# The Proximal Operator

**Instructor: Jin Zhang**

Department of Mathematics
Southern University of Science and Technology

*Fall 2023*

# Contents

**Definition (proximal mapping)**

Given $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the proximal mapping of $f$ is the operator given by

$$\mathrm{prox}_f(x) = \arg\min_{u \in \mathbb{R}^n} \left\{ f(u) + \frac{1}{2} \|u - x\|^2 \right\} \text{ for any } x \in \mathbb{R}^n.$$

Let $0 < \alpha < \infty$, the mapping $\mathcal{H}_\alpha : \mathbb{R} \rightrightarrows \mathbb{R}$ is the so-called hard thresholding operator defined by

$$\mathcal{H}_\alpha(s) = \begin{cases} \{0\}, & |s| < \alpha, \\ \{s\}, & |s| > \alpha, \\ \{0, s\} & |s| = \alpha. \end{cases}$$

**Example**

Let $\lambda > 0$, consider the univariate function $g : \mathbb{R} \to \mathbb{R}$, given by $g(x) = 0$ if $x \neq 0$ and $g(x) = -\lambda$ otherwise. Then

$$\text{prox}_g(x) = \mathcal{H}_{\sqrt{2\lambda}}(x).$$

**Theorem (first prox theorem)**

*Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex. Then $\mathrm{prox}_f(x)$ is a singleton for any $x \in \mathbb{R}^n$.*
*In this case, we treat $\mathrm{prox}_f$ as a single-valued mapping from $\mathbb{R}^n \to \mathbb{R}^n$, meaning that we write $\mathrm{prox}_f(x) = y$ and not $\mathrm{prox}_f(x) = \{y\}$.*

Recall: A proper function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is coercive if

$$\lim_{\|x\| \to \infty} g(x) = \infty.$$

**Theorem (nonemptiness of the prox under closedness and coerciveness)**

*Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and closed, and assume that the following condition is satisfied:*

$$\text{the function } u \mapsto f(u) + \frac{1}{2}\|u - x\|^2 \text{ is coercive for any } x \in \mathbb{R}^n.$$

*Then $\mathrm{prox}_f(x)$ is nonempty for any $x \in \mathbb{R}^n$.*

### Example (Affine)

Let $f(x) = \langle a, x \rangle + b$, where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then

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

### Example (Convex Quadratic)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(x) = \frac{1}{2}x^T Ax + b^T x + c$, where $0 \leq A \in \mathbb{R}^{n \times n}$ is symmetric, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Then

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

## Example (One-Dimensional Examples)

1. Let $g_1(x) = \mu x$ if $x > 0$ and $g_1(x) = \infty$ otherwise with $\mu \in \mathbb{R}$. Then

$$\mathrm{prox}_{g_1}(x) = [x - \mu]_+ .$$

2. Let $g_2(x) = \lambda |x|$ with $0 < \lambda < \infty$. Then

$$\mathrm{prox}_{g_2}(x) = [|x| - \lambda]_+ \mathrm{sgn}(x).$$

   By the way, the univariate function $\mathcal{T}_\lambda(\cdot) = [|\cdot| - \lambda]_+ \mathrm{sgn}(\cdot)$ is called the soft thresholding function.

3. Let $g_3(x) = \lambda x^3$ if $x \geq 0$ and $g_3(x) = \infty$ otherwise with $0 \leq \lambda < \infty$. Then

$$\mathrm{prox}_{g_3}(x) = \frac{-1 + \sqrt{1 + 12\lambda [x]_+}}{6\lambda}.$$

## Example (One-Dimensional Examples)

4. Let $g_4(x) = -\lambda \log x$ if $x > 0$ and $g_4(x) = \infty$ otherwise with $0 \leq \lambda < \infty$. Then

$$\mathsf{prox}_{g_4}(x) = \frac{x + \sqrt{x^2 + 4\lambda}}{2}.$$

5. Let $g_5(x) = \delta_{[0,\eta]\cap\mathbb{R}}(x)$ with $0 \leq \eta \leq \infty$. Then

$$\mathsf{prox}_{g_5}(x) = \min\left\{\max\left\{x, 0\right\}, \eta\right\}.$$

**Theorem (prox of separable functions)**

*Suppose that* $f : \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m} \to \overline{\mathbb{R}}$ *is given by*

$$f(x_1, \cdots, x_m) = \sum_{i=1}^{m} f_i(x_i) \text{ for any } x_i \in \mathbb{R}^{n_i}, \quad i = 1, \cdots, m.$$

*Then for any* $x_1 \in \mathbb{R}^{n_1}, \cdots, x_m \in \mathbb{R}^{n_m}$,

$$\mathrm{prox}_f(x_1, \cdots, x_m) = \mathrm{prox}_{f_1}(x_1) \times \cdots \times \mathrm{prof}_{f_m}(x_m).$$

**Remark**

If $f : \mathbb{R}^n \to \mathbb{R}$ is closed convex and separable,

$$f(x) = \sum_{i=1}^{n} f_i(x_i),$$

with $f_i : \mathbb{R} \to \mathbb{R}$ being closed and convex univariate functions. Then

$$\mathrm{prox}_f(x) = \left(\mathrm{prox}_{f_i}(x_i)\right)_{i=1}^{n}.$$

## Example

- **($l_1$-norm).** Let $g(x) = \lambda \|x\|_1$ for any $x \in \mathbb{R}^n$ with $0 < \lambda < \infty$,

$$\mathrm{prox}_g(x) = \left(\mathcal{T}_\lambda(x_i)\right)_{i=1}^{\infty}.$$

- **($l_0$-norm).** Let $f(x) = \lambda \|x\|_0$ for any $x \in \mathbb{R}^n$ with $0 < \lambda < \infty$,

$$\mathrm{prox}_f(x) = \mathcal{H}_{\sqrt{2\lambda}}(x_1) \times \cdots \times \mathcal{H}_{\sqrt{2\lambda}}(x_n).$$

## Theorem (scaling and translation)

*Let* $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be proper. Let* $\lambda \neq 0$ *and* $a \in \mathbb{R}^n$*. Define* $f(x) = g(\lambda x + a)$*. Then*

$$\text{prox}_f(x) = \frac{1}{\lambda}\left(\text{prox}_{\lambda^2 g}(\lambda x + a) - a\right).$$

## Theorem (prox of $\lambda g(\cdot/\lambda)$)

*Let* $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be proper. Let* $\lambda \neq 0$ *and define* $f(x) = \lambda g(\frac{x}{\lambda})$*. Then*

$$\text{prox}_f(x) = \lambda \text{prox}_{\frac{g}{\lambda}}\left(\frac{x}{\lambda}\right).$$

### Theorem (quadratic perturbation)

*Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper, and let $f(x) = g(x) + \frac{c}{2}\|x\|^2 + \langle a, x\rangle + \gamma$, where $c > 0$, $a \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}$. Then*

$$\text{prox}_f(x) = \text{prox}_{\frac{1}{c+1}g}\left(\frac{x-a}{c+1}\right).$$

### Theorem (composition with an affine mapping)

*Let $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ be proper closed and convex, and let*

$$f(x) = g\left(\mathcal{A}(x) + b\right),$$

*where $b \in \mathbb{R}^m$ and $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation satisfying $\mathcal{A} \circ \mathcal{A}^T = \alpha\mathcal{I}$ for some constant $\alpha > 0$. Then*

$$\text{prox}_f(x) = x + \frac{1}{\alpha}\mathcal{A}^T\left(\text{prox}_{\alpha g}\left(\mathcal{A}(x) + b\right) - \mathcal{A}(x) - b\right).$$

**Theorem**

*Let $C \subset \mathbb{R}^n$ be nonempty. Then* $\text{prox}_{\delta_C}(x) = P_C(x)$ *for any* $x \in \mathbb{R}^n$.

**Theorem (first projection theorem)**

*Let $C \subset \mathbb{R}^n$ be nonempty closed and convex. Then $P_C(x)$ is a singleton for any $x \in \mathbb{R}^n$.*

### Lemma

*Following are pairs of nonempty closed and convex sets and their corresponding orthogonal projections:*

| | | |
|---|---|---|
| *nonnegative orthant* | $\mathbb{R}^n_+$, | $[x]_+$, |
| *box* | *Box*$[l, u]$, | $\left(\min\left\{\max\left\{x_i, l_i\right\}, u_i\right\}\right)_{i=1}^n$, |
| *affine set* | $\{x \in \mathbb{R}^n : Ax = b\}$, | $x - A^T\left(AA^T\right)^{-1}(Ax - b)$, |
| $l_2$ *ball* | $\mathbb{B}_{\|\cdot\|_2}[c, r]$, | $c + \frac{r}{\max\left\{\|x-c\|_2, r\right\}}(x - c)$, |
| *half-space* | $\left\{x \in \mathbb{R}^n : a^T x \leq \alpha\right\}$, | $x - \frac{\left[a^T x - \alpha\right]_+}{\|a\|^2} a$, |

*where* $l \in [-\infty, \infty)^n$, $u \in (-\infty, \infty]^n$ *are such that* $l \leq u$, $A \in \mathbb{R}^{m \times n}$ *has full row rank,* $b \in \mathbb{R}^m$, $r > 0$, $a \in \mathbb{R}^n \setminus \{0\}$, *and* $\alpha \in \mathbb{R}$.

### Theorem (projection onto the intersection of a hyperplane and a box)

*Let* $C \subseteq \mathbb{R}^n$ *be given by*

$$C = H_{a,b} \cap \mathrm{Box}[l,u] = \left\{ x \in \mathbb{R}^n \middle| a^T x = b, l \le x \le u \right\},$$

*where* $a \in \mathbb{R}^n \setminus \{0\}$, $b \in \mathbb{R}$, $l \in [-\infty, \infty)^n$, $u \in (-\infty, \infty]^n$. *Assume that* $C \neq \varnothing$. *Then*

$$P_C(x) = P_{\mathrm{Box}[l,u]}(x - \mu^* a),$$

*where* $\mu^*$ *is a solution of the equation*

$$a^T P_{\mathrm{Box}[l,u]}(x - \mu a) = b.$$

### Corollary (orthogonal projection onto the unit simplex)

*Recall that the* $n$*-length column vectors of all ones is denoted by* $e \in \mathbb{R}^n$. *Then* $P_{\Delta_n}(x) = [x - \mu^* e]_+$, *where* $\mu^*$ *is a root of the equation* $e^T [x - \mu e]_+ = 1$.

**Theorem (orthogonal projection onto level sets)**

*Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper closed and convex, and $\alpha \in \mathbb{R}$. Assume that there exists $\hat{x} \in \mathbb{R}^n$ for which $f(\hat{x}) < \alpha$. Then*

$$P_{\mathrm{Lev}(f,\alpha)}(x) = \begin{cases} P_{\mathrm{dom}(f)}(x), & f\left(P_{\mathrm{dom}(f)}(x)\right) < \alpha, \\ \mathrm{prox}_{\lambda^* f}(x), & \text{else}, \end{cases}$$

*where $\lambda^*$ is any positive root of the equation*

$$\varphi(\lambda) \equiv f\left(\mathrm{prox}_{\lambda f}(x)\right) - \alpha = 0.$$

*In addition, the function $\varphi$ is nonincreasing.*

## Example (projection onto the intersection of a half-space and a box)

Let

$$C = H^{-}_{a,b} \cap \mathrm{Box}\,[l,u] = \left\{ x \in \mathbb{R}^n \middle| a^T x \le b, l \le x \le u \right\},$$

where $a \in \mathbb{R}^n \backslash \{0\}$, $b \in \mathbb{R}$, $l \in [-\infty, \infty)^n$, $u \in (-\infty, \infty]^n$. Assume that $C \neq \varnothing$. Then

$$P_C(x) = \begin{cases} P_{\mathrm{Box}[l,u]}(x), & a^T P_{\mathrm{Box}[l,u]}(x) \le \alpha, \\ P_{\mathrm{Box}[l,u]}\left(x - \lambda^* a\right), & a^T P_{\mathrm{Box}[l,u]}(x) > \alpha, \end{cases}$$

where $\lambda^*$ is any positive root of the nonincreasing function

$$\varphi(\lambda) = a^T P_{\mathrm{Box}[l,u]}\left(x - \lambda a\right) - b.$$

Define the soft thresholding mapping $\mathcal{T}_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ associated with $\lambda > 0$ by

$$\mathcal{T}_\lambda(x) = [|x| - \lambda e]_+ \odot \mathrm{sgn}(x).$$

**Example (projection onto the $l_1$ ball)**

Let $C = \mathbb{B}_{\|\cdot\|_1}[0, \alpha]$ where $\alpha > 0$. Then

$$P_C(x) = \begin{cases} x, & \|x\|_1 \le \alpha, \\ \mathcal{T}_{\lambda^*}, & \|x\|_1 > \alpha, \end{cases}$$

where $\lambda^*$ is any positive root of the nonincreasing function

$$\varphi(\lambda) = \|\mathcal{T}_\lambda(x)\|_1 - \alpha.$$

Define the two-sided soft thresholding mapping $\mathcal{S}_{a,b}$ associated with $a, b \in (-\infty, \infty]^n$ by

$$\mathcal{S}_{a,b}(x) = \left(\min\left\{\max\left\{|x_i| - a_i, 0\right\}, b_i\right\} \operatorname{sgn}(x_i)\right)_{i=1}^{n}.$$

**Example** ( projection onto the intersection of weighted $l_1$ ball and a box)

Let

$$C = \left\{ x \in \mathbb{R}^n \middle| \sum_{i=1}^{n} w_i |x_i| \le \beta, -\alpha \le x \le \alpha \right\},$$

where $w \in \mathbb{R}_+^n$, $\alpha \in [0, \infty]^n$, and $\beta \in \mathbb{R}_{++}$ Then

$$P_C(x) = \begin{cases} P_{\mathrm{Box}[-\alpha,\alpha](x)}, & w^T \left|\mathcal{S}_{\lambda w,\alpha}(x)\right|_1 \le \beta, \\ \mathcal{S}_{\lambda^* w,\alpha}(x), & w^T \left|\mathcal{S}_{\lambda w,\alpha}(x)\right|_1 > \beta, \end{cases}$$

where $\lambda^*$ is any positive root of the nonincreasing function

$$\varphi(\lambda) = w^T \left|\mathcal{S}_{\lambda w,\alpha}(x)\right|_1 - \beta.$$

## Example

Let

$$C = \left\{ x \in \mathbb{R}^n_{++} \,\middle|\, \prod_{i=1}^{n} x_i \geq \alpha \right\},$$

where $\alpha > 0$. Then

$$P_C(x) = \begin{cases} x, & x \in C, \\ \left( \left( \frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2} \right)^n \right)_{j=1}, & x \notin C, \end{cases}$$

where $\lambda^*$ is any positive root of the nonincreasing function

$$\varphi(\lambda) = -\sum_{j=1}^{n} \log \left( \frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2} \right) + \log \alpha.$$

**Theorem (orthogonal projection onto epigraphs)**

*Let $g : \mathbb{R}^n \to \mathbb{R}$ be real-valued and convex. Then*

$$P_{\text{epi}(g)}((x, s)) = \begin{cases} (x, s), & g(x) \leq s, \\ \left(\text{prox}_{\lambda^* g}(x), s + \lambda^*\right), & g(x) > s, \end{cases}$$

