Notes for Lecture 8

Scribe: Tingkai Jia

1 Proximal Operator

1.1 ℓ_1 Norm Example

Example 1. *If* $h(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$, *then*

$$(\text{prox}_h(\mathbf{x}))_i = \psi_{st}(x_i; \lambda)$$

where

$$\psi_{st}(x; \lambda) = \begin{cases} x - \lambda, & \textit{if } x > \lambda \\ x + \lambda, & \textit{if } x < -\lambda \\ 0, & \textit{otherwise} \end{cases}$$

Proof. Consider the subgradient of l_1 norm function, let $\mathbf{g} \in \partial h(\mathbf{z})$, we can have

$$g_i \in \begin{cases} \lambda, & \text{if } z_i > 0 \\ -\lambda, & \text{if } z_i < 0 \\ [-\lambda, \lambda], & \text{if } z_i = 0 \end{cases}$$

Let $f(\mathbf{z}) = \frac{1}{2}\|\mathbf{z} - \mathbf{x}\|_2^2 + \lambda\|\mathbf{z}\|_1$, when $\nabla f(\mathbf{z}) = 0$, which also means $z_i - x_i + g_i = 0$, we have

$$(\text{prox}_h(\mathbf{x}))_i = z_i = \begin{cases} x_i - \lambda, & \text{if } x_i > \lambda \\ x_i + \lambda, & \text{if } x_i < -\lambda \\ 0, & \text{otherwise} \end{cases}$$

□

1.2 Basic Rules of Proximal Operator

Quadratic Addition: if $f(\mathbf{x}) = g(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x} - \mathbf{a}\|_2^2$, then

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

Proof. It follows that

$$\begin{aligned} \text{prox}_f(\mathbf{x}) &= \arg\min_{\mathbf{z}}\left\{\frac{1}{2}\|\mathbf{z} - \mathbf{x}\|_2^2 + g(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{z} - \mathbf{a}\|_2^2\right\} \\ &= \arg\min_{\mathbf{z}}\left\{\frac{1+\rho}{2}\|\mathbf{z}\|_2^2 - \langle \mathbf{z}, \mathbf{x} + \rho\mathbf{a}\rangle + g(\mathbf{z})\right\} \\ &= \arg\min_{\mathbf{z}}\left\{\frac{1}{2}\|\mathbf{z}\|_2^2 - \frac{1}{1+\rho}\langle \mathbf{z}, \mathbf{x} + \rho\mathbf{a}\rangle + \frac{1}{1+\rho}g(\mathbf{z})\right\} \\ &= \arg\min_{\mathbf{z}}\left\{\frac{1}{2}\left\|\mathbf{z} - \left(\frac{1}{1+\rho}\mathbf{x} + \frac{\rho}{1+\rho}\mathbf{a}\right)\right\|_2^2 + \frac{1}{1+\rho}g(\mathbf{z})\right\} \end{aligned}$$

$$= \text{prox}_{\frac{1}{1+\rho}g}\left(\frac{1}{1+\rho}\mathbf{x} + \frac{\rho}{1+\rho}\mathbf{a}\right).$$

□

Scaling and Translation: if $f(\mathbf{x}) = g(a\mathbf{x} + b)$ with $a \neq 0$, then

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

Proof. Let $\mathbf{w} = a\mathbf{z} + b$, it follows that

$$\begin{aligned}
\text{prox}_f(\mathbf{x}) &= \arg\min_{\mathbf{z}} \left\{ \frac{1}{2}\|\mathbf{z} - \mathbf{x}\|_2^2 + g(a\mathbf{z} + b) \right\} \\
&= \frac{1}{a}\left(\arg\min_{\mathbf{w}} \left\{ \frac{1}{2}\left\| \frac{\mathbf{w} - b}{a} - \mathbf{x} \right\|_2^2 + g(\mathbf{w}) \right\} - b \right) \\
&= \frac{1}{a}\left(\arg\min_{\mathbf{w}} \left\{ \frac{1}{2}\left\| \mathbf{w} - (b + a\mathbf{x}) \right\|_2^2 + a^2 g(\mathbf{w}) \right\} - b \right) \\
&= \frac{1}{a}\left(\text{prox}_{a^2 g}(a\mathbf{x} + b) - b \right).
\end{aligned}$$

□

norm composition: if$f(\mathbf{x}) = g(\|\mathbf{x}\|_2)$ with $\text{dom} g = [0, +\infty)$, then

$$\text{prox}_f(\mathbf{x}) = \text{prox}_g(\|\mathbf{x}\|_2)\frac{\mathbf{x}}{\|\mathbf{x}\|_2}, \quad \forall \mathbf{x} \neq 0$$

Proof. Note that

$$\begin{aligned}
\min_{\mathbf{z}} \left\{ \frac{1}{2}\|\mathbf{z} - \mathbf{x}\|_2^2 + g(\|\mathbf{z}\|_2) \right\} &= \min_{\mathbf{z}} \left\{ \frac{1}{2}\|\mathbf{z}\|_2^2 + \frac{1}{2}\|\mathbf{x}\|_2^2 - \langle \mathbf{z}, \mathbf{x} \rangle + g(\|\mathbf{z}\|_2) \right\} \\
&= \min_{\alpha \geq 0} \min_{\mathbf{z}:\|\mathbf{z}\|_2 = \alpha} \left\{ \frac{1}{2}\alpha^2 + \frac{1}{2}\|\mathbf{x}\|_2^2 - \langle \mathbf{z}, \mathbf{x} \rangle + g(\alpha) \right\} \\
&= \min_{\alpha \geq 0} \left\{ \frac{1}{2}\alpha^2 + \frac{1}{2}\|\mathbf{x}\|_2^2 - \alpha\|\mathbf{x}\|_2 + g(\alpha) \right\} \\
&= \min_{\alpha \geq 0} \left\{ \frac{1}{2}\left(\alpha - \|\mathbf{x}\|_2\right)^2 + g(\alpha) \right\},
\end{aligned}$$

