9 Proximal mapping

The proximal mapping is a "functional" generalization of the projection mapping. Given a convex function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$, the proximal mapping associated to f is

$$\mathbf{prox}_h(y) = \operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{ h(x) + \frac{1}{2}\|x-y\|_2^2 \right\}. \tag{1}$$

Clearly the proximal operator of the indicator function I_C of a closed convex set is precisely the projection operator.

The next proposition guarantees that $\mathbf{prox}_h$ is well-defined under mild conditions on f. A function h is *lower-semicontinuous* (lsc) if $h(x) \leq \liminf_{i\to\infty} f(x_i)$ for any sequence (x_i) converging to x.

EXERCISE: Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$. Prove that the following are equivalent: (i) h is lower-semicontinuous, (ii) $\mathbf{epi}(h)$ is closed, (iii) all the sublevel sets $h^{-1}((-\infty, a])$ are closed.

Proposition 9.1. *If h is lower-semicontinuous, then $\mathbf{prox}_h(y)$ is well-defined for all $y \in \mathbb{R}^n$.*

Proof. Let $g(x) = h(x) + (1/2)\|x-y\|_2^2$. Since g is strongly convex, any minimizer is necessarily unique. It remains to show that a minimizer exists. First note that g is bounded below: since h is convex it can be lower bounded by an affine function $h(x) \geq \langle a, x\rangle + b$, and so $g(x) \geq \langle a, x\rangle + b + (1/2)\|x-y\|_2^2 \geq \min_{x\in\mathbb{R}^n}\{\langle a, x\rangle + b + (1/2)\|x-y\|_2^2\} = c > -\infty$. Also note that the sublevel sets of g are all bounded since $g(x) \leq t \implies \langle a, x\rangle + b + (1/2)\|x-y\|_2^2 \leq t \iff \|x-(y-a)\|_2^2 \leq C$ for some constant $C > 0$. Now let (x_i) be a sequence so that $g(x_i) \downarrow \inf_{x\in\mathbb{R}^n} g(x)$. The sequence (x_i) lives in the sublevel set $\{x : g(x) \leq g(x_1)\}$ which is closed and bounded. Thus we can extract from (x_i) a converging subsequence, that converges to some x. Since g is lower semicontinuous we have $g(x) \leq \liminf_i g(x_i) = \inf g$, and so x is a minimizer of g. □

Note that

$$x = \mathbf{prox}_h(y) \iff 0 \in \partial h(x) + (x-y) \iff y \in x + \partial h(x). \tag{2}$$

Remark 1. *If h is smooth, we see that $x = \mathbf{prox}_h(y)$ is a solution to the nonlinear equation $x + \nabla h(x) = y$, i.e., it satisfies $x = (I + \nabla h)^{-1}(y)$.*

Just like with the projection, one can prove that the proximal map is nonexpansive, i.e., that

$$\|\mathbf{prox}_h(y_1) - \mathbf{prox}_h(y_2)\|_2 \leq \|y_1 - y_2\|_2.$$

To see why, let $x_1 = \mathbf{prox}_h(y_1)$ and $x_2 = \mathbf{prox}_h(y_2)$. Then $y_1 - x_1 \in \partial h(x_1)$, and so we can write:

$$h(x_2) \geq h(x_1) + \langle y_1 - x_1, x_2 - x_1\rangle .$$

Similarly, from $y_2 - x_2 \in \partial h(x_2)$, we get

$$h(x_1) \geq h(x_2) + \langle y_2 - x_2, x_1 - x_2\rangle .$$

Summing the two inequalities, we get $0 \geq \langle x_1 - y_1 + y_2 - x_2, x_1 - x_2\rangle$ which corresponds to

$$\|x_1 - x_2\|_2^2 \leq \langle y_1 - y_2, x_1 - x_2\rangle \tag{3}$$

and which, by Cauchy-Schwarz implies $\|x_1 - x_2\|_2 \leq \|y_1 - y_2\|_2$ as desired.

Example Let $h(x) = |x|$ defined on $\mathbb{R}$. Then one can verify (exercise!) that for any $t > 0$,

$$\mathbf{prox}_{th}(y) = \operatorname*{argmin}_{x\in\mathbb{R}} \left\{|x| + 1/(2t)(x-y)^2\right\} = S_t(y) := \begin{cases} y+t & \text{if } y \leq -t \\ 0 & \text{if } |y| < t \\ y-t & \text{if } y \geq t. \end{cases} \tag{4}$$

This function is known as *soft-thresholding*. See Figure 1.

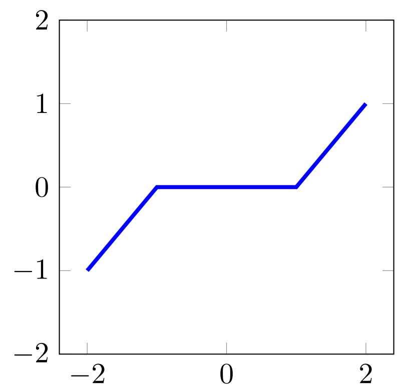

Figure 1: The soft-thresholding function (4) for $t = 1$.

Observe that if $h(x) = \sum_{i=1}^n h_i(x_i)$, then the **prox** of h decomposes:

$$(\mathbf{prox}_h(y))_i = \mathbf{prox}_{h_i}(y_i).$$

This implies for example that the prox operator of the ℓ_1 norm function is a componentwise soft-thresholding:

$$\mathbf{prox}_{t\|\cdot\|_1}(y) = [S_t(y_i)]_{1\leq i\leq n}$$

EXERCISE: Compute the proximal operators for the following functions: (i) $h(x) = (1/2)x^T Ax$ where A is symmetric positive definite; (ii) $h(x) = -\sum_{i=1}^n \log x_i$ for $x \in \mathbb{R}^n_{++}$.

References

[BT09] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009. 5

[PB14] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.