*where $\lambda^*$ is any positive root of the function*

$$\psi(\lambda) = g\left(\text{prox}_{\lambda g}(x)\right) - \lambda - s.$$

*In addition, $\psi$ is nonincreasing.*

## Example (projection onto the Lorentz cone)

Consider the Lorentz cone

$$L^n = \left\{ (x, t) \in \mathbb{R}^n \times \mathbb{R} \middle| \, \|x\|_2 \leq t \right\}.$$

Then

$$P_{L^n}(x, s) = \begin{cases} \left( \frac{\|x\|_2 + s}{2\|x\|_2}, \frac{\|x\|_2 + s}{2} \right), & \|x\|_2 > |s|, \\ (0, 0), & s < \|x\|_2 < -s, \\ (x, s), & \|x\|_2 \leq s. \end{cases}$$

## Example (projection onto the epigraph of the $l_1$-norm)

Let

$$C = \left\{ (x,t) \in \mathbb{R}^n \times \mathbb{R} \middle| \, \|x\|_1 \leq t \right\}.$$

Then

$$P_C((x,s)) = \begin{cases} (x,s), & \|x\|_1 \leq s, \\ \left(\mathcal{T}_{\lambda^*}(x), s+\lambda^*\right), & \|x\|_1 > s, \end{cases}$$

where $\lambda^*$ is any positive root of the nonincreasing function

$$\psi(\lambda) = \|\mathcal{T}_\lambda(x)\|_1 - \lambda - s.$$

# 5. The Second Prox Theorem

**Theorem (second prox theorem)**

*Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex. Then for any $x, u \in \mathbb{R}^n$, the following three claims are equivalent:*

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

(ii). $x - u \in \partial f(u)$.

(iii). $\langle x - u, y - u \rangle \leq f(y) - f(u)$ *for any* $y \in \mathbb{R}^n$.

**Corollary:**

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex. Then $x$ is a minimizer of $f$ if and only if $x = \text{prox}_f(x)$.

**Theorem (first projection theorem)**

*Let $C \subset \mathbb{R}^n$ be nonempty closed and convex. Then $P_C(x)$ is a singleton for any $x \in \mathbb{R}^n$.*

**Theorem (second projection theorem)**

*Let $C \subset \mathbb{R}^n$ be nonempty closed and convex. Let $u \in \mathbb{R}^n$. Then $u = P_C(x)$ if and only if*

$$\langle x - u, y - u \rangle \leq 0 \text{ for any } y \in C.$$

**Theorem (firm nonexpansivity of the prox operator)**

*Let $f$ be proper closed and convex. Then for any $x, y \in \mathbb{R}^n$,*

(a). ***(firm nonexpansivity)***

$$\langle x - y, \mathrm{prox}_f(x) - \mathrm{prox}_f(y) \rangle \geq \left\| \mathrm{prox}_f(x) - \mathrm{prox}_f(y) \right\|^2 .$$

(b). ***(nonexpansivity)***

$$\left\| \mathrm{prox}_f(x) - \mathrm{prox}_f(y) \right\| \leq \| x - y \| .$$

## Lemma (prox of the distance function).

Let $C \subset \mathbb{R}^n$ be nonempty closed and convex. Let $\lambda > 0$. Then for any $x \in \mathbb{R}^n$,

$$\mathrm{prox}_{\lambda d_C}(x) = \begin{cases} (1-\theta)x + \theta P_C(x), & d_C(x) > \lambda, \\ P_C(x), & d_C(x) \leq \lambda, \end{cases}$$

where

$$\theta = \frac{\lambda}{d_C(x)}.$$

# 6. Moreau Decomposition

**Theorem (Moreau decomposition)**

*Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex. Then for any $x \in \mathbb{R}^n$,*

