

Proximal Gradient Method

Lecturer: Jiaming Liang

September 25, 2025

1 Proximal operator

Definition 1. Given a function f , the proximal mapping of f is given by

$$\text{prox}_f(x) = \operatorname{argmin}_{u \in \mathbb{R}^n} \left\{ f(u) + \frac{1}{2} \|u - x\|^2 \right\}, \quad \forall x \in \mathbb{R}^n.$$

Note that if f is closed and convex then $\text{prox}_f(x)$ is a singleton for any $x \in \mathbb{R}^n$.

Example: soft-thresholding, for some $\lambda > 0$, the proximal mapping for the one-dimensional function $\lambda |\cdot|$ is

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

Hence, the proximal mapping for $f(x) = \lambda \|x\|_1$ is

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

where \odot denotes componentwise multiplication.

Theorem 1. Let $Q \subseteq \mathbb{R}^n$ be nonempty. Then, $\text{prox}_{I_Q}(x) = \text{proj}_Q(x)$ for any $x \in \mathbb{R}^n$. Let $Q \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Then, $\text{prox}_{I_Q}(x) = \text{proj}_Q(x)$ is a singleton for any $x \in \mathbb{R}^n$.

Theorem 2. Let f be a closed and convex function. Then for any $x, y \in \mathbb{R}^n$, we have

$$(i) \quad \|\text{prox}_f(x) - \text{prox}_f(y)\|^2 \leq \langle \text{prox}_f(x) - \text{prox}_f(y), x - y \rangle;$$

$$(ii) \quad \|\text{prox}_f(x) - \text{prox}_f(y)\| \leq \|x - y\|.$$

Proof. (a) Let $u = \text{prox}_f(x)$ and $v = \text{prox}_f(y)$. It follows from the definition of proximal mapping that

$$u = \operatorname{argmin}_{w \in \mathbb{R}^n} \left\{ f(w) + \frac{1}{2} \|w - x\|^2 \right\}$$

and

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

The inclusion is equivalent to

$$f(w) \geq f(u) + \langle x - u, w - u \rangle \quad \forall w \in \mathbb{R}^n.$$

Taking $w = v$, we have

$$f(v) \geq f(u) + \langle x - u, v - u \rangle.$$

Following the same argument for $v = \text{prox}_f(y)$, we have

$$f(u) \geq f(v) + \langle y - v, u - v \rangle$$

Adding the above two inequalities, we obtain

$$0 \geq \langle y - x + u - v, u - v \rangle,$$

i.e.,

$$\langle x - y, u - v \rangle \geq \|u - v\|^2.$$

Plugging $u = \text{prox}_f(x)$ and $v = \text{prox}_f(y)$ into the above inequality, we prove (a).

(b) This statement simply follows from (a) using the Cauchy-Schwarz inequality. \square

2 Moreau envelope

Theorem 3. (Moreau decomposition) Let f be a closed and convex function. Then for any $x \in \mathbb{R}^n$, we have

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

Proof. Let $u = \text{prox}_f(x)$. It is equivalent to $x - u \in \partial f(u)$. Using Theorem 2 of Lecture 5, we have $u \in \partial_{f^*}(x - u)$, which is equivalent to $x - u = \text{prox}_{f^*}(x)$. Therefore,

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

\square

Theorem 4. (extended Moreau decomposition) Let f be a closed and convex function and $\lambda > 0$. Then for any $x \in \mathbb{R}^n$, we have

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

Definition 2. Let f be a closed and convex function and $\mu > 0$. The Moreau envelope of f is

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

The parameter μ is called the smoothing parameter.

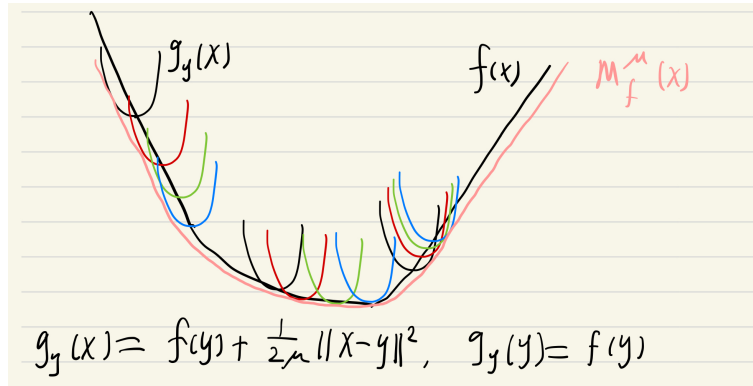


Figure 1: Moreau envelope

Properties

- $M_f^\mu(x) \leq f(x)$, plot, geometric interpretation: Moreau envelope M_f^μ is an envelope underneath f that smoothifies f but may not convexifies f

•

$$\nabla M_f^\mu(x) = \frac{1}{\mu} (x - \text{prox}_{\mu f}(x))$$

- ∇M_f^μ is $\frac{1}{\mu}$ -Lipschitz continuous, M_f^μ is $\frac{1}{\mu}$ -smooth

$$\begin{aligned}
 \|\nabla M_f^\mu(x) - \nabla M_f^\mu(y)\|^2 &= \frac{1}{\mu^2} \|x - \text{prox}_{\mu f}(x) - y + \text{prox}_{\mu f}(y)\|^2 \\
 &= \frac{1}{\mu^2} \left(\|x - y\|^2 + \|\text{prox}_{\mu f}(x) - \text{prox}_{\mu f}(y)\|^2 - 2\langle \text{prox}_{\mu f}(x) - \text{prox}_{\mu f}(y), x - y \rangle \right) \\
 &\leq \frac{1}{\mu^2} \left(\|x - y\|^2 - \|\text{prox}_{\mu f}(x) - \text{prox}_{\mu f}(y)\|^2 \right) \\
 &\leq \frac{1}{\mu^2} \|x - y\|^2
 \end{aligned}$$

- M_f^μ maintains convexity if f is convex. This is because partial minimization $g(x) = \min_y f(x, y)$ preserves convexity.

Noting that applying gradient descent on the Moreau envelope is equivalent to the proximal point method for f . Indeed,

$$x_{k+1} = x_k - \lambda \nabla M_f^\lambda(x_k) = x_k - (x_k - \text{prox}_{\lambda f}(x_k)) = \text{prox}_{\lambda f}(x_k).$$

Proximal Gradient Method-3

3 Proximal gradient method

3.1 Composite optimization

$$\min\{\phi(x) := f(x) + h(x)\}$$

- h is closed and convex;
- f is closed and convex, $\text{dom } f$ is convex, $\text{dom } h \subseteq \text{int}(\text{dom } f)$, and f is L -smooth over $\text{int}(\text{dom } f)$;
- the optimal set X_* is nonempty.

3.2 Proximal gradient

Algorithm 1 Proximal gradient method

Input: Initial point $x_0 \in \text{dom } h$, stepsize $\lambda > 0$

for $k \geq 0$ **do**

 Compute $x_{k+1} = \text{prox}_{\lambda h}(x_k - \lambda \nabla f(x_k))$.

end for

Theorem 5. *Functions f and h are as assumed in Subsection 3.1. Choose $\lambda \in (0, 1/L]$. Then, the proximal gradient method generates a sequence of points $\{x_k\}$ satisfying*

$$\phi(x_k) - \phi_* \leq \frac{\|x_0 - x_*\|^2}{2\lambda k}, \quad \forall k \geq 1.$$

Proof. It is easy to verify that one iteration of the proximal gradient method can be written as

$$x_{k+1} = \min_{x \in \mathbb{R}^n} \left\{ \ell_f(x; x_k) + h(x) + \frac{1}{2\lambda} \|x - x_k\|^2 \right\}.$$

Using Theorem 5 of Lecture 3 and the fact that the above objective function is $(1/\lambda)$ -strongly convex, we have for every $x \in \text{dom } h$,

$$\begin{aligned} \ell_f(x; x_k) + h(x) + \frac{1}{2\lambda} \|x - x_k\|^2 &\geq \ell_f(x_{k+1}; x_k) + h(x_{k+1}) + \frac{1}{2\lambda} \|x_{k+1} - x_k\|^2 + \frac{1}{2\lambda} \|x - x_{k+1}\|^2 \\ &\geq \ell_f(x_{k+1}; x_k) + h(x_{k+1}) + \frac{L}{2} \|x_{k+1} - x_k\|^2 + \frac{1}{2\lambda} \|x - x_{k+1}\|^2 \\ &\geq f(x_{k+1}) + h(x_{k+1}) + \frac{1}{2\lambda} \|x - x_{k+1}\|^2, \end{aligned}$$

where the second inequality is due to $\lambda \leq 1/L$ and the last inequality is due to Lemma 1(ii) of Lecture 3. It then follows from the convexity of f that

$$f(x) + h(x) + \frac{1}{2\lambda} \|x - x_k\|^2 \geq f(x_{k+1}) + h(x_{k+1}) + \frac{1}{2\lambda} \|x - x_{k+1}\|^2.$$

Taking $x = x_k$, we have

$$f(x_k) + h(x_k) \geq f(x_{k+1}) + h(x_{k+1}) + \frac{1}{2\lambda} \|x_{k+1} - x_*\|^2 \geq f(x_{k+1}) + h(x_{k+1})$$

which shows that the function value of the iterates is a nonincreasing sequence. Taking $x = x_*$, we have

$$f(x_*) + h(x_*) + \frac{1}{2\lambda} \|x_k - x_*\|^2 \geq f(x_{k+1}) + h(x_{k+1}) + \frac{1}{2\lambda} \|x_{k+1} - x_*\|^2,$$

i.e.,

$$(f + h)(x_{k+1}) - (f + h)(x_*) \leq \frac{1}{2\lambda} \|x_k - x_*\|^2 - \frac{1}{2\lambda} \|x_{k+1} - x_*\|^2.$$

Summing the above inequality and using the monotonicity of $\{(f + h)(x_k)\}$, we obtain

$$k[(f + h)(x_k) - (f + h)(x_*)] \leq \sum_{i=0}^{k-1} (f + h)(x_{i+1}) - (f + h)(x_*) \leq \frac{1}{2\lambda} \|x_0 - x_*\|^2 - \frac{1}{2\lambda} \|x_k - x_*\|^2.$$

Thus, the claim of the theorem follows. □