ISyE8813

Renato D.C. Monteiro

January 2023

1 Inexact proximal point framework

This section presents an inexact proximal point (IPP) framework for solving the convex optimization problem

$$\phi_* = \min\{\phi(x) : x \in \mathbb{R}^n\} \tag{1}$$

where $\phi \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$ and gives general results about the sequence of iterates generated by it. Assume that the optimal solution set X_* of (1) is nonempty.

IPP Framework

---

0) $\tau \geq 0$ and $x_0 \in X$ is given

1) For $k = 1, 2, \ldots$, do

   - choose stepsize $\lambda_k > 0$
   - compute $(x_k, \varepsilon_k) \in \mathbb{R}^n \times \mathbb{R}_+$ such that

$$v_k := \frac{x_{k-1} - x_k}{\lambda_k} \in \partial_{\varepsilon_k}\phi(x_k), \quad 2\lambda_k\varepsilon_k \leq \|x_k - x_{k-1}\|^2 + 2\lambda_k\tau \tag{2}$$

The main results about the IPP framework are:

Proposition 1.1 *Assume that $K \geq 1$ and $\bar{x}_K \in \mathrm{conv}\{x_1, \ldots, x_K\}$ is a point such that*

$$\phi(\bar{x}_K) \leq \frac{1}{\Lambda_K}\sum_{k=1}^{K}\lambda_k\phi(x_k) \tag{3}$$

where $\Lambda_K := \sum_{k=1}^{K}\lambda_k$. Then,

$$\phi(\bar{x}_K) - \phi_* \leq \frac{d_0^2}{2\Lambda_K} + \tau \tag{4}$$

$$\|x_K - x_*\|^2 \leq \|x_0 - x_*\|^2 + 2\Lambda_K\tau, \qquad \forall x_* \in X_* \tag{5}$$

Proposition 1.2 *Let $\bar{x}_K \in \mathrm{conv}\{x_1, \ldots, x_K\}$ be a point such that* (3) *holds and define*

$$\bar{v}_K := \frac{x_0 - x_K}{\Lambda_K}, \qquad \bar{\varepsilon}_K := \frac{1}{2\Lambda_K}\left(\|x_0 - \bar{x}_K\|^2 - \|x_K - \bar{x}_K\|^2\right) + \tau \tag{6}$$

Then, for every $K \geq 1$, we have

$$\bar{v}_K \in \partial_{\bar{\varepsilon}_K}\phi(\bar{x}_K)$$

and the following bounds hold:

$$\|\bar{v}_K\| \leq \frac{2d_0}{\Lambda_K} + \sqrt{\frac{2\tau}{\Lambda_k}}, \qquad \bar{\varepsilon}_K \leq \frac{2d_0^2}{\Lambda_K} + 3\tau$$

Remark: Two immediate examples of $\bar{x}_K$ satisfying (3) are:

$$\bar{x}_K = \frac{1}{\Lambda_K}\sum_{k=1}^{K} \lambda_k x_k$$

and

$$\bar{x}_K \in \mathrm{Argmin}\left\{\phi(x) : x \in \{x_1, \ldots, x_K\}\right\}$$

2 Special Instances of the IPP Framework

2.1 Composite Gradient Method

Assume that

$$\phi = f + h$$

where

- $h \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$
- f is convex and L-smooth on $\mathrm{dom}\, h$

Consider the following method.

Composite Gradient Method (CGM)

---

0) Let $x_0 \in X$ be given
1) For $k = 1, 2, \ldots,$ do

$$x_k = \mathrm{argmin} \left\{ \ell_f(u; x_{k-1}) + h(u) + \frac{L}{2}\|u - x_{k-1}\|^2 \right\} \tag{7}$$

where

$$\ell_f(u; x_{k-1}) = f(x_{k-1}) + \langle \nabla f(x_{k-1}), u - x_{k-1} \rangle$$

We will now show that the CGM is a special case of IPP in which

$$\tau = 0, \quad \lambda_k = \frac{1}{L} \ \ \forall k \geq 1 \tag{8}$$

The optimality condition for (7) implies that

$$0 \in \nabla f(x_{k-1}) + \partial h(x_k) + L(x_k - x_{k-1})$$

It is easy to see that

$$\nabla f(x_{k-1}) \in \partial_{\varepsilon_k} f(x_k)$$

where

$$\varepsilon_k = f(x_k) - f(x_{k-1}) - \langle \nabla f(x_{k-1}), x_k - x_{k-1} \rangle = f(x_k) - \ell_f(x_k; x_{k-1})$$