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

**Theorem (extended Moreau decomposition).**

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex, and let $\lambda > 0$. Then for any $x \in \mathbb{R}^n$,

$$\text{prox}_{\lambda f}(x) + \text{prox}_{\lambda^{-1} f^*}\left(\frac{x}{\lambda}\right) = x.$$

## Theorem: prox of support functions

Let $C \subset \mathbb{R}^n$ be nonempty closed and convex, and let $\lambda > 0$. Then for any $x \in \mathbb{R}^n$,

$$\mathrm{prox}_{\lambda\sigma_C}(x) = x - \lambda P_C\left(\frac{x}{\lambda}\right).$$

## Example (prox of norms)

$$\text{prox}_{\lambda\|\cdot\|}(x) = x - \lambda P_{\mathbb{B}_{\|\cdot\|_*}[0,1]}\left(\frac{x}{\lambda}\right).$$

$$\text{prox}_{\lambda\|\cdot\|_\infty}(x) = x - \lambda P_{\mathbb{B}_{\|\cdot\|_1}[0,1]}\left(\frac{x}{\lambda}\right).$$

## Example (prox of max function)

Let $\max(x) = \max\{x_1, \cdots, x_n\}$. Then

$$\text{prox}_{\lambda\max(\cdot)}(x) = x - \lambda P_{\Delta_n}\left(\frac{x}{\lambda}\right).$$

## Example (prox of the sum-of-$k$-largest-values function)

Let

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[n]},$$

where $k \in \{1, 2, \cdots, n\}$ and for any $i$, $x_{[i]}$ denotes $i$th largest value in the vector $x$. Then

$$\text{prox}_{\lambda f}(x) = x - \lambda P_{\{y: e^T y = k, 0 \leq y \leq e\}} \left( \frac{x}{\lambda} \right).$$

## Example (prox of the sum-of-$k$-largest-absolute-values function)

Let

$$f(x) = \left| x_{\langle 1 \rangle} \right| + \left| x_{\langle 2 \rangle} \right| + \cdots + \left| x_{\langle n \rangle} \right|,$$

where $k \in \{1, 2, \cdots, n\}$ and for any $i$, $x_{\langle 1 \rangle}$ denotes $i$th largest absolute value in the vector $x$. Then

$$\text{prox}_{\lambda f}(x) = x - \lambda P_{\{z: \|z\|_1 \leq k, -e \leq z \leq e\}} \left( \frac{x}{\lambda} \right).$$

**Definition: (Moreau envelope)**

Given a proper closed convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\mu > 0$, the Moreau envelope of $f$ is the function

$$M_f^{\mu}(x) = \min_{u \in \mathbb{R}^n} \left\{ f(u) + \frac{1}{2\mu} \|x - u\|^2 \right\},$$

where the parameter $\mu$ is called the smoothing parameter. By the first prox theorem, $M_f^{\mu}(x)$ is always a real number and

$$M_f^{\mu}(x) = f\left(\mathrm{prox}_{\mu f}(x)\right) + \frac{1}{2\mu} \left\|x - \mathrm{prox}_{\mu f}(x)\right\|^2.$$

Recall: (first prox theorem). Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex. Then $\mathrm{prox}_g(x)$ is a singleton for any $x \in \mathbb{R}^n$.

## Example

Moreau envelope of indicators Let $C \subset \mathbb{R}^n$ be nonempty closed and convex. Then for any $x \in \mathbb{R}^n$,

$$M^{\mu}_{\delta_C} = \frac{1}{2\mu} d_C^2.$$

## Example (Huber function)

For any $\mu > 0$,

$$M^{\mu}_{\|\cdot\|} = H_{\mu} = \begin{cases} \frac{1}{2\mu} \|x\|^2, & \|x\| \leq \mu, \\ \|x\| - \frac{\mu}{2}, & \|x\| > \mu. \end{cases}$$

## Theorem

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex, and let $\omega_\mu(x) = \frac{1}{2\mu} \|x\|^2$, where $\mu > 0$. Then

(a). $M_f^\mu = f \Box \omega_\mu$;

(b). $M_f^\mu : \mathbb{R}^n \to \mathbb{R}$ is real-valued and convex.

