4 Proximity Operators

We will refer to $\mathcal{H}$ as a "finite dimensional real **Hilbert space**." When we say "Hilbert space," we are referring to both a vector space and an inner product $(\mathcal{H}, \langle\cdot \mid \cdot\rangle)$, which also gives rise to a norm via $\|\cdot\| = \sqrt{\langle\cdot \mid \cdot\rangle}$ (see Exercise 1.12).

Definition 4.1 Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$. Then f is **strictly convex** if, for *distinct* points $x, y \in \mathcal{H}$ $(x \neq y)$, we have

$$(\forall \alpha \in \left]0,1\right[) \quad f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y). \tag{49}$$

Exercise 4.2 Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be strictly convex and suppose that at least one minimizer exists. Prove that the minimizer is unique.

Proposition 4.3 *Let $f \in \Gamma_0(\mathcal{H})$ and let $x \in \mathcal{H}$. The following optimization problem has a unique solution*

$$\underset{z \in \mathcal{H}}{\text{minimize}} \;\; f(z) + \frac{1}{2}\|x - z\|^2 \tag{$*$}$$

The proof idea of Proposition 4.3 comes from the fact that the norm is strictly convex, which implies that the objective function in $(*)$ is convex.

Definition 4.4 Let $f \in \Gamma_0(\mathcal{H})$ and let $x \in \mathcal{H}$. The solution to $(*)$ is the **proximal point** of f at x, denoted $\text{Prox}_f(x) \in \mathcal{H}$. This defines an operator $\text{Prox}_f\colon \mathcal{H} \to \mathcal{H}$ where x maps to $\text{Prox}_f(x)$.

Example 4.5 Let C be a nonempty closed convex subset of $\mathcal{H}$. Then, for $x \in \mathcal{H}$, we have

$$\min_{z \in \mathcal{H}} \iota_C(z) + \frac{1}{2}\|x - z\|^2 = \min_{z \in C} \frac{1}{2}\|x - z\|^2. \tag{50}$$

However, the minimizer of (50) will be the same as in (24). Therefore, the proximity operator of ι_C is actually just the projection operator!

$$\text{Prox}_{\iota_C} = \text{Proj}_C. \tag{51}$$

Exercise 4.6 Let $f \in \Gamma_0(\mathcal{H})$ and let $x, p \in \mathcal{H}$. Show that

$$p = \text{Prox}_f x \qquad \Leftrightarrow \qquad x - p \in \partial f(p). \tag{52}$$

Definition 4.7 Let $\mathcal{H}_1, \ldots, \mathcal{H}_m$ be real Hilbert spaces with inner products denoted $\langle\cdot \mid \cdot\rangle_{\mathcal{H}_1}, \ldots, \langle\cdot \mid \cdot\rangle_{\mathcal{H}_m}$. We can construct another Hilbert space, called the **direct sum** (sometimes also called a *product space*) as follows. Our vector space is denoted $\bigoplus_{i=1}^m \mathcal{H}_i = \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$, and the inner product on the direct sum is given by

$$\langle\cdot \mid \cdot\rangle_{\bigoplus_{i=1}^m \mathcal{H}_i}\colon \bigoplus_{i=1}^m \mathcal{H}_i \times \bigoplus_{i=1}^m \mathcal{H}_i \to \mathbb{R}\colon ((x_i)_{i=1}^m, (y_i)_{i=1}^m) \mapsto \sum_{i=1}^m \langle x_i \mid y_i\rangle_{\mathcal{H}_i}. \tag{53}$$

If, for every $i \in I$, $f_i \colon \mathcal{H}_i \to \left]-\infty, +\infty\right]$, then the function

$$f\colon \bigoplus_{i=1}^{m} \mathcal{H}_i \to \left]-\infty, +\infty\right] : (x_i)_{i=1}^{m} \mapsto \sum_{i=1}^{m} f_i(x_i) \tag{54}$$

is called a **separable sum**.

Proposition 4.8 (Proximity Operator of a Separable Sum) *Let $\mathcal{H}_1, \ldots, \mathcal{H}_m$ be real Hilbert spaces, and for every $i \in \{1, \ldots, m\}$, let $f_i \in \Gamma_0(\mathcal{H}_i)$. Let f be the separable sum of $(f_i)_{i=1}^{m}$ in the form (54). Then*

$$\left(\forall (x_i)_{i=1}^{m} \in \bigoplus_{i=1}^{m} \mathcal{H}_i\right) \quad \mathrm{Prox}_{f}\left((x_i)_{i=1}^{m}\right) = \left(\mathrm{Prox}_{f_1}(x_1), \mathrm{Prox}_{f_2}(x_2), \ldots, \mathrm{Prox}_{f_m}(x_m)\right). \tag{55}$$