Thus, using the two observations above and the fact that $\lambda_k = 1/L$ for all k, we have

$$\begin{aligned} \frac{x_{k-1} - x_k}{\lambda_k} &= L(x_{k-1} - x_k) \in \nabla f(x_{k-1}) + \partial h(x_k) \\ &\subset \partial_{\varepsilon_k} f(x_k) + \partial h(x_k) \subset \partial_{\varepsilon_k} (f + h)(x_k) = \partial_{\varepsilon_k} \phi(x_k) \end{aligned}$$

which shows that the inclusion in (2) holds.

Second, since f is L-smooth, we have

$$\varepsilon_k = f(x_k) - \ell_f(x_k; x_{k-1}) \leq \frac{L}{2}\|x_k - x_{k-1}\|^2$$

and hence

$$2\lambda_k \varepsilon_k = \frac{2\varepsilon_k}{L} \leq \|x_k - x_{k-1}\|^2$$

where the equality is due to the fact that $\lambda_k = 1/L$ for every $k \geq 1$. Hence, the inequality in (2) holds with $\tau = 0$. We have thus shown that CGM is a special case of IPP with $\tau = 0$ and $\lambda_k = 1/L$ for every $k \geq 1$.

Corollary 2.1 *For every $K \geq 1$, the K-th iterate of CGM satisfies*

$$\phi(x_K) - \phi_* \leq \frac{Ld_0^2}{2K}$$

Proof: Follows immediately from Proposition 1.1 and the conclusion that CGM is a special case of IPP with τ and $\{\lambda_k\}$ given by (8). ■

2.2 Hybrid Composite Subgradient Method

Assume that

$$\phi = f + h$$

where

- $h \in \overline{\text{Conv}}\,(\mathbb{R}^n)$
- f is convex on $\text{dom}\, h$
- there exists a function $s : \text{dom}\, h \to \mathbb{R}^n$ satisfying the following properties:
  - $s(x) \in \partial f(x)$ for all $x \in \text{dom}\, h$
  - there exist $M, L \geq 0$ such that for every $x, x' \in \text{dom}\, h$,

$$\|s(x) - s(x')\| \leq 2M + L\|x - x'\| \tag{9}$$

Lemma 2.2 *For every $x, x' \in \text{dom}\, h$, we have*

$$f(x') - \ell_f(x'; x) \leq 2M\|x' - x\| + \frac{L}{2}\|x' - x\|^2$$

where

$$\ell_f(x'; x) := f(x) + \langle s(x), x' - x\rangle$$

Proof: Let $x, x' \in \text{dom}\, h$ be given and, for any $t \in \mathbb{R}$, define $x_t := (1 - t)x + tx'$. Then, the integral mean value theorem for subgradients, the Cauchy-Schwarz inequality, inequality (9) and the definition of x_t imply that

$$\begin{aligned}
f(x') - \ell_f(x'; x) &= f(x') - f(x) - \langle s(x), x' - x\rangle \\
&= \left(\int_0^1 \langle s(x_t), x' - x\rangle dt\right) - \langle s(x), x' - x\rangle \\
&= \int_0^1 \langle s(x_t) - s(x), x' - x\rangle dt \leq \int_0^1 \|s(x_t) - s(x)\|\, \|x' - x\| dt \\
&\overset{(9)}{\leq} \|x' - x\| \int_0^1 \left(2M + L\|x_t - x\|\right) dt \\
&= \|x' - x\| \left(2M + \int_0^1 Lt\|x' - x\| dt\right)
\end{aligned}$$

and hence that the conclusion of the lemma holds ∎

We now state the hybrid composite subgradient method (H-CSM)

H-CSM

---

0) Let $x_0 \in X$ be given and set

$$\lambda = \frac{1}{L + 4M^2/\bar{\varepsilon}} \tag{10}$$

1) For $k = 1, 2, \ldots,$ do

   - set $s_{k-1} = s(x_{k-1})$
   - compute

$$x_k = \operatorname{argmin} \left\{ \ell_f(u; x_{k-1}) + h(u) + \frac{1}{2\lambda} \|u - x_{k-1}\|^2 \right\} \tag{11}$$

   where

$$\ell_f(u; x_{k-1}) := f(x_{k-1}) + \langle s_{k-1}, u - x_{k-1} \rangle$$

We will now show that the H-CSM is a special case of IPP with

$$\tau = 0, \quad \lambda_k = \frac{1}{L + 4M^2/\bar{\varepsilon}} \quad \forall k \geq 1$$

