PART I: THEORETICAL BACKGROUND

Lecture 1–Extended Real-Valued Functions

- An extended real-valued function is a function defined over the entire underlying space that can take any real value, as well as the infinite values $-\infty$ and ∞.
- **Infinite values arithmetic:**

$$\begin{aligned}
a+\infty=\infty+a &= \infty & (-\infty < a < \infty),\\
a-\infty=-\infty+a &= -\infty & (-\infty < a < \infty),\\
a\cdot\infty=\infty\cdot a &= \infty & (0 < a < \infty),\\
a\cdot(-\infty)=(-\infty)\cdot a &= -\infty & (0 < a < \infty),\\
a\cdot\infty=\infty\cdot a &= -\infty & (-\infty < a < 0),\\
a\cdot(-\infty)=(-\infty)\cdot a &= \infty & (-\infty < a < 0),\\
0\cdot\infty=\infty\cdot 0=0\cdot(-\infty)=(-\infty)\cdot 0 &= 0. &
\end{aligned}$$

- For an extended real-valued function $f : \mathbb{E} \to [-\infty, \infty]$, the effective domain or just the domain is the set

$$\mathrm{dom}(f) = \{\mathbf{x} \in \mathbb{E} : f(\mathbf{x}) < \infty\}.$$

- For any subset $C \subseteq \mathbb{E}$, the indicator function of C is

$$\delta_C(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \in C, \\ \infty & \mathbf{x} \notin C. \end{cases}$$

Closedness

- The epigraph of an extended real-valued function $f : \mathbb{E} \to [-\infty, \infty]$ is

$$\text{epi}(f) = \{(\mathbf{x}, y) : f(\mathbf{x}) \le y, \mathbf{x} \in \mathbb{E}, y \in \mathbb{R}\} \subseteq \mathbb{E} \times \mathbb{R}.$$

- $f : \mathbb{R}^n \to [-\infty, \infty]$ is proper if it does not attain the value $-\infty$ and $\text{dom}(f) \neq \emptyset$.
- $f : \mathbb{E} \to [-\infty, \infty]$ is called closed if its epigraph is closed.

Theorem. The indicator function δ_C is closed if and only if C is closed.

Proof.

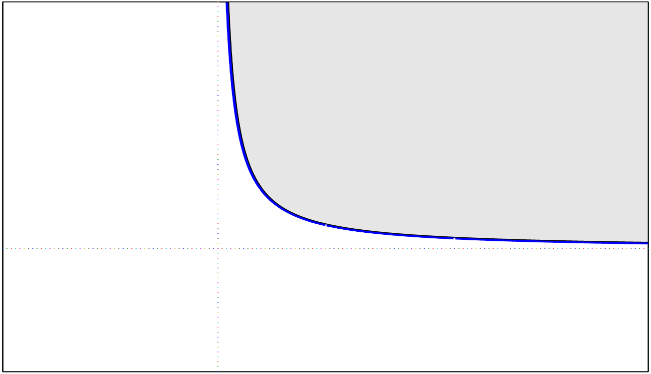
$$\text{epi}(f) = \{(\mathbf{x}, y) \in \mathbb{E} \times \mathbb{R} : \delta_C(\mathbf{x}) \le y\} = C \times \mathbb{R}_+,$$

which is evidently closed if and only if C is closed. $\square$

Example

$$f(x) = \begin{cases} \frac{1}{x}, & x > 0, \\ \infty, & \text{else.} \end{cases}$$

f is closed.



Lower Semicontinuity

Definition

- A function $f : \mathbb{E} \to [-\infty, \infty]$ is called lower semicontinuous at $\mathbf{x} \in \mathbb{E}$ if

$$f(\mathbf{x}) \leq \liminf_{n \to \infty} f(\mathbf{x}_n),$$

for any sequence $\{\mathbf{x}_n\}_{n \geq 1} \subseteq \mathbb{E}$ for which $\mathbf{x}_n \to \mathbf{x}$ as $n \to \infty$.

- $f : \mathbb{E} \to [-\infty, \infty]$ is lower semicontinuous if it is lower semicontinuous at each point in $\mathbb{E}$.

Theorem. Let $f : \mathbb{E} \to [-\infty, \infty]$. Then the following three claims are equivalent:

(i) f is lower semicontinuous.

(ii) f is closed.

(iii) for any $\alpha \in \mathbb{R}$, the level set

$$\text{Lev}(f, \alpha) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \leq \alpha\}$$

is closed.

Operations Preserving Closedness

Theorem.

(a) Let $\mathcal{A} : \mathbb{E} \to \mathbb{V}$ be a linear transformation and $\mathbf{b} \in \mathbb{V}$, and let $f : \mathbb{V} \to (-\infty, \infty]$ be closed. Then the function $g : \mathbb{E} \to [-\infty, \infty]$ given by

$$g(\mathbf{x}) = f(\mathcal{A}(\mathbf{x}) + \mathbf{b})$$

is closed.

(b) Let $f_1, f_2, \ldots, f_m : \mathbb{E} \to (-\infty, \infty]$ be extended real-valued closed functions, and let $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}_+$. Then the function $f = \sum_{i=1}^m \alpha_i f_i$ is closed.

(c) Let $f_i : \mathbb{E} \to (-\infty, \infty], i \in I$ be extended real-valued closed functions, where I is a given index set. Then the function

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

is closed.

Closedness Vs. Continuity

Theorem Let $f : \mathbb{E} \rightarrow [-\infty, \infty]$ be an extended real-valued function that is continuous over its domain, and suppose that $\text{dom}(f)$ is closed. Then f is closed.

Examples

- $f : \mathbb{R} \to (-\infty, \infty]$ is given by

$$f_\alpha(x) = \begin{cases} \alpha, & x = 0, \\ x, & 0 < x \leq 1, \\ \infty, & \text{else.} \end{cases}$$

  for which values of α is the function closed? continuous over its domain?

- Consider the l_0-norm function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(\mathbf{x}) = \|\mathbf{x}\|_0 \equiv \#\{i : x_i \neq 0\}.$$

  f is closed but not continuous.

Weierstrass theorem for closed functions

Theorem. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed function, and assume that C is a compact set satisfying $C \cap \text{dom}(f) \neq \emptyset$. Then

(a) f is bounded below over C.

(b) f attains a minimizer over C.

Convex Extended Real-Valued Functions

- An extended real-valued function is called convex if $\mathrm{epi}(f)$ is convex.
- $f : \mathbb{E} \to (-\infty, \infty]$ is convex $\Leftrightarrow$ dom(f) is convex and the real-valued function $\tilde{f}$: dom$(f) \to \mathbb{R}$ which is the restriction of f to dom(f) is convex over dom(f).
- Result: A proper function $f : \mathbb{E} \to (-\infty, \infty]$ is convex iff

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}) \text{ for all } \lambda \in [0,1], \mathbf{x}, \mathbf{y} \in \mathbb{E}$$

- Jensen's inequality

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

for any $\boldsymbol{\lambda} \in \Delta_k$ (k being an arbitrary positive integer), $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{E}$.

Are Closed and Convex Functions Continuous?

Not in general, but it is correct in the 1D case:

Theorem (continuity of 1D closed convex functions) Let $f : \mathbb{R} \to (-\infty, \infty]$ be a proper closed and convex function whose domain is $\text{dom} f = [a, b]$, where $a, b \in \mathbb{R}, a < b$. Then f is continuous over $[a, b]$.

Proof. technical and long.

Support Functions

- Let $C \subseteq \mathbb{E}$ be nonempty. Then the support function of C, $\sigma_C : \mathbb{E} \to (-\infty, \infty]$ is given by

$$\sigma_C(\mathbf{y}) \equiv \max_{\mathbf{y} \in C} \langle \mathbf{y}, \mathbf{x} \rangle.$$

Theorem. Let $C \subseteq \mathbb{E}$ be a nonempty set. Then σ_C is a closed and convex function.

Proof. σ_C is a maximum of convex functions.

Examples of Support Functions

C	$\sigma_C(\mathbf{y})$	assumptions	Example No.
$\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$	$\max_{i=1,2,\ldots,n}\langle \mathbf{b}_i, \mathbf{y}\rangle$	$\mathbf{b}_i \in \mathbb{E}$	1
K	$\delta_{K^\circ}(\mathbf{y})$	K – cone	2
$\mathbb{R}^n_+$	$\delta_{\mathbb{R}^n_-}(\mathbf{y})$	$\mathbb{E} = \mathbb{R}^n$	3
Δ_n	$\max\{y_1, y_2, \ldots, y_n\}$	$\mathbb{E} = \mathbb{R}^n$	4
$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq \mathbf{0}\}$	$\delta_{\{\mathbf{A}^T\boldsymbol{\lambda}:\boldsymbol{\lambda}\in\mathbb{R}^m_+\}}(\mathbf{y})$	$\mathbb{E} = \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$	5
$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{B}\mathbf{x} = \mathbf{b}\}$	$\langle \mathbf{y}, \mathbf{x}_0\rangle + \delta_{\mathrm{Range}(\mathbf{B}^T)}(\mathbf{y})$	$\mathbb{E} = \mathbb{R}^n$, $\mathbf{B} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{B}\mathbf{x}_0 = \mathbf{b}$	6
$B_{\|\cdot\|}[\mathbf{0}, 1]$	$\|\mathbf{y}\|_*$	$\|\cdot\|$ - arbitrary norm	7

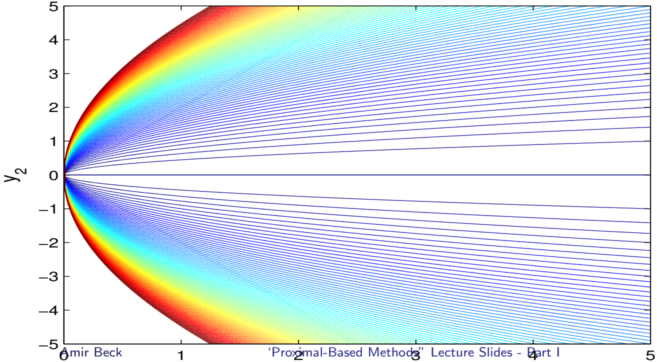
A Discontinuous Closed and Convex Function

If

$$C = \left\{ (x_1, x_2) : x_1 + \frac{x_2^2}{2} \leq 0 \right\}.$$

Then

$$\sigma_C(\mathbf{y}) = \begin{cases} \frac{y_2^2}{2y_1}, & y_1 > 0 \\ 0, & y_1 = y_2 = 0 \\ \infty, & \text{else.} \end{cases}$$



Lecture 2 - Subgradients

- Definition: Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper function, and let $\mathbf{x} \in \text{dom}(f)$. A vector $\mathbf{g} \in \mathbb{E}$ is called a subgradient of f at $\mathbf{x}$ if

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

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

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

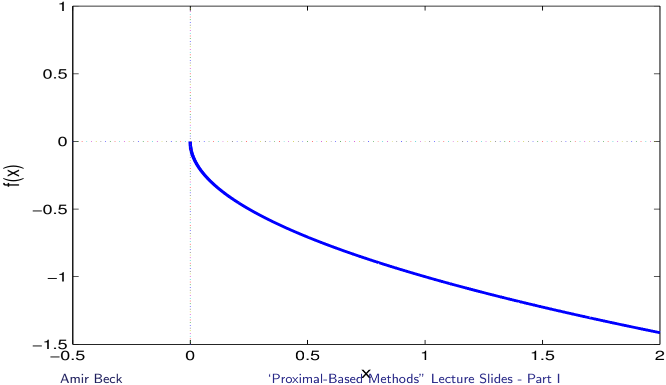
When $\mathbf{x} \notin \text{dom}(f)$, we define $\partial f(\mathbf{x}) = \emptyset$.

Subdifferentiability

- If $\partial f(\mathbf{x}) \neq \emptyset$, f it is called subdifferentiable at $\mathbf{x}$.
- $$\mathrm{dom}(\partial f) \equiv \{\mathbf{x} \in \mathbb{E} : \partial f(\mathbf{x}) \neq \emptyset\}.$$

Example:

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



Existence and Boundedness of $\partial f(\mathbf{x})$

Theorem. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper convex function, and assume that $\tilde{\mathbf{x}} \in \text{int}(\text{dom}(f))$. Then $\partial f(\tilde{\mathbf{x}})$ is nonempty and compact.

Corollary. Let $f : \mathbb{E} \to \mathbb{R}$ be a convex function. Then f is subdifferentiable over $\mathbb{E}$.

The Subdifferential at Differentiability Points

Theorem. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper convex function, and let $\mathbf{x} \in \text{int}(\text{dom}(f))$. If f is differentiable at $\mathbf{x}$, then $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$. Conversely, if f has a unique subgradient at $\mathbf{x}$, then f is differentiable at $\mathbf{x}$ and $\partial f(\mathbf{x}) = \{\nabla f(\mathbf{x})\}$.

Example: $f(\mathbf{x}) = \|\mathbf{x}\|_2$ $(\mathbb{E} = \mathbb{R}^n)$. Then $\partial f(\mathbf{x}) = \begin{cases} \left\{ \frac{\mathbf{x}}{\|\mathbf{x}\|_2} \right\}, & \mathbf{x} \neq \mathbf{0}, \\ B_{\|\cdot\|_2}[\mathbf{0}, 1], & \mathbf{x} = \mathbf{0}. \end{cases}$

Fermat's Optimality Condition

Theorem. Let $f : \mathbb{E} \to (-\infty, \infty]$ be an extended real-valued convex function. Then

$$\mathbf{x}^* \in \operatorname{argmin}\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{E}\} \tag{1}$$

if and only if

$$\mathbf{0} \in \partial f(\mathbf{x}^*)$$

Proof. $\mathbf{0} \in \partial f(\mathbf{x}^*)$ is satisfied iff

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \langle \mathbf{0}, \mathbf{x} - \mathbf{x}^* \rangle \text{ for any } \mathbf{x} \in \operatorname{dom}(f),$$

which is the the same as (1).

Optimality Conditions for the Composite Model (Mixed Convex/Nonconvex)

Theorem. Let $f : \mathbb{E} \to (-\infty, \infty]$ be proper, and let $g : \mathbb{E} \to (-\infty, \infty]$ be a proper convex function such that $\text{dom}(g) \subseteq \text{int}(\text{dom}(f))$. Consider the problem

$$(\text{P}) \quad \min f(\mathbf{x}) + g(\mathbf{x}).$$

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

$$-\nabla f(\mathbf{x}^*) \in \partial g(\mathbf{x}^*). \tag{2}$$

(b) **(necessary and sufficient condition for convex problems)** Suppose that f is convex. If f is differentiable at $\mathbf{x}^* \in \text{dom}(g)$, then $\mathbf{x}^*$ is a global optimal solution of (P) if and only if (2) is satisfied.

Stationarity in Composite Models

$$\text{(P)} \quad \min f(\mathbf{x}) + g(\mathbf{x}).$$

- $f : \mathbb{E} \to (-\infty, \infty]$ proper.
- $g : \mathbb{E} \to (-\infty, \infty]$ proper convex.
- $\text{dom}(g) \subseteq \text{int}(\text{dom}(f))$.

Definition A point $\mathbf{x}^* \in \text{dom}\, g$ in which f is differentiable is called a stationarity point of (P) if $-\nabla f(\mathbf{x}^*) \in \partial g(\mathbf{x}^*)$

Example: If $g(\mathbf{x}) = \delta_C(\mathbf{x})$ for convex C, then stationarity is the same as

$$\langle \nabla f(\mathbf{x}^*), \mathbf{x} - \mathbf{x}^* \rangle \geq 0$$

Example: $\min f(\mathbf{x}) + \lambda \|\mathbf{x}\|_1$ $(f : \mathbb{R}^n \to \mathbb{R})$

Lecture 3 - Conjugate Functions

Definition. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper extended real-valued function. The function $f : \mathbb{E} \to [-\infty, \infty]$ defined by

$$f^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x})\}.$$

is called the conjugate function of f.

Result: Conjugate functions are **always** closed and convex (regardless of the properties of f). Why?

Example: $f = \delta_C$, where $C \subseteq \mathbb{E}$ is nonempty. Then for any $\mathbf{y} \in \mathbb{E}$

$$f^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} \rangle - \delta_C(\mathbf{x})\} = \max_{\mathbf{x} \in C} \langle \mathbf{y}, \mathbf{x} \rangle = \sigma_C(\mathbf{y}).$$

$$\boxed{\delta_C^* = \sigma_C.}$$

The Biconjugate

The conjugacy operation can be invoked twice resulting with the biconjugacy operation. Specifically, for a function f we define

$$f^{**}(\mathbf{x}) = \max_{\mathbf{y}\in\mathbb{E}} \langle \mathbf{x}, \mathbf{y}\rangle - f^*(\mathbf{y})$$

Theorem ($f \geq f^{**}$). Let $f : \mathbb{E} \to [-\infty, \infty]$ be an extended real-valued function. Then $f(\mathbf{x}) \geq f^{**}(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{E}$.

Proof. For any $\mathbf{x} \in \mathbb{E}$:

- $f^*(\mathbf{y}) \geq \langle \mathbf{y}, \mathbf{x}\rangle - f(\mathbf{x})$
- $f(\mathbf{x}) \geq \langle \mathbf{y}, \mathbf{x}\rangle - f^*(\mathbf{y})$.
- $f(\mathbf{x}) \geq \max_{\mathbf{y}\in\mathbb{E}}\{\langle \mathbf{y}, \mathbf{x}\rangle - f^*(\mathbf{y})\} = f^{**}(\mathbf{x})$.

Fenchel's Inequality

Theorem. Let $f : \mathbb{E} \to (-\infty, \infty]$ be an extended real-valued proper function. Then for any $\mathbf{x} \in \mathbb{E}, \mathbf{y} \in \mathbb{E}$

$$f(\mathbf{x}) + f^*(\mathbf{y}) \geq \langle \mathbf{y}, \mathbf{x} \rangle.$$

$f^{**} = f$

Theorem. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a closed and proper extended real-valued function. Then $f^{**} = f$.

Examples

- $f = \sigma_C$, where C is nonempty closed and convex. Then

$$f^* = \sigma_C^* = (\delta_C^*)^* = \delta_C^{**} = \delta_C.$$

- $f(\mathbf{x}) = \max\{x_1, x_2, \ldots, x_n\}$. Then $f^* = \delta_{\Delta_n}$. Why?

Simple Algebraic Rules

function definition	conjugate
$g(\mathbf{x}_1, \ldots, \mathbf{x}_m) = \sum_{i=1}^m f_i(\mathbf{x}_i)$	$g^*(\mathbf{y}_1, \ldots, \mathbf{y}_m) = \sum_{i=1}^m f_i^*(\mathbf{y}_i)$
$g(\mathbf{x}) = \alpha f(\mathbf{x})$	$g^*(\mathbf{y}) = \alpha f^*(\mathbf{y}/\alpha)$
$g(\mathbf{x}) = \alpha f(\mathbf{x}/\alpha)$	$g^*(\mathbf{y}) = \alpha f^*(\mathbf{y})$
$f(\mathcal{A}(\mathbf{x} - \mathbf{a})) + \langle \mathbf{b}, \mathbf{x} \rangle + c$	$f^*\left((\mathcal{A}^T)^{-1}(\mathbf{y} - \mathbf{b})\right) + \langle \mathbf{a}, \mathbf{y} \rangle - c - \langle \mathbf{a}, \mathbf{b} \rangle$

Conjugates of Simple Functions

function (f)	dom f	conjugate (f^*)	assumptions
$\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}+\mathbf{b}^T\mathbf{x}+c$	$\mathbb{R}^n$	$\frac{1}{2}(\mathbf{y}-\mathbf{b})^T\mathbf{A}^{-1}(\mathbf{y}-\mathbf{b})-c$	$\mathbf{A}\succ\mathbf{0}, \mathbf{A}\in\mathbb{R}^{n\times n}$, $\mathbf{b}\in\mathbb{R}^n$, $c\in\mathbb{R}$
$\sum_{i=1}^n x_i\log x_i$	$\mathbb{R}^n_+$	$\sum_{i=1}^n e^{y_i-1}$	–
$\sum_{i=1}^n x_i\log x_i$	Δ_n	$\log\left(\sum_{i=1}^n e^{y_i}\right)$	–
$\log\left(\sum_{i=1}^n e^{x_i}\right)$	$\mathbb{R}^n$	$\sum_{i=1}^n y_i\log y_i$ (dom $f^*=\Delta_n$)	–
$\delta_C(\mathbf{x})$	C	$\sigma_C(\mathbf{x})$	$\emptyset\neq C$ arbitrary
$\sigma_C(\mathbf{x})$	$\mathbb{R}^n$	$\delta_C(\mathbf{x})$	$\emptyset\neq C$ closed, convex
$\|\mathbf{x}\|$	$\mathbb{R}^n$	$\delta_{B_{\|\cdot\|_*}[\mathbf{0},1]}$	$\|\cdot\|$ arbitrary norm
$-\sqrt{1-\|\mathbf{x}\|^2}$	$B_{\|\cdot\|}[\mathbf{0},1]$	$\sqrt{\|\mathbf{y}\|_*^2+1}$	$\|\cdot\|$ arbitrary norm
$\frac{1}{p}\|x\|^p$	$\mathbb{R}$	$\frac{1}{q}\|y\|^q$	$p>1, \frac{1}{p}+\frac{1}{q}=1$
$\frac{1}{2}\|\mathbf{x}\|^2$	$\mathbb{R}^n$	$\frac{1}{2}\|\mathbf{y}\|_*^2$	$\|\cdot\|$ arbitrary norm

Conjugate Subgradient Theorem

Theorem. Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper convex extended real-valued function. The following two claims are equivalent for any $\mathbf{x} \in \mathbb{E}, \mathbf{y} \in \mathbb{E}$:

(i) $\langle \mathbf{x}, \mathbf{y} \rangle = f(\mathbf{x}) + f^*(\mathbf{y})$.

(ii) $\mathbf{y} \in \partial f(\mathbf{x})$.

If, in addition f is closed, then (i) and (ii) are equivalent to

(iii) $\mathbf{x} \in \partial f^*(\mathbf{y})$.

Proof.

- $\mathbf{y} \in \partial f(\mathbf{x})$ iff
$$f(\mathbf{z}) \geq f(\mathbf{x}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle \text{ for all } \mathbf{z} \in \mathbb{E},$$
- iff $\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}) \geq \langle \mathbf{y}, \mathbf{z} \rangle - f(\mathbf{z})$ for all $\mathbf{z} \in \mathbb{E}$.
- iff $\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}) \geq f^*(\mathbf{y})$,
- which combined with Fenchel's inequality is equivalent to $\langle \mathbf{x}, \mathbf{y} \rangle = f(\mathbf{x}) + f^*(\mathbf{y})$.
- If f is closed, then (i) is the same as $\langle \mathbf{x}, \mathbf{y} \rangle = g(\mathbf{y}) + g^*(\mathbf{x})$, where $g = f^*$, and thus it is equivalent to $\mathbf{x} \in \partial g(\mathbf{y}) = \partial f^*(\mathbf{y})$.

Conjugate Subgradient Theorem Contd.

- The equivalence $(i) \Leftrightarrow (ii)$ can be compactly written as
$$\partial f^*(\mathbf{y}) = \operatorname*{argmax}_{\mathbf{x}} \{\langle \mathbf{y}, \mathbf{x}\rangle - f(\mathbf{x})\}$$
- If f is closed, the equivalence $(i) \Leftrightarrow (iii)$ is the same as
$$\partial f(\mathbf{x}) = \operatorname*{argmax}_{\mathbf{y}} \{\langle \mathbf{x}, \mathbf{y}\rangle - f^*(\mathbf{y})\}$$
- If f is differentiable at a point $\mathbf{x}$, then
$$f(\mathbf{x}) + f^*(\nabla f(\mathbf{x})) = \langle \mathbf{x}, \nabla f(\mathbf{x})\rangle.$$
- In particular,
$$\partial f^*(\mathbf{0}) = \operatorname*{argmin}_{\mathbf{x}} f(\mathbf{x}).$$
- Example: $g(\mathbf{x}) = \|\mathbf{x}\|$ (arbitrary norm). What is $\partial g(\mathbf{0})$?
- Exercise 1: f proper closed and convex, $\lambda \in \mathbb{R} \setminus \{0\}, \mathbf{a} \in \mathbb{R}^n$. Define $g(\mathbf{x}) = f(\lambda \mathbf{x} + \mathbf{a})$. Find a formula for g^* in terms of f^*.
- Exercise 2: Let $f(x_1, x_2) = \max\{1 - 2x_1, 0\} + 2e^{3x_2}$. Find f^*
- Exercise 3: Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a closed proper convex function. Suppose that $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, and define $g(\mathbf{x}) = f^*(\mathbf{A}\mathbf{x} + \mathbf{b})$. Explain how to compute a member in $\partial g(\mathbf{x})$ (in terms of $\mathbf{A}, \mathbf{b}$ and f).

Fenchel's Duality Theorem

$$\text{(P)} \ \min_{\mathbf{x}\in\mathbb{E}} f(\mathbf{x}) + g(\mathbf{x}).$$

Lagrangian duality:

- $\min_{\mathbf{x},\mathbf{z}\in\mathbb{E}}\{f(\mathbf{x}) + g(\mathbf{z}) : \mathbf{x} = \mathbf{z}\}$
- Lagrangian:
  $L(\mathbf{x},\mathbf{z};\mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \langle \mathbf{y}, \mathbf{z}-\mathbf{x}\rangle = -\left[\langle \mathbf{y},\mathbf{x}\rangle - f(\mathbf{x})\right] - \left[\langle -\mathbf{y},\mathbf{z}\rangle - g(\mathbf{z})\right].$
- Dual objective function: $q(\mathbf{y}) = \min_{\mathbf{x},\mathbf{z}} L(\mathbf{x},\mathbf{z};\mathbf{y}) = -f^*(\mathbf{y}) - g^*(-\mathbf{y})$

Fenchel's dual problem:

$$\text{(D)} \ \max_{\mathbf{y}\in\mathbb{E}^*}\{-f^*(\mathbf{y}) - g^*(-\mathbf{y})\}.$$

Theorem (Fenchel's duality theorem) Let $f, g : \mathbb{E} \to (-\infty,\infty]$ be proper convex functions. If $\text{ri}(\text{dom}(f)) \cap \text{ri}(\text{dom}(g)) \neq \emptyset$, then

$$\min_{\mathbf{x}\in\mathbb{E}}\{f(\mathbf{x}) + g(\mathbf{x})\} = \max_{\mathbf{y}\in\mathbb{E}^*}\{-f^*(\mathbf{y}) - g^*(-\mathbf{y})\},$$

and the maximum in the right-hand problem is attained whenever it is finite.

Lecture 4 - The Proximal Operator

Definition. Given a closed, proper and convex function g, the proximal mapping of g is defined by

$$\mathrm{prox}_g(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{u}\in\mathbb{E}} \left\{ g(\mathbf{u}) + \frac{1}{2}\|\mathbf{u}-\mathbf{x}\|^2 \right\}.$$

Examples

- **Constant.** If $f \equiv c$ for some $c \in \mathbb{R}$, then

$$\mathrm{prox}_f(\mathbf{x}) = \underset{\mathbf{u}\in\mathbb{E}}{\operatorname{argmin}} \left\{ c + \frac{1}{2}\|\mathbf{u}-\mathbf{x}\|^2 \right\} = \mathbf{x}$$

  The identity mapping.

- **Affine.** Let $f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x}\rangle + b$, where $\mathbf{a} \in \mathbb{E}$ and $b \in \mathbb{R}$. Then

$$\begin{aligned}\mathrm{prox}_f(\mathbf{x}) &= \underset{\mathbf{u}\in\mathbb{E}}{\operatorname{argmin}} \left\{ \langle \mathbf{a}, \mathbf{u}\rangle + b + \frac{1}{2}\|\mathbf{u}-\mathbf{x}\|^2 \right\} \\ &= \mathbf{x} - \mathbf{a}.\end{aligned}$$

- Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$, where $\mathbf{A} \in \mathbb{S}^n_+, \mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The vector $\mathrm{prox}_f(\mathbf{x})$ is the solution of

$$\min_{\mathbf{u}\in\mathbb{E}} \left\{ \frac{1}{2}\mathbf{u}^T\mathbf{A}\mathbf{u} + \mathbf{b}^T\mathbf{u} + c + \frac{1}{2}\|\mathbf{u}-\mathbf{x}\|^2 \right\}.$$

  The optimal solution is attained at $\mathbf{u}$ satisfying $(\mathbf{A}+\mathbf{I})\mathbf{u} = \mathbf{x} - \mathbf{b}$, and hence

$$\mathrm{prox}_f(\mathbf{x}) = \mathbf{u} = (\mathbf{A}+\mathbf{I})^{-1}(\mathbf{x}-\mathbf{b}).$$

The Orthogonal Projection

- Definition. Given a nonempty closed and convex set $C \subseteq \mathbb{E}$ and $\mathbf{x} \in \mathbb{E}$, the orthogonal projection operator $P_C : \mathbb{E} \to C$ is defined by

$$P_C(\mathbf{x}) \equiv \operatorname*{argmin}_{\mathbf{y} \in C} \|\mathbf{y} - \mathbf{x}\|.$$

First projection theorem. Let $C \subseteq \mathbb{E}$ be a nonempty closed convex set. Then $P_C(\mathbf{x})$ is a singleton.

Prox of Indicator = Orthogonal Projection

- If $C \subseteq \mathbb{E}$ is nonempty, then $\mathrm{prox}_{\delta_C} = P_C$

$$\mathrm{prox}_{\delta_C}(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{u}\in\mathbb{E}} \left\{ \delta_C(\mathbf{u}) + \frac{1}{2}\|\mathbf{u}-\mathbf{x}\|^2 \right\} = \operatorname*{argmin}_{\mathbf{u}\in C} \|\mathbf{u}-\mathbf{x}\|^2 = P_C(\mathbf{x}).$$

First prox theorem. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed and convex function. Then $\mathrm{prox}_f(\mathbf{x})$ is a singleton for any $\mathbf{x} \in \mathbb{E}$.

Necessity of the Conditions in the First Prox Theorem

• When f is not convex and/or closed, the prox is not guaranteed to uniquely exist, or even to exist at all.

$$\begin{aligned} g_1(x) &\equiv 0, \\ g_2(x) &= \begin{cases} 0, & x \neq 0, \\ -\lambda, & x = 0, \end{cases} \\ g_3(x) &= \begin{cases} 0, & x \neq 0, \\ \lambda, & x = 0. \end{cases} \end{aligned}$$

$$\mathrm{prox}_{g_1}(x) = x, \mathrm{prox}_{g_2}(x) = \begin{cases} \{0\}, & |x| < \sqrt{2\lambda}, \\ \{x\}, & |x| > \sqrt{2\lambda}, \\ \{0, x\}, & |x| = \sqrt{2\lambda}. \end{cases}, \mathrm{prox}_{g_3}(x) = \begin{cases} \{x\}, & x \neq 0, \\ \emptyset, & x = 0. \end{cases}$$

- Uniquness is not guaranteed in any case.
- Existence is guaranteed whenever f is proper closed and the function $\mathbf{u} \mapsto f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2$ is coercive.

Basic Calculus Rules

$f(\mathbf{x})$	$\mathrm{prox}_f(\mathbf{x})$	assumptions
$\sum_{i=1}^m f_i(\mathbf{x}_i)$	$\mathrm{prox}_{f_1}(\mathbf{x}_1) \times \cdots \times \mathrm{prox}_{f_m}(\mathbf{x}_m)$	
$g(\lambda\mathbf{x} + \mathbf{a})$	$\frac{1}{\lambda}\left[\mathrm{prox}_{\lambda^2 g}(\mathbf{a} + \lambda\mathbf{x}) - \mathbf{a}\right]$	$\lambda \neq 0, \mathbf{a} \in \mathbb{E}$, g proper
$\lambda g(\mathbf{x}/\lambda)$	$\lambda\mathrm{prox}_{g/\lambda}(\mathbf{x}/\lambda)$	$\lambda > 0$, g proper
$g(\mathbf{x}) + \frac{c}{2}\\|\mathbf{x}\\|^2 + \langle \mathbf{a}, \mathbf{x}\rangle + \gamma$	$\mathrm{prox}_{\frac{1}{c+1}g}\left(\frac{\mathbf{x}-\mathbf{a}}{c+1}\right)$	$\mathbf{a} \in \mathbb{E}, c > 0, \gamma \in \mathbb{R}$, g proper
$g(\mathcal{A}(\mathbf{x}) + \mathbf{b})$	$\mathbf{x} + \frac{1}{\alpha}\mathcal{A}^T(\mathrm{prox}_{\alpha g}(\mathcal{A}(\mathbf{x}) + \mathbf{b}) - \mathcal{A}(\mathbf{x}) - \mathbf{b})$	$\mathbf{b} \in \mathbb{R}^m$, $\mathcal{A} : \mathbb{V} \to \mathbb{R}^m$, g closed proper convex, $\mathcal{A} \circ \mathcal{A}^T = \alpha I$, $\alpha > 0$
$g(\\|\mathbf{x}\\|)$	$\mathrm{prox}_g(\\|\mathbf{x}\\|)\frac{\mathbf{x}}{\\|\mathbf{x}\\|}, \quad \mathbf{x} \neq \mathbf{0}$ $\{\mathbf{u} : \\|\mathbf{u}\\| = \mathrm{prox}_g(0)\}, \quad \mathbf{x} = \mathbf{0}$	g proper closed convex, $\mathrm{dom}(g) \subseteq [0, \infty)$

Proof of One Property

Let $g : \mathbb{E} \to (-\infty, \infty]$ be proper, and let $\lambda \neq 0$. Define $f(\mathbf{x}) = \lambda g(\mathbf{x}/\lambda)$. Then $\mathrm{prox}_f(\mathbf{x}) = \lambda \mathrm{prox}_{g/\lambda}(\mathbf{x}/\lambda)$

Proof.

- Note that

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

- Making the change of variables $\mathbf{z} = \frac{\mathbf{u}}{\lambda}$, we can continue to write

$$\begin{aligned}
\mathrm{prox}_f(\mathbf{x}) &= \lambda \operatorname*{argmin}_{\mathbf{z}} \left\{ \lambda g(\mathbf{z}) + \frac{1}{2}\|\lambda \mathbf{z} - \mathbf{x}\|^2 \right\} \\
&= \lambda \operatorname*{argmin}_{\mathbf{z}} \left\{ \lambda^2 \left[\frac{g(\mathbf{z})}{\lambda} + \frac{1}{2}\left\|\mathbf{z} - \frac{\mathbf{x}}{\lambda}\right\|^2 \right] \right\} \\
&= \lambda \operatorname*{argmin}_{\mathbf{z}} \left\{ \frac{g(\mathbf{z})}{\lambda} + \frac{1}{2}\left\|\mathbf{z} - \frac{\mathbf{x}}{\lambda}\right\|^2 \right\} \\
&= \lambda \mathrm{prox}_{g/\lambda}(\mathbf{x}/\lambda).
\end{aligned}$$

Examples or Prox Computations

f	dom f	prox_f	assumptions
$\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}+\mathbf{b}^T\mathbf{x}+c$	$\mathbb{R}^n$	$(\mathbf{A}+\mathbf{I})^{-1}(\mathbf{x}-\mathbf{b})$	$\mathbf{A}\in\mathbb{S}^n_{++}$, $\mathbf{b}\in\mathbb{R}^n$, $c\in\mathbb{R}$
$\lambda\lVert\mathbf{x}\rVert$	$\mathbb{E}$	$\left[1-\frac{\lambda}{\lVert\mathbf{x}\rVert}\right]_+\mathbf{x}$	Euclidean norm, $\lambda>0$
$\lambda\lVert\mathbf{x}\rVert_1$	$\mathbb{R}^n$	$[\lvert\mathbf{x}\rvert-\lambda\mathbf{e}]_+\circ\mathrm{sgn}(\mathbf{x})$	$\lambda>0$
$-\lambda\sum_{j=1}^n\log x_j$	$\mathbb{R}^n_{++}$	$\left(\frac{x_j+\sqrt{x_j^2+4\lambda}}{2}\right)_{j=1}^n$	$\lambda>0$
$\delta_C(\mathbf{x})$	$\mathbb{E}$	$P_C(\mathbf{x})$	$C\subseteq\mathbb{E}$
$\lambda\sigma_C(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x}-\lambda P_C(\mathbf{x}/\lambda)$	C closed and convex
$\lambda\lVert\mathbf{x}\rVert$	$\mathbb{E}$	$\mathbf{x}-\lambda P_{B_{\lVert\cdot\rVert_*}[\mathbf{0},1]}(\mathbf{x}/\lambda)$	arbitrary norm
$\lambda\max\{x_1,x_2,\ldots,x_n\}$	$\mathbb{R}^n$	$\mathbf{x}-\mathrm{prox}_{\Delta_n}(\mathbf{x}/\lambda)$	$\lambda>0$
$\lambda d_C(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x}+\min\left\{\frac{\lambda}{d_C(\mathbf{x})},1\right\}(P_C(\mathbf{x})-\mathbf{x})$	C closed convex
$\frac{\lambda}{2}d_C(\mathbf{x})^2$	$\mathbb{E}$	$\frac{\lambda}{\lambda+1}P_C(\mathbf{x})+\frac{1}{\lambda+1}\mathbf{x}$	C closed convex

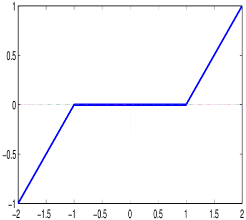
Prox of l_1-Norm

- $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$ $(\lambda > 0)$
- $g(\mathbf{x}) = \sum_{i=1}^n \varphi(x_i)$, where $\varphi(t) = \lambda|t|$.

- $\mathrm{prox}_\varphi(s) = \mathcal{T}_\lambda(s)$, where $\mathcal{T}_\lambda$ is defined as

$$\mathcal{T}_\lambda(y) = [|y| - \lambda]_+ \mathrm{sgn}(y) = \begin{cases} y - \lambda, & y \geq \lambda, \\ 0, & |y| < \lambda, \\ y + \lambda, & y \leq -\lambda \end{cases}$$

is the soft thresholding operator.



- By the separability of the l_1-norm, $\mathrm{prox}_g(\mathbf{x}) = (\mathcal{T}_\lambda(x_j))_{j=1}^n$. We expend the definition of the soft thresholding operator and write

$$\mathrm{prox}_g(\mathbf{x}) = \mathcal{T}_\lambda(\mathbf{x}) \equiv (\mathcal{T}_\lambda(x_j))_{j=1}^n = [|\mathbf{x}| - \lambda\mathbf{e}]_+ \odot \mathrm{sgn}(\mathbf{x}).$$

Moreau Decomposition

Theorem. Let f be a closed, proper and extended real-valued convex function. Then for any $\mathbf{x} \in \mathbb{E}$

$$\mathrm{prox}_f(\mathbf{x}) + \mathrm{prox}_{f^*}(\mathbf{x}) = \mathbf{x}.$$

Extended Moreau decomposition. same setting with $\lambda > 0$. For any $\mathbf{x} \in \mathbb{E}$

$$\mathrm{prox}_{\lambda f}(\mathbf{x}) + \lambda \mathrm{prox}_{\lambda^{-1} f^*}(\mathbf{x}/\lambda) = \mathbf{x}.$$

Exercise: Suppose that g is a proper closed and convex function. Define $G(\mathbf{y}) \equiv g^*(-\mathbf{y})$. For any $\lambda > 0$, write $\mathrm{prox}_{\lambda G}$ in terms of $\mathrm{prox}_{\alpha g}$ for some $\alpha > 0$.

Prox of Support Functions

Let C be a nonempty closed and convex set, and let $\lambda > 0$. Then

$$\operatorname{prox}_{\lambda\sigma_C}(\mathbf{x}) = \mathbf{x} - \lambda P_C(\mathbf{x}/\lambda).$$

Proof. By the extended Moreau decomposition formula

$$\operatorname{prox}_{\lambda\sigma_C}(\mathbf{x}) = \mathbf{x} - \lambda \operatorname{prox}_{\lambda^{-1}\sigma_C^*}(\mathbf{x}/\lambda) = \mathbf{x} - \lambda \operatorname{prox}_{\lambda^{-1}\delta_C}(\mathbf{x}/\lambda) = \mathbf{x} - \lambda P_C(\mathbf{x}/\lambda)$$

Examples:

- $\operatorname{prox}_{\lambda\|\cdot\|_\alpha}(\mathbf{x}) = \mathbf{x} - \lambda P_{B_{\|\cdot\|_{\alpha,*}}[\mathbf{0},1]}(\mathbf{x}/\lambda)$. ($\|\cdot\|_\alpha$ - arbitrary norm)
- $\operatorname{prox}_{\lambda\|\cdot\|_\infty}(\mathbf{x}) = \mathbf{x} - \lambda P_{B_{\|\cdot\|_1}[\mathbf{0},1]}(\mathbf{x}/\lambda)$.

Spectral Functions over $\mathbb{S}^n$

- Given a matrix $\mathbf{X} \in \mathbb{S}^n$, its eigenvalues vector is denoted by $\boldsymbol{\lambda}(\mathbf{X})$, where

$$\lambda_1(\mathbf{X}) \geq \lambda_2(\mathbf{X}) \geq \cdots \geq \lambda_n(\mathbf{X})$$

- F is a **spectral function** if it is of the form

$$F(\mathbf{x}) = g(\lambda_1(\mathbf{X}), \lambda_2(\mathbf{X}), \ldots, \lambda_n(\mathbf{X})),$$

  g is the "outer function".
- if g is symmetric w.r.t. permutations, then F is a **symmetric spectral function**
- **Examples:** $\lambda_{\max}(\mathbf{X})$, $\|\mathbf{X}\|_F$, $\|\mathbf{X}\|_2$, $-\log\det(\mathbf{X})$
- **Spectral Proximal Formula:** $F = f \circ \boldsymbol{\lambda}$ spectral symmetric $\Rightarrow$

$$\mathrm{prox}_F(\mathbf{X}) = \mathbf{U}\,\mathrm{diag}(\mathrm{prox}_f(\boldsymbol{\lambda}(\mathbf{X})))\mathbf{U}^T$$

  where $\mathbf{X} = \mathbf{U}\,\mathrm{diag}(\boldsymbol{\lambda}(\mathbf{X}))\mathbf{U}^T$ (spectral deocomposition)

Spectral Functions over $\mathbb{R}^{m\times n}$

- Given a matrix $\mathbf{X} \in \mathbb{S}^n$, its singular values vector is denoted by $\boldsymbol{\sigma}(\mathbf{X})$, where

$$\sigma_1(\mathbf{X}) \geq \sigma_2(\mathbf{X}) \geq \cdots \geq \sigma_r(\mathbf{X}) \quad r = \min\{m, n\}$$

- F is a **spectral function** if it of the form

$$F(\mathbf{x}) = g(\sigma_1(\mathbf{X}), \sigma_2(\mathbf{X}), \ldots, \sigma_r(\mathbf{X})),$$

  g is the outer function.
- if g is symmetric w.r.t. permutations and $g(\mathbf{x}) \equiv g(|\mathbf{x}|)$, then F is an **absolutely symmetric spectral function**
- **Examples:** $\sigma_{\max}(\mathbf{X})$, $\|\mathbf{X}\|_F$, $\|\mathbf{X}\|_2$, $\|\mathbf{X}\|_* = \sum_{i=1}^r \sigma_i(\mathbf{X})$
- **Spectral Proximal Formula:** $F = f \circ \boldsymbol{\lambda}$ spectral symmetric $\Rightarrow$

$$\mathrm{prox}_F(\mathbf{X}) = \mathbf{U}\,\mathrm{diag}(\mathrm{prox}_f(\boldsymbol{\lambda}(\mathbf{X})))\mathbf{U}^T$$

  where $\mathbf{X} = \mathbf{U}\,\mathrm{diag}(\boldsymbol{\lambda}(\mathbf{X}))\mathbf{U}^T$ (spectral deocomposition)

Exercises

For each of the following functions, find an expression for $\mathrm{prox}_{\lambda f}(\mathbf{x})$ for any $\lambda > 0$ and $\mathbf{x}$.

1. elastic net: $f(\mathbf{x}) = \lambda\|\mathbf{x}\|_2^2 + \mu\|\mathbf{x}\|_1$ $(\lambda, \mu > 0)$ (just $\mathrm{prox}_f(\mathbf{x})$)
2. $f(t) = \max\{0, t\}$
3. $f(t) = \max\{t, 1 - 3t\}$
4. $f(t) = [\max\{0, t\}]^2$
5. $f(\mathbf{x}) = \mathbf{x}_{[1]} \equiv \max\{x_1, x_2, \ldots, x_n\}$
6. $f(\mathbf{x}) = 2x_{[1]} + x_{[2]}$. Write a code implementing prox_f . Use this code to find $\mathrm{prox}_f((2, 1, 4, 1, 2, 1))$ Final answer: $(1.5, 1, 2, 1, 1.5, 1)$
7. $f(\mathbf{x}) = |\mathbf{a}^T\mathbf{x}|$, $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$
8. $f(t) = \begin{cases} \frac{1}{t} & t > 0, \\ \infty & \text{else.} \end{cases}$
9. $f(\mathbf{X}) = \begin{cases} \mathrm{tr}(\mathbf{X}^{-1}) & \mathbf{X} \succ \mathbf{0} \\ \infty & \text{else} \end{cases}$ (over $\mathbb{S}^n$). Write a code implementing prox_f .

   Use this code to find $\mathrm{prox}_f\left[\begin{pmatrix} 3 & 1 \\ 1 & 4 \end{pmatrix}\right]$ Final answer: $\begin{pmatrix} 3.1251 & 0.9511 \\ 0.9511 & 4.0762 \end{pmatrix}$
10. $f(\mathbf{x}) = (\|\mathbf{x}\|_2 - 1)^2$

References

- A. Beck, *First Order Methods in Optimization* (2017).
- D. P. Bertsekas, A. Nedic and A. E. Ozdaglar, *Convex analysis and optimization* (2013).
- J. M. Borwein and A. S. Lewis, *Convex analysis and nonlinear optimization* (2006).
- P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward backward splitting*, Multiscale Model. Simul. (2005).
- J. B. Hiriart-Urruty and C. Lemarechal. *Convex analysis and minimization algorithms. I* (1996).
- J. J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France (1965).
- Y. Nesterov. *Introductory lectures on convex optimization* (2004).
- N. Parikh and S. Boyd, *Proximal algorithms*, Foundations and Trends in Optimization (2014).
- R. T. Rockafellar, *Convex analysis* (1970).