Proof. For notational convenience, set $\mathcal{H} = \bigoplus_{i=1}^{m} \mathcal{H}_i$. Let $x = (x_i)_{i=1}^{m} \in \mathcal{H}$. Then

$$\min_{z \in \mathcal{H}} f(z) + \frac{1}{2}\|x - z\|_{\mathcal{H}}^2 = \min_{\substack{z_1 \in \mathcal{H}_1 \\ \vdots \\ z_m \in \mathcal{H}_m}} \sum_{i=1}^{m} f_i(z_i) + \frac{1}{2}\|x_i - z_i\|_{\mathcal{H}_i}^2 \tag{56}$$

$$= \sum_{i=1}^{m} \min_{z_i \in \mathcal{H}_i} f_i(z_i) + \frac{1}{2}\|x_i - z_i\|_{\mathcal{H}_i}^2. \tag{57}$$

Since all of the summands in (56) are only concerned with a single optimization variable z_i (as opposed to the entire vector $z \in \mathcal{H}$), we actually have a sum of independent optimization problems! This observation permits us to commute the sum and the minimization between (56) and (57). Finally, we see that the solution in each of the subproblems in (57) is precisely the proximity operator of f_i at x_i. Since the solution in each component is $\mathrm{Prox}_{f_i}(x_i)$, we have are done since we have just shown (55). □

Example 4.9 Let $\gamma > 0$. Let's compute the proximity operator of my favorite nonsmooth function $\gamma|\cdot|$. Let $x, p \in \mathbb{R}$. Combiniing Example 3.12 and the characterization in Exercise 4.6, we find

$$p = \mathrm{Prox}_{\gamma|\cdot|} x \Leftrightarrow x - p \in \partial(\gamma|\cdot|)(p) \tag{58}$$

$$\Leftrightarrow x - p \in \partial(|\cdot|)(p) = \begin{cases} -\gamma & \text{if } p < 0 \\ [-\gamma, \gamma] & \text{if } p = 0 \\ \gamma & \text{if } p > 0. \end{cases} \tag{59}$$

Now, (59) looks a bit funny, since we don't know what p is to begin with. However, if we cobble together the case analysis, we will be able to formally invert this inclusion. Let's consider the first case: If $p < 0$, then

$$x - p \in \{-\gamma\} \quad \Leftrightarrow \quad p = x + \gamma, \tag{60}$$

which means $p < 0$ if and only if $x < -\gamma$. Note we have translated a condition on p to a condition on x. Similarly, for $p > 0$

$$x - p \in \{\gamma\} \quad \Leftrightarrow \quad p = x - \gamma, \tag{61}$$

and $p > 0$ if and only if $x > \gamma$. Finally, if $p = 0$

$$x - p \in [-\gamma, \gamma] \quad \Leftrightarrow \quad x \in [-\gamma, \gamma]. \tag{62}$$

Note that for each of the three cases, we were able to translate the condition on p in (59) to a condition on x! Combining these three observations, we find that

$$p = \operatorname{Prox}_{\gamma|\cdot|}(x) \begin{cases} x + \gamma & \text{if } \ x < -\gamma \\ 0 & \text{if } \ -\gamma \leq x \leq \gamma \\ x - \gamma & \text{if } \ \gamma \leq x. \end{cases} \tag{63}$$

This operation in (63) is known as the **soft thresholder**.

Exercise 4.10 Compute the proximity operator of the **one norm**

$$\|\cdot\|_1 \colon \mathbb{R}^n \to \mathbb{R} \colon (x_i)_{i=1}^m \mapsto \sum_{i=1}^m |x_i| \tag{64}$$

hint: Proposition 4.8

Remark 4.11 Collectively, humans know how to compute lot of proximity operators. Two libraries I have used are below.

(i) `proximity-operator.net`

(ii) ProximalOperators.jl (on Github)

Definition 4.12 Let $T \colon \mathcal{H} \to \mathcal{H}$ be an operator and let $x \in \mathcal{H}$. If $Tx = x$, x is a **fixed point** of T.

Proposition 4.13 *Let $f \in \Gamma_0(\mathcal{H})$ and let $x \in \mathcal{H}$. Then*

$$\operatorname{Prox}_f(x) = x \quad \Leftrightarrow \quad x \textit{ minimizes } f. \tag{65}$$

Proof. First, note that $\operatorname{dom}\left(\frac{1}{2}\|\cdot - x\|^2\right) = \mathcal{H}$, so by the Sum Rule (Theorem 3.14(ii)), we know that

$$\partial\left(f + \frac{1}{2}\|\cdot - x\|^2\right) = \partial f + \partial\left(\frac{1}{2}\|\cdot - x\|^2\right).$$

In a matter which proceeds similarly to Exercise 3.5, we observe from Remark 3.15 that, for every $z \in \mathcal{H}$, $\partial\left(\frac{1}{2}\|\cdot - x\|^2\right)(z) = \{z - x\}$. Combining these facts, along with using Fermat's rule (Theorem 3.13) twice, we find

$$\begin{aligned}
\operatorname{Prox}_f(x) = x &\Leftrightarrow x \text{ minimizes } f(\cdot) + \frac{1}{2}\|\cdot - x\|^2 && (66)\\
&\Leftrightarrow 0 \in \partial\left(f(\cdot) + \frac{1}{2}\|\cdot - x\|^2\right)(x) && (67)\\
&\Leftrightarrow 0 \in \partial f(x) + \partial\left(\frac{1}{2}\|\cdot - x\|^2\right) && (68)\\
&\Leftrightarrow 0 \in \partial f(x) + \{x - x\} && (69)\\
&\Leftrightarrow x \text{ minimizes } f. && (70)
\end{aligned}$$

□

Proposition 4.13 tells us that the fixed-points of a proximity operator are precisely the minimizers of our function. Spoiler alert: this motivates a fixed-point algorithm! It turns out that, if we just repeatedly apply the proximity operator, we will converge to a solution of our problem. This will be detailed in Theorem 4.15.

There are a **lot** of awesome properties of the proximity operator. However, in the interest of time (and practicability), we now shift focus towards algorithms.

4.1 The Proximal Point Algorithm

In optimization, there are a variety of methods to quantify "convergence" of an algorithm. One method – *primal convergence* shows that the function value of the objective (1) approaches the optimal value, i.e., $f(x_n) - \inf_{x \in C} f(x) \to 0$. However, this could be misleading.

Exercise 4.14 *(extra credit)* Construct a convex function f with at least one minimizer and a sequence $(x_n)_{n\in\mathbb{N}}$ such that $f(x_n) - \inf_{x\in\mathcal{H}} f(x) \to 0$, and for every minimizer z, $\|x_n - z\| \not\to 0$.

There are solutions to Exercise 4.14 (e.g., in Bauschke/Combettes' book mentioned in the introduction) which demonstrate that, no matter how low the objective value can be, our iterates could still be arbitrarily far away from the actual minimizers. As a result, some like to instead show that that, for a minimizer $z \in C$, the sequence of iterates actually approaches the minimizer, i.e., $\|x_n - z\| \to 0$. For the proximal point algorithm, we have both.

Theorem 4.15 (Proximal Point Algorithm (Martinet, 1970)) *Let $f \in \Gamma_0(\mathcal{H})$ have at least one minimizer. Let $\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$, and let $x_0 \in \mathcal{H}$. Set*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{Prox}_{\gamma_n f}(x_n). \tag{71}$$

Then the following hold.

(i) $f(x_n)_{n\in\mathbb{N}}$ *converges monotonically to* $\min f(\mathcal{H})$.

(ii) $(x_n)_{n\in\mathbb{N}}$ *converges to a minimizer of* f.

Remark 4.16 (Comment on convergence proofs for prox-based algorithms) The proximal point algorithm (and many of its relatives) can be proven to converge using the following template[4].

(i) Show that the algorithm is **Fejér monotone**, i.e., for a solution to your problem $z \in \mathcal{H}$, we have

$$\|x_{n+1} - z\| \leq \|x_n - z\| \tag{72}$$

Oftentimes, the properties of $(T_n)_{n\in\mathbb{N}}$ actually reveal a strictly negative term being added to $\|x_n - z\|$ on the upper bound in (72). This can sometimes be used to obtain specific rates of convergence.

(ii) Assume that a cluster point of the algorithm exists. Show that it is a solution of our problem.

Using existing theory about Fejér monotonicity, we can conclude $(x_n)_{n\in\mathbb{N}}$ converges to a solution of our problem. Note that we never actually had to prove the algorithm converges.

Remark 4.17 (A retrospective on Theorem 4.15) The Proximal Point Algorithm tells us that, as long as we can compute the proximity operator of our objective function, we can declare victory. That's it, right? Well, it turns out that the story is not quite so simple. In practice, one can typically compute the proximity operator of each summand in an optimization problem. However, computing the prox of their sum is much trickier. Tune in next time, for an introduction to *splitting algorithms*!

[4] unfortunately we do not have time to detail the full convergence proofs