First, note that the optimality condition for (7) implies that

$$0 \in s_{k-1} + \partial h(x_k) + \frac{x_k - x_{k-1}}{\lambda}$$

It is easy to see that

$$s_{k-1} \in \partial_{\varepsilon_k} f(x_k)$$

where

$$\varepsilon_k = f(x_k) - f(x_{k-1}) - \langle s_{k-1}, x_k - x_{k-1} \rangle = f(x_k) - \ell_f(x_k; x_{k-1})$$

Thus

$$\begin{aligned} \frac{x_{k-1} - x_k}{\lambda} &\in s_{k-1} + \partial h(x_k) \subset \partial_{\varepsilon_k} f(x_k) + \partial h(x_k) \\ &\subset \partial_{\varepsilon_k} (f + h)(x_k) = \partial_{\varepsilon_k} \phi(x_k) \end{aligned}$$

which shows that the inclusion in (2) holds with $\lambda_k = \lambda$.

Second, we have

$$\varepsilon_k = f(x_k) - \ell_f(x_k; x_{k-1}) \leq 2M \|x_k - x_{k-1}\| + \frac{L}{2} \|x_k - x_{k-1}\|^2$$

Hence

$$\begin{aligned}
& 2\lambda\varepsilon_k - \|x_k - x_{k-1}\|^2 \\
& \quad \le 4\lambda M\|x_k - x_{k-1}\| + \lambda L\|x_k - x_{k-1}\|^2 - \|x_k - x_{k-1}\|^2 \\
& \quad = 4\lambda M\|x_k - x_{k-1}\| + (\lambda L - 1)\|x_k - x_{k-1}\|^2 \\
\text{due to (10)} & \quad = 4\lambda M\|x_k - x_{k-1}\| - \frac{4\lambda M^2}{\bar{\varepsilon}}\|x_k - x_{k-1}\|^2 \\
& \quad \le 4\lambda M \max\left\{t - \frac{M}{\bar{\varepsilon}}t^2 : t \in \mathbb{R}\right\} \\
& \quad = (4\lambda M)\left(\frac{\bar{\varepsilon}}{4M}\right) = \lambda\bar{\varepsilon} \qquad (12)
\end{aligned}$$

Hence, the inequality in (2) holds with $\tau = \bar{\varepsilon}/2$ and $\lambda_k = \lambda$. We have thus shown that H-CSM is a special case of IPP with $\tau = \bar{\varepsilon}/2$ and $\lambda_k = \lambda$ for all $k \ge 1$ where λ is as in (10).

Corollary 2.3 *The sequence of iterates $\{x_k\}$ generated by H-CSM satisfies the following property: for any $K \ge 1$ and any point $\bar{x}_K \in \mathrm{conv}\{x_1, \dots, x_K\}$ such that*

$$\phi(\bar{x}_K) \le \frac{1}{K}\sum_{k=1}^{K} f(x_k),$$

there holds

$$\phi(\bar{x}_K) - \phi_* \le \frac{1}{2K}\left(L + \frac{M^2}{\bar{\varepsilon}}\right)d_0^2 + \frac{\bar{\varepsilon}}{2}.$$

Proof: Since H-CSM is a special case of IPP with $\tau = \bar{\varepsilon}/2$ and $\lambda_k = \lambda$ for all $k \ge 1$ where λ is as in (10), it follows from Prop 1.1 that

$$\phi(\bar{x}_K) - \phi_* \le \frac{d_0^2}{2\lambda K} + \frac{\bar{\varepsilon}}{2},$$

and hence that the result holds due to (10). ∎

It follows from the above result that the iteration complexity of H-CSM to find an iterate x_K such that $\phi(x_K) - \phi_* \le \bar{\varepsilon}$ is

$$\mathcal{O}\left(\left(\frac{L}{\bar{\varepsilon}} + \frac{M^2}{\bar{\varepsilon}^2}\right)d_0^2\right)$$

