 0 & \text{if } |x_i| < \sqrt{2\lambda}, \end{cases}$$

which is known as *hard-thresholding*.

(4) Given a closed convex set Ω, define the *indicator function* $\mathrm{I}_\Omega(\mathbf{x})$

$$\mathrm{I}_\Omega(\mathbf{x}) = \begin{cases} 0, & \text{if } \mathbf{x} \in \Omega, \\ +\infty, & \text{otherwise.} \end{cases}$$

We have

$$\mathrm{prox}_{\mathrm{I}_\Omega}(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{u}} \left\{ \mathrm{I}_\Omega(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 \right\} = \operatorname*{argmin}_{\mathbf{u}\in\Omega} \|\mathbf{u} - \mathbf{x}\|_2,$$

which is simply the projection of $\mathbf{x}$ onto the set Ω.