6 Splitting methods

When a sum of several functions is minimized, some of them may be easier to minimize than others. In addition, optimization objects may be difficult to compute for sums of functions compared to individual ones. In this chapter, we cover how to split optimization objects into atomic ones.

Let us consider problems of the form

$$\min_{x\in\mathbb{R}^n} g(x) + h(x).$$

where g and h are two convex functions, that are potentially nonsmooth.

6.1 The Proximal Gradient

A first simple case is when one of the functions is smooth. In this case, the problem writes as

$$\min_{x\in\mathbb{R}^n} f(x) + g(x). \tag{1}$$

where f is smooth and convex, while g is only convex proper and lower semi-continuous.

Since the proximity operator is difficult to compute in general, a rule of thumb is to use a gradient method as soon as possible. Furthermore, in many signal processing or machine learning problems, the objective is of the form $f + g$, with f a smooth loss function that measure the fit between the model and the data and g a nonsmooth regularization, chosen so that the proximity operator is easy to compute.

6.1.1 Algorithm

The *proximal gradient* algorithm consists in iterating

$$x_{k+1} = \mathbf{prox}_{\gamma g}\left(x_k - \gamma\nabla f(x_k)\right) \tag{Proximal gradient}$$

for some $\gamma > 0$ and starting point x_0.

It is worth noticing that this composition can actually be seen as the minimization of a first-order approximation of f plus g. Indeed:

$$\begin{aligned}
x_{k+1} &= \mathbf{prox}_{\gamma g}\left(x_k - \gamma\nabla f(x_k)\right)\\
&= \operatorname{argmin}_{u\in\mathbb{R}^n}\left\{g(u) + \frac{1}{2\gamma}\|u - x_k + \gamma\nabla f(x_k)\|^2\right\}\\
&= \operatorname{argmin}_{u\in\mathbb{R}^n}\left\{g(u) + \langle u - x_k, \nabla f(x_k)\rangle + \frac{1}{2\gamma}\|u - x_k\|^2 + \frac{\gamma}{2}\|\nabla f(x_k)\|^2\right\}\\
&= \operatorname{argmin}_{u\in\mathbb{R}^n}\left\{g(u) + f(x_k) + \langle u - x_k, \nabla f(x_k)\rangle + \frac{1}{2\gamma}\|u - x_k\|^2\right\} \qquad (2)
\end{aligned}$$

where in the last inequality we remove terms independent of u. We notice that the first term is g while the next two terms approximate f.

This helps us put together the tools for the algorithm's descent lemma.

Lemma 1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex L-smooth function and let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex lower semi-continuous proper function. Then, the Proximal gradient method with $\gamma \in (0, 1/L]$ verifies $f(x_{k+1})+g(x_{k+1}) \leq f(x_k)+g(x_k)$ and for any minimizer $x^\star$*

$$f(x_k) + g(x_k) - (f(x^\star) + g(x^\star)) \leq \frac{\|x^\star - x_0\|^2}{2\gamma k}.$$

Proof. By (2), x_{k+1} is the minimizer of the right hand side, which is a $1/\gamma$-strongly convex function, hence for any $z \in \mathbb{R}^n$,

$$\begin{aligned}
& f(x_k) + \langle x_{k+1} - x_k, \nabla f(x_k)\rangle + g(x_{k+1}) + \frac{1}{2\gamma}\|x_{k+1} - x_k\|^2 \\
\leq & f(x_k) + \langle z - x_k, \nabla f(x_k)\rangle + g(z) + \frac{1}{2\gamma}\|z - x_k\|^2 - \frac{1}{2\gamma}\|z - x_{k+1}\|^2 \\
\leq & f(z) + g(z) + \frac{1}{2\gamma}\|z - x_k\|^2 - \frac{1}{2\gamma}\|z - x_{k+1}\|^2
\end{aligned}$$

where the second inequality comes from the convexity of f.

Now, the smoothness of f, implies that

$$f(x_{k+1}) \leq f(x_k) + \langle x_{k+1} - x_k, \nabla f(x_k)\rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$$

and using the first set of inequalities, we get

$$\begin{aligned}
& f(x_{k+1}) + g(x_{k+1}) \\
& \leq f(x_k) + \langle x_{k+1} - x_k, \nabla f(x_k)\rangle + g(x_{k+1}) + \frac{1}{2\gamma}\|x_{k+1} - x_k\|^2 + \frac{1}{2}\left(L - \frac{1}{\gamma}\right)\|x_{k+1} - x_k\|^2 \\
& \leq f(z) + g(z) + \frac{1}{2\gamma}\|z - x_k\|^2 - \frac{1}{2\gamma}\|z - x_{k+1}\|^2 + \frac{1}{2}\left(L - \frac{1}{\gamma}\right)\|x_{k+1} - x_k\|^2.
\end{aligned}$$

Using $z = x_k$, we get that the sequence of functional values is decreasing and with $z = x^\star$, we obtain the rate with the same proof as for the proximal point method. □

6.1.2 A first splitting method

An alternative (and mode general) construction can be provided for the Proximal gradient, closely linked to *splitting methods.*

First, we notice that a minimizer of problem (1) is a point x satisfying

$$0 \in \nabla f(x) + \partial g(x). \tag{3}$$

Second, we recall that:

- a gradient step on f leads to $u = x - \gamma\nabla f(x)$
- a proximal step on g generates y such that $y + \gamma\partial g(y) \ni z$

which are our two ingredients.

Now, it suffices to notice that

$$\begin{aligned}
& 0 \in \nabla f(x) + \partial g(x) \\
\Leftrightarrow & 0 \in \gamma\nabla f(x) + \gamma\partial g(x) \\
\Leftrightarrow & 0 \in -(x - \gamma\nabla f(x)) + (x + \gamma\partial g(x)) \\
\Leftrightarrow & \begin{cases} u & = x - \gamma\nabla f(x) \\ x + \gamma\partial g(x) & \ni u \end{cases}
\end{aligned} \tag{4}$$

where we have split the two functions and thus decoupled the difficult problem of finding x verifying (3). Thus, we need to iteratively find two points x, u verifying (4).

Here, the *order* of the iterates will be important: if we are given x_k, we can compute u_k (by a gradient step), and then a new value x_{k+1} (by a proximal step). We obtain

$$\begin{aligned}
& \begin{cases} u_k & = x_k - \gamma\nabla f(x_k) \\ x_{k+1} + \gamma\partial g(x_{k+1}) & \ni u_k \end{cases} \\
\Leftrightarrow & \begin{cases} u_k & = x_k - \gamma\nabla f(x_k) \\ x_{k+1} & = \mathbf{prox}_{\gamma g}(u_k) \end{cases} \\
\Leftrightarrow & x_{k+1} = \mathbf{prox}_{\gamma g}(x_k - \gamma\nabla f(x_k))
\end{aligned}$$

which is indeed our Proximal gradient.

6.2 Splitting of two nonsmooth functions

We get back to our original problem

$$\min_{x \in \mathbb{R}^n} g(x) + h(x).$$

where g and h are two convex functions, that are potentially nonsmooth.

A splitting method will find a point x such that $0 \in \partial g(x) + \partial h(x)$ and one can thus wonder when such a point is a minimizer of $g + h$. Fortunately, this is the case under rather mild assumptions.

Lemma 2. *Let g and h be two proper lower semi-continuous convex functions. If (i)* $\operatorname{dom} g \cap \operatorname{int} \operatorname{dom} h \neq \emptyset$ *or (ii)* $\operatorname{ri} \operatorname{dom} g \cap \operatorname{ri} \operatorname{dom} h \neq \emptyset$*, then*

$$\partial(g + h) = \partial g + \partial h.$$

Hence, in this section, we will build iterative methods that solve

$$0 \in \partial g(x) + \partial h(x). \tag{5}$$

6.2.1 Construction

Since both functions are nonsmooth, our only ingredient is the proximal step:

- a proximal step on g generates x such that $x + \gamma \partial g(x) \ni u$;
- a proximal step on h generates y such that $y + \gamma \partial h(y) \ni z$.

$$\begin{aligned}
& 0 \in \partial g(x) + \partial h(x) \\
\Leftrightarrow & 0 \in \gamma \partial g(x) + \gamma \partial h(x) \\
\Leftrightarrow & \begin{cases} 0 & \in \gamma \partial g(x) + \gamma \partial h(y) \\ x & = y \end{cases} \\
\Leftrightarrow & \begin{cases} u - x & \in \gamma \partial g(x) \\ y - u & \in \gamma \partial h(y) \\ x & = y \end{cases} \\
\Leftrightarrow & \begin{cases} u & \in x + \gamma \partial g(x) \\ 2y - u & \in y + \gamma \partial h(y) \\ x & = y \end{cases} \\
\Leftrightarrow & \begin{cases} x & = \mathbf{prox}_{\gamma g}(u) \\ y & = \mathbf{prox}_{\gamma h}(2y - u) \\ x & = y \end{cases} \\
\Leftrightarrow & \begin{cases} x & = \mathbf{prox}_{\gamma g}(u) \\ y & = \mathbf{prox}_{\gamma h}(2x - u) \\ x & = y \end{cases} \\
\Leftrightarrow & \begin{cases} x & = \mathbf{prox}_{\gamma g}(u) \\ y & = \mathbf{prox}_{\gamma h}(2x - u) \\ u & = u + (y - x) \end{cases}
\end{aligned} \tag{6}$$

where the guiding principle was to make the proximity operators appear and then make the iterations computable in order. For the last one, since u is not updated and we have no mean to enforce $x = y$ at that time, we add u on both sides (we could also add an additional parameter there).

This way, starting from top right to bottom left, we obtain the following method, called Douglas-Rachford splitting:

$$\begin{cases} x_k & = \mathbf{prox}_{\gamma g}(u_k) \\ y_k & = \mathbf{prox}_{\gamma h}(2x_k - u_k) \\ u_{k+1} & = u_k + (y_k - x_k) \end{cases}$$

which can also be seen as

$$\begin{aligned} u_{k+1} &= u_k + (\mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u_k) - u_k) - \mathbf{prox}_{\gamma g}(u_k)) \\ &:= \mathsf{T}(u_k). \end{aligned}$$

It is thus interesting to look at the properties of the operator T.

6.2.2 Properties of the operator T

The operator T is a $\mathbb{R}^n \to \mathbb{R}^n$ mapping defined for all $u \in \mathbb{R}^n$ as

$$\mathsf{T}(u) \coloneqq u + (\mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma g}(u)).$$

The goal of this section is to show that T has virtually the same *contraction properties* as a proximity operator. For this recall that for any convex lower semi-continuous proper function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$, the following propositions are equivalent:

i) $x = \mathbf{prox}_{\gamma g}(y)$;

ii) $(y - x)/\gamma \in \partial g(x)$;

iii) $g(u) \geq g(x) + \langle (y - x)/\gamma, u - x\rangle$ for any $u \in \mathbb{R}^n$.

This enables us to show the following lemma, which highlights the proximity operator is *firmly non-expansive.*

Lemma 3. *Let $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a convex lower semi-continuous proper function, then for any $u, z \in \mathbb{R}^n$, $\gamma > 0$,*

$$\begin{aligned} &\|\mathbf{prox}_{\gamma g}(u) - \mathbf{prox}_{\gamma g}(z)\|^2 \leq \langle u - z, \mathbf{prox}_{\gamma g}(u) - \mathbf{prox}_{\gamma g}(z)\rangle \\ \Leftrightarrow &\|\mathbf{prox}_{\gamma g}(u) - \mathbf{prox}_{\gamma g}(z)\|^2 \leq \|u - z\|^2 - \|u - \mathbf{prox}_{\gamma g}(u) - z + \mathbf{prox}_{\gamma g}(z)\|^2 \end{aligned}$$

Proof. Take point (iii) above with $y \leftarrow u, u \leftarrow \mathbf{prox}_{\gamma g}(z)$ and $y \leftarrow z, u \leftarrow \mathbf{prox}_{\gamma g}(u)$. Summing both inequalities gives the first result. The second one comes directly afterwards. □

Now, we can prove the same result for the Douglas-Rachford operator T.

Theorem 4. *Let $g, h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be two convex lower semi-continuous proper functions, then for any $u, z \in \mathbb{R}^n$, $\gamma > 0$,*

$$\begin{aligned} &\|\mathsf{T}(u) - \mathsf{T}(z)\|^2 \leq \langle u - z, \mathsf{T}(u) - \mathsf{T}(z)\rangle \\ \Leftrightarrow &\|\mathsf{T}(u) - \mathsf{T}(z)\|^2 \leq \|u - z\|^2 - \|u - \mathsf{T}(u) - z + \mathsf{T}(z)\|^2 \end{aligned}$$

Proof. $(\star)$ Let $u, z \in \mathbb{R}^n$,

$$
\begin{aligned}
&\langle \mathsf{T}(u) - \mathsf{T}(z), u - z\rangle \\
&= \langle u - z, u - z\rangle \\
&\quad + \langle \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(z) - z), u - z\rangle \\
&\quad - \langle \mathbf{prox}_{\gamma g}(u) - \mathbf{prox}_{\gamma g}(z), u - z\rangle \\
&\geq \|u - z\|^2 - 2\langle \mathbf{prox}_{\gamma g}(u) - \mathbf{prox}_{\gamma g}(z), u - z\rangle + \|\mathbf{prox}_{\gamma g}(u) - \mathbf{prox}_{\gamma g}(z)\|^2 \\
&\quad + \langle \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(z) - z), u - z\rangle \\
&= \|u - \mathbf{prox}_{\gamma g}(u) - (z - \mathbf{prox}_{\gamma g}(z))\|^2 \\
&\quad + \langle \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(z) - z), u - z\rangle \\
&= \|u - \mathbf{prox}_{\gamma g}(u) - (z - \mathbf{prox}_{\gamma g}(z))\|^2 \\
&\quad - \langle \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(z) - z), 2\mathbf{prox}_{\gamma g}(u) - u - (2\mathbf{prox}_{\gamma g}(z) - z)\rangle \\
&\quad + \langle \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(z) - z), 2\mathbf{prox}_{\gamma g}(u) - 2\mathbf{prox}_{\gamma g}(z)\rangle \\
&\geq \|u - \mathbf{prox}_{\gamma g}(u) - (z - \mathbf{prox}_{\gamma g}(z))\|^2 \\
&\quad - 2\langle \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(z) - z), 2\mathbf{prox}_{\gamma g}(u) - u - (2\mathbf{prox}_{\gamma g}(z) - z)\rangle \\
&\quad + \|\mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(z) - z)\|^2 \\
&\quad + 2\langle \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma h}(\mathbf{prox}_{\gamma g}(z) - z), \mathbf{prox}_{\gamma g}(u) - \mathbf{prox}_{\gamma g}(z)\rangle \\
&= \|u - \mathbf{prox}_{\gamma g}(u) - (z - \mathbf{prox}_{\gamma g}(z))\|^2 \\
&\quad + 2\langle \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma h}(\mathbf{prox}_{\gamma g}(z) - z), u - \mathbf{prox}_{\gamma g}(u) - (z - \mathbf{prox}_{\gamma g}(z))\rangle \\
&\quad + \|\mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(z) - z)\|^2 \\
&= \|u + \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(u) - u) - \mathbf{prox}_{\gamma g}(u) - (z + \mathbf{prox}_{\gamma h}(2\mathbf{prox}_{\gamma g}(z) - \mathbf{prox}_{\gamma g}(z))\|^2 \\
&= \|\mathsf{T}(u) - \mathsf{T}(z)\|^2
\end{aligned}
$$

where the inequalities come from Lemma 3. □

6.2.3 Convergence

Now that we have highlighted our base property, let us show the convergence of our method. Note that this kind of proof is ubiquitous in optimization with *monotone operators*.

Theorem 5. *Let $g, h : \mathbb{R}^n \to \overline{\mathbb{R}}$ be two convex lower semi-continuous proper functions such that:*

- *(i)* $\operatorname{dom} g \cap \operatorname{int} \operatorname{dom} h \neq \emptyset$ *or (ii)* $\operatorname{ri} \operatorname{dom} g \cap \operatorname{ri} \operatorname{dom} h \neq \emptyset$ *;*
- *$g + h$ has a minimizer.*

Then, the Douglas-Rachford algorithm

$$
\begin{cases}
x_k &= \mathbf{prox}_{\gamma g}(u_k) \\
y_k &= \mathbf{prox}_{\gamma h}(2x_k - u_k) \\
u_{k+1} &= u_k + (y_k - x_k)
\end{cases}
$$

produces a sequence (x_k) that converges to a minimizer of $g + h$.

Proof. The two assumptions in the theorem imply the existence of minimizers that are the points verifying $0 \in \partial g(x) + \partial h(x)$ (see Eq. (5)).

Let $x^\star$ be a point such that $0 \in \partial g(x^\star) + \partial h(x^\star)$. Then, by Eq. (6), there is a $u^\star$ such that:

- $u^\star = \mathsf{T}(u^\star)$, i.e., $u^\star$ is a fixed point of T;
- $x^\star = \mathbf{prox}_{\gamma g}(u^\star)$.

Then, from Theorem 4, we get that

$$\|u_{k+1} - u^\star\|^2 = \|\mathsf{T}(u_k) - \mathsf{T}(u^\star)\|^2 \leq \|u_k - u^\star\|^2 - \|\mathsf{T}(u_k) - u_k\|^2$$

which means that $(\|u_k - u^\star\|^2)$ is non-increasing and thus convergent.

Also, since $\|u_k - u^\star\|^2 \leq \|u_0 - u^\star\|^2$, (u_k) is bounded and thus has a converging subsequence. Furthermore, the limit u of this subsequence must verify $\mathsf{T}(u) - u = 0$ (i.e., it has to be a fixed point of T).

Since u is a fixed point of T, we can replace $u^\star$ above by u and obtain that $(\|u_k - u\|^2)$ converges, to 0.

Finally, we get that (u_k) converges to u. This implies that (x_k) converges to $x = \mathbf{prox}_{\gamma g}(u)$, which is a point verifying $0 \in \partial g(x) + \partial h(x)$ by Eq. (6). □

Exercices

Exercise 1. Let us consider the $\mathbb{R} \to \mathbb{R}$ function

$$F(x) = \sum_{i=1}^{n} |x - i|$$

for some positive integer n.

1. Show that the minimization of F is equivalent to solving

$$\min_{y \in \mathbb{R}^n} G(y) + H(y) \text{ with } G(y) = \sum_{i=1}^{n} |y_i - i| \text{ and } H(y) = \begin{cases} 0 & \text{if } y_1 = y_2 = \cdots = y_n \\ +\infty & \text{otherwise} \end{cases}$$

2. How is the proximity operator of G related to the proximity operators of the functions (r_i).
3. Compute the proximity operator of H.
4. How can the Douglas-Rachford method be used to minimize F?

References