Corollary 2.4 *For every $K \geq 1$, the K-th triple $(\bar{x}_K, \bar{v}_K, \bar{\varepsilon}_K)$ where $\bar{x}_K$ is as in the previous corollary, $(\bar{v}_K, \bar{\varepsilon}_K)$ is given by*

$$\bar{v}_K = \frac{x_0 - x_K}{K}, \qquad \bar{\varepsilon}_K = \frac{1}{2\lambda K}\left(\|x_0 - \bar{x}_K\|^2 - \|x_K - \bar{x}_K\|^2\right) + \frac{\bar{\varepsilon}}{2},$$

and λ is as in (10)*, satisfies*

$$\bar{v}_K \in \partial_{\bar{\varepsilon}_K}\phi(\bar{x}_K)$$

and the following bounds hold:

$$\begin{aligned}
\|\bar{v}_K\| &\leq 2\left(L + \frac{M^2}{\bar{\varepsilon}}\right)\frac{d_0}{K} + \sqrt{\left(L + \frac{M^2}{\bar{\varepsilon}}\right)\frac{\bar{\varepsilon}}{K}}, \\
\bar{\varepsilon}_K &\leq \frac{1}{K}\left[2\left(L + \frac{M^2}{\bar{\varepsilon}}\right)d_0^2\right] + \frac{3\bar{\varepsilon}}{2}.
\end{aligned}$$

It is easy to see that if $K = \Omega(1/\bar{\varepsilon}^2)$ then

$$\max\{\|\bar{v}_K\|, \bar{\varepsilon}_K\} = \mathcal{O}(\varepsilon)$$

3 Analysis of the IPP framework

Lemma 3.1 *For every $k \geq 1$ and $u \in \mathbb{R}^n$, we have:*

$$\phi(x_k) - \phi(u) \leq \langle v_k, x_k - u \rangle + \varepsilon_k$$

Proof: Follows immediately from the fact that $v_k \in \partial_{\varepsilon_k} \phi(x_k)$ and the definition of the ε-subdifferential. ■

Lemma 3.2 *For every $k \geq 1$, we have*

$$2\lambda_k[\phi(x_{k-1}) - \phi(x_k)] \geq \|\lambda_k v_k\|^2 - 2\lambda_k \tau \tag{13}$$

Proof: Taking $u = x_{k-1}$ in Lemma 3.1 and noting the definition of v_k, we have

$$\phi(x_k) - \phi(x_{k-1}) \leq -\frac{1}{\lambda_k}\|x_k - x_{k-1}\|^2 + \varepsilon_k$$

and hence

$$2\lambda_k[\phi(x_k) - \phi(x_{k-1})] \leq -2\|x_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq -\|x_k - x_{k-1}\|^2 + 2\lambda_k \tau$$

The conclusion of the lemma now follows from the above inequality and the definition of v_k. ■

Lemma 3.3 *For every $k \geq 1$ and $u \in \mathbb{R}^n$, we have*

$$2\lambda_k[\phi(x_k) - \phi(u)] \leq \|x_{k-1} - u\|^2 - \|x_k - u\|^2 + 2\lambda_k \tau$$

Proof: Have

$$\begin{aligned}
\|x_{k-1} - u\|^2 - \|x_k - u\|^2 &= \|x_{k-1} - x_k + x_k - u\|^2 - \|x_k - u\|^2 \\
&= \|x_{k-1} - x_k\|^2 + 2\langle x_{k-1} - x_k, x_k - u \rangle \\
&= \|x_{k-1} - x_k\|^2 + 2\lambda_k \langle v_k, x_k - u \rangle \\
&\geq \|x_{k-1} - x_k\|^2 + 2\lambda_k[\phi(x_k) - \phi(u) - \varepsilon_k] \quad \text{(Lemma 3.1)} \\
&\geq -2\lambda_k \tau + 2\lambda_k[\phi(x_k) - \phi(u)]
\end{aligned}$$

where the last inequality is due to the inequality in (2). ■

Proposition 3.4 *For every $K \geq 1$, define*

$$\Lambda_K := \sum_{k=1}^{K} \lambda_k. \tag{14}$$

Then, the following statements hold for every $K \geq 1$ and $u \in \mathbb{R}^n$:

a) have

$$2\sum_{k=1}^{K} \lambda_k[\phi(x_k) - \phi(u)] