Convex Optimisation

Matthew Tam
(University of Melbourne)
matthew.tam@unimelb.edu.au

ACE Network
S1 2021

---

* Assignment 4
  - due June 3rd @ 5pm

The Plan

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

convex, but not differentiable

$$\Downarrow$$

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

→ subgradient method to h. ✓

→ apply the proximal part of proximal-gradient method to h.

$\Longleftrightarrow$ compute

$$\text{prox}_h = \text{prox}_{f+g}$$

prox_f, prox_g

3.5 Douglas-Rachford splitting

We consider the problem

$$\min_{x \in \mathbb{R}^n} f(x) + g(x) \tag{3.20}$$

where $f, g : \mathbb{R}^n \to (-\infty, +\infty]$ proper, lsc and convex. That is, neither function need be differentiable.

Remark: x^* solves (3.20) iff

$$0 \in \partial(f+g)(x^*) = \partial f(x^*) + \partial g(x^*)$$

assuming the sum rule hold

Given an initial $z_0 \in \mathbb{R}^n$, the Douglas-Rachford method generates sequences according to

$$\begin{cases} \overset{x}{x_k} = \mathrm{prox}_f(\overset{z}{z_k}) \\ \overset{y}{y_k} = \mathrm{prox}_g(\overset{2x-z}{2x_k - z_k}) \quad \forall k \in \mathbb{N} \\ \underbrace{z_{k+1} = z_k}_{z \,=\, z} + y_k - x_k \end{cases} \tag{3.21}$$

$z = z + y - x \iff x = y.$

Throughout today's lecture, we denote

$$\Omega = \{ z \in \mathbb{R}^n : \ x := \operatorname{prox}_f(z) = \operatorname{prox}_g(2x - z) \}.$$

Given operator T, we denote

$$\operatorname{zer}(T) := \{ z \in \mathbb{R}^n : \ 0 \in T(z) \}.$$

Lemma 3.5.1

$$\underbrace{\operatorname{zer}(\partial f + \partial g)}_{\exists x \text{ st } 0 \in \partial f(x) + \partial g(x)} \neq \phi \text{ iff } \Omega \neq \phi.$$

Proof

Let $x \in \operatorname{zer}(\partial f + \partial g)$. Since $0 \in \partial f(x) + \partial g(x)$, there exists $z' \in \mathbb{R}^n$ such that

$$z' \in \partial f(x) \text{ and } -z' \in \partial g(x).$$

By setting $z := z' + x$ and note that $\nabla\left(\frac{1}{2}\|\cdot - z\|^2\right) = \cdot - z$, we obtain

$$z - x \in \partial f(x)$$

$$\Rightarrow 0 \in \partial f(x) + x - z$$

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

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

$\Rightarrow x$ is the soln to $\min_u f(u) + \frac{1}{2}\|u - z\|^2$

$$\Rightarrow \quad x = \mathrm{prox}_f(z).$$

Similarly, noting that

$$\nabla\left(\frac{1}{2}\|\cdot - (2x - z)\|^2\right) = \cdot - (2x - z),$$

we have

$$x - z \in \partial g(x)$$

$$\Rightarrow \; 0 \in \partial g(x) + \overbrace{2 - x}^{\nabla h(x)}$$

$$= \partial\Big(g + \underbrace{\tfrac{1}{2}\|\cdot - (2x-z)\|^2}_{h}\Big)(x)$$

$\Rightarrow$ x is a soln of $\min_u g(u) + \frac{1}{2}\|u-(2x-z)\|^2$

$$\Rightarrow \; x = \operatorname{prox}_g(2x-z).$$

Thus, $z \in \Omega \Rightarrow \Omega \neq \phi$.

The converse is left as an exercise.

QED.

Aside: $h(u) = \frac{1}{2}\|u - (2x-z)\|^2$

$$\nabla h(u) = u - (2x-z).$$

$$\Rightarrow \nabla h(x) = x - (2x-z)$$

$$= z - x.$$

Lemma 3.5.2

Let $z \in \Omega$. Then the sequence (z_k) given (3.21) satisfies

$$\|z_{k+1} - z\|^2 + \|z_{k+1} - z_k\|^2 \leq \|\underbrace{z_k - z}_{(z_k - z_{k+1}) + (z_{k+1} - z)}\|^2 \quad \forall k \in \mathbb{N}.$$

Proof

Since $x := \mathrm{prox}_f(z)$ and $x_k = \mathrm{prox}_f(z_k)$, Corollary 2.3.8 applied to prox_f gives

$$0 \leq \langle \underbrace{x - x_k}_{\mathrm{prox}_f(z) - \mathrm{prox}_g(x_k)}, \underbrace{(z_k - x_k)}_{= (I - \mathrm{prox}_f)(z_k)} - \underbrace{(z - x)}_{(I - \mathrm{prox}_f)(z)} \rangle \qquad (3.22)$$

Since $x = \mathrm{prox}_g(2x - z)$ and $y_k = \mathrm{prox}_g(2x_k - z_k)$, Corollary 2.3.8 applied to prox_g gives

$$z_{k+1} = z_k + y_k - x_k.$$

$$\Leftrightarrow \quad -z_{k+1} = x_k - z_k - y_k$$

$$0 \cancel{\leq} \langle x - y_n, (2x_n - z_n - y_n) - (x - z) \rangle$$

$$= \langle x \overset{-x_n + x_n}{} - y_n, (x_n - z_{n+1}) - (x - z) \rangle \qquad (3.23)$$

$$= \langle x - x_n, (\cancel{x_n} - z_{n+1}) - (\cancel{x} - z) \rangle$$

$$+ \langle x_n - y_n, (x_n - z_{n+1}) - (x - z) \rangle$$

Adding (3.22) and (3.23) gives

$$0 \leq \langle x - x_n, z_n - z_{n+1} \rangle$$

$$+ \langle \underbrace{x_n - y_n}_{= z_n - z_{n+1}}, x_n - z_{n+1} - (x - z) \rangle$$

$$= \langle x - x_n, z_n - z_{n+1} \rangle$$

$$+ \langle z_n - z_{n+1}, x_n - z_{n+1} - (x - z) \rangle$$

$$= \langle x - x_n, z_n - z_{n+1} \rangle$$

$$+ \langle z_n - z_{n+1}, x_n - x \rangle$$

$$+ \langle z_n - z_{n+1}, z - z_{n+1} \rangle$$

$$= \langle \underline{z_n - z_{n+1}}, \underline{z - z_{n+1}} \rangle$$

The result then follows from the Polarisation identity. ∎

Theorem 3.5.3

Suppose $\operatorname{zer}(\partial f + \partial g) \neq \emptyset$. Given an initial point $z_0 \in \mathbb{R}^n$, the sequence (z_k) given by (3.21) converges to a point $z \in \Omega$. Consequently,

$$\lim_{k\to\infty} \underbrace{\operatorname{prox}_f(z_k)}_{= x_k} = \operatorname{prox}_f(z) \in \operatorname*{argmin}_{u \in \mathbb{R}^n} f(u) + g(u).$$

Proof.

By Lemma 3.5.1, there exists $z' \in \Omega$ and hence applying Lemma 3.5.2 gives

$$\|z_{k+1} - z'\|^2 + \|z_{k+1} - z_k\|^2 \leq \|z_k - z'\|^2 \quad \forall k \in \mathbb{N}. \tag{3.24}$$

Thus the sequence $\left(\|z_k - z'\|^2\right)$ is nonincreasing and bounded below by 0, and hence converges by the MCT to some $L \in \mathbb{R}$. Taking the limit in (3.24) as $k \to \infty$ gives

$$L + \lim_{k \to \infty} \|z_{k+1} - z_k\|^2 \leq L$$

$$\Rightarrow \lim_{k \to \infty} \|z_{k+1} - z_k\|^2 = 0.$$

Since (3.24) implies (z_k) is bounded, there exists a subsequence (z_{k_n}) which converges to some point $z \in \mathbb{R}^n$. Taking the limit along this subsequence gives

$$x := \mathrm{prox}_f(z) = \mathrm{prox}_g(2x - z)$$

$$\overset{\to x}{x_{k_n}} = \mathrm{prox}_f(\overset{\to z}{z_{k_n}})$$

$$\overset{\to y}{y_{k_n}} = \mathrm{prox}_g(2\overset{\to x}{x_{k_n}} - \overset{\to z}{z_{k_n}})$$

$$\underset{\to 0}{z_{k_n+1} - z_{k_n}} = \overset{\to y}{y_{k_n}} - \overset{\to x}{x_{k_n}}$$

$$\Leftrightarrow 0 = y - x \Rightarrow y = x$$

This implies that $z \in \Omega$.

Repeating the above argument with z in place of z' implies that $(\|z_k - z\|^2)$ is non increasing. And further, we have $\|z_{k_n} - z\|^2 \to 0$ as $k \to \infty$. Combining these two facts implies $\|z_k - z\|^2 \to 0$ $\Rightarrow$ $z_k \to z$ as $k \to \infty$. The last part follows by taking limits and using continuity of prox_f. $\boxed{\text{QED}}$

3.5.1 Product Space Reformulations

Consider the problem

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^{m} f_i(x)$$

where $f_i : \mathbb{R}^n \to (-\infty, +\infty]$ are proper, lsc, and convex. By making m copies of the variable $x\in\mathbb{R}$, this can be formulated as

$$\min_{x_1,\dots,x_m\in\mathbb{R}^n} \sum_{i=1}^{m} f_i(x_i)$$

$$\text{s.t.} \quad x_1 = x_2 = \dots = x_m$$

Letting $\mathbf{x} = (x_1,\dots,x_m) \in \mathbb{R}^n \times \dots \times \mathbb{R}^n \equiv (\mathbb{R}^n)^m$,

$f(\mathbf{x}) = \sum_{i=1}^{m} f_i(x_i)$ and

$$D := \{(x_1,\dots,x_m) \in \mathbb{R}^n \times \dots \times \mathbb{R}^n : x_1 = \dots = x_m\},$$

we can write this as

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + i_D(\boldsymbol{x}).$$

This is in the form of problems considered in section 3.5. The algorithm (3.21) takes the form

$$\begin{cases} \boldsymbol{x}_k = \operatorname{prox}_{i_D}(\boldsymbol{z}_k) = P_D(\boldsymbol{z}_k) \\ \boldsymbol{y}_k = \operatorname{prox}_f(2\boldsymbol{x}_k - \boldsymbol{z}_k) \\ \boldsymbol{z}_{k+1} = \boldsymbol{z}_k + \boldsymbol{y}_k - \boldsymbol{x}_k \end{cases} \qquad (3.26)$$

Under the assumptions of Theorem 3.5.?,

$$\boldsymbol{z}_k \to \boldsymbol{z} \in \Omega \text{ and } \boldsymbol{x}_k = P_D(\boldsymbol{z}_k) \to \underbrace{P_D(\boldsymbol{z})}_{=:\,\boldsymbol{x}}$$

Thus, since $\boldsymbol{x} \in D$, we have $\boldsymbol{x} = (x, \ldots, x)$ for some $x \in \mathbb{R}^n$ and solves the original problem