From the above calculation, we know that the optimal point is

$$\alpha_* = \text{prox}_g(\|\mathbf{x}\|_2) \quad \text{and} \quad \mathbf{z}_* = \alpha_* \frac{\mathbf{x}}{\|\mathbf{x}\|_2} = \text{prox}_g(\|\mathbf{x}\|_2)\frac{\mathbf{x}}{\|\mathbf{x}\|_2}.$$

. □

1.3 Nonexpansiveness of Proximal Operator

Theorem 1 (firm nonexpansiveness)**.**

$$\langle \text{prox}_h(\mathbf{x}_1) - \text{prox}_h(\mathbf{x}_2), \mathbf{x}_1 - \mathbf{x}_2 \rangle \geq \|\text{prox}_h(\mathbf{x}_1) - \text{prox}_h(\mathbf{x}_2)\|_2^2$$

Proof. Let $\mathbf{z}_1 = \text{prox}_h(\mathbf{x}_1)$, $\mathbf{z}_2 = \text{prox}_h(\mathbf{x}_2)$. It is true that

$$\mathbf{x}_1 - \mathbf{z}_1 \in \partial h(\mathbf{z}_1) \quad \text{and} \quad \mathbf{x}_2 - \mathbf{z}_2 \in \partial h(\mathbf{z}_2).$$

Consider that h is convex, then we have

$$\begin{cases} h(\mathbf{z}_2) \geq h(\mathbf{z}_1) + \langle \mathbf{x}_1 - \mathbf{z}_1, \mathbf{x}_2 - \mathbf{z}_2 \rangle \\ h(\mathbf{z}_1) \geq h(\mathbf{z}_2) + \langle \mathbf{x}_2 - \mathbf{z}_2, \mathbf{x}_1 - \mathbf{z}_1 \rangle \end{cases}$$

add these two inequalities, we obtain

$$\langle \mathbf{x}_1 - \mathbf{x}_2, \mathbf{z}_1 - \mathbf{z}_2 \rangle \geq \|\mathbf{z}_1 - \mathbf{z}_2\|_2^2.$$

□

2 Moreau Decomposition

Theorem 2. *Suppose f is closed convex, and $f^*(\mathbf{x}) = \sup_{\mathbf{z}}\{\langle \mathbf{x}, \mathbf{z} \rangle - f(\mathbf{z})\}$ is convex conjugate of f. Then*

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

Example 2 (prox of support function). *For any closed and convex set $\mathcal{C}$, the support function $S_{\mathcal{C}}(\mathbf{x}) = \sup_{\mathbf{z} \in \mathcal{C}} \langle \mathbf{z}, \mathbf{x} \rangle$. Then*

$$\text{prox}_{S_{\mathcal{C}}}(\mathbf{x}) = \mathbf{x} - \mathcal{P}_{\mathcal{C}}(\mathbf{x})$$

Proof. First it is true that

$$S_{\mathcal{C}}^*(\mathbf{x}) = \mathbf{1}_{\mathcal{C}}(\mathbf{x}),$$

then the Moreau decomposition gives

$$\begin{aligned} \text{prox}_{S_{\mathcal{C}}}(\mathbf{x}) &= \mathbf{x} - \text{prox}_{S_{\mathcal{C}}^*}(\mathbf{x}) \\ &= \mathbf{x} - \text{prox}_{\mathbf{1}_{\mathcal{C}}}(\mathbf{x}) \\ &= \mathbf{x} - \mathcal{P}_{\mathcal{C}}(\mathbf{x}). \end{aligned} \tag{1}$$

□

Example 3 (ℓ_∞ norm)**.**

$$\text{prox}_{\|\cdot\|_\infty}(\mathbf{x}) = \mathbf{x} - \mathcal{P}_{\mathcal{B}_{\|\cdot\|_1}}(\mathbf{x}),$$

where $\mathcal{B}_{\|\cdot\|_1} := \{\mathbf{z} | \|\mathbf{z}\|_1 \leq 1\}$ is the unit l_1 ball.

Proof. Since $\|\mathbf{x}\|_\infty = \sup_{\mathbf{z}:\|\mathbf{z}\|_1 \leq 1} \langle \mathbf{z}, \mathbf{x} \rangle = S_{\mathcal{B}_{\|\cdot\|_1}}(\mathbf{x})$, we can use Example 2 to get

$$\text{prox}_{\|\cdot\|_\infty}(\mathbf{x}) = \text{prox}_{S_{\mathcal{B}_{\|\cdot\|_1}}}(\mathbf{x}) = \mathbf{x} - \mathcal{P}_{\mathcal{B}_{\|\cdot\|_1}}(\mathbf{x}).$$

□

Example 4 (max function)**.** *Let $g(\mathbf{x}) = \max\{x_1, x_2, \cdots, x_n\}$, then*

$$\text{prox}_g(\mathbf{x}) = \mathbf{x} - \mathcal{P}_{\triangle}(\mathbf{x})$$

where $\triangle := \{\mathbf{z} \in \mathbb{R}_+^n | \mathbf{1}^\top \mathbf{z} = 1\}$ is a probability simplex.

Proof. Since $\triangle := \{\mathbf{z} \in \mathbb{R}_+^n | \mathbf{1}^\top \mathbf{z} = 1\}$, we can use Example 2 to get

$$\text{prox}_g(\mathbf{x}) = \mathbf{x} - \mathcal{P}_{\triangle}(\mathbf{x}).$$

□