## Corollary

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex, and let $\omega_\mu(x) = \frac{1}{2\mu} \|x\|^2$, where $\mu > 0$. Then

$$\left(M_f^\mu\right)^* = f^* + \omega_{\frac{1}{\mu}}.$$

**Lemma**

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex, and let $\lambda, \mu > 0$. Then for any $x \in \mathbb{R}^n$,

$$\lambda M_f^{\mu}(x) = M_{\lambda f}^{\mu/\lambda}(x).$$

**Theorem (Moreau envelope of separable functions).**

Suppose that $f : \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m} \to \overline{\mathbb{R}}$ is given by

$$f(x_1, \cdots, x_m) = \sum_{i=1}^{m} f_i(x_i) \text{ for any } x_i \in \mathbb{R}^{n_i}, \quad i = 1, \cdots, m,$$

with $f_i : \mathbb{R}^{n_m} \to \overline{\mathbb{R}}$ being proper closed and convex for any $i$. Then given $\mu > 0$, for any $x_1 \in \mathbb{R}^{n_1}, \cdots, x_m \in \mathbb{R}^{n_m}$,

$$M_f^{\mu}(x_1, \cdots, x_m) = \sum_{i=1}^{m} M_{f_i}^{\mu}(x_i).$$

## Example (Moreau envelope of the $l_1$-norm)

For any $\mu > 0$,

$$M^{\mu}_{\|\cdot\|_1} = \sum_{i=1}^{n} H_{\mu}(x_i).$$

**Theorem (smoothness of the Moreau envelope).**

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. Let $\mu > 0$. Then $M_f^\mu$ is $\frac{1}{\mu}-$smooth over $\mathbb{R}^n$, and for any $x \in \mathbb{R}^n$,

$$\nabla M_f^\mu(x) = \frac{1}{\mu}\left(x - \mathrm{prox}_{\mu f}(x)\right).$$

**Example ($1-$smoothness of $\frac{1}{2}d_C^2$).**

Let $C \subset \mathbb{R}^n$ be nonempty closed and convex. Recall that $\frac{1}{2}d_C^2 = M_{\delta_C}^1$. Then $\frac{1}{2}d_C^2$ is $1-$smooth and

$$\nabla\left(\frac{1}{2}d_C^2\right)(x) = x - \mathrm{prox}_{\delta_C}(x) = x - P_C(x).$$

**Theorem (prox of Moreau envelope)**

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex, and let $\mu > 0$. Then for any $x \in \mathbb{R}^n$,

$$\operatorname{prox}_{M_f^\mu}(x) = x + \frac{1}{\mu+1}\left(\operatorname{prox}_{(\mu+1)f}(x) - x\right).$$

**Corollary**

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex, and let $\lambda, \mu > 0$. Then for any $x \in \mathbb{R}^n$,

$$\operatorname{prox}_{\lambda M_f^\mu}(x) = x + \frac{\lambda}{\mu+\lambda}\left(\operatorname{prox}_{(\mu+\lambda)f}(x) - x\right).$$

**Example: (prox of $\frac{\lambda}{2}d_C^2$).**

Let $C \subset \mathbb{R}^n$ be nonempty closed and convex and let $\lambda > 0$. Then

$$\mathrm{prox}_{\frac{\lambda}{2}d_C^2}(x) = \frac{\lambda}{\lambda+1}P_C(x) + \frac{1}{\lambda+1}x.$$

**Theorem: (Moreau envelope decomposition).**

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper closed and convex, and let $\mu > 0$. Then for any $x \in \mathbb{R}^n$,

$$M_f^{\mu}(x) + M_{f^*}^{1/\mu}\left(\frac{x}{\mu}\right) = \frac{1}{2\mu}\left\|x\right\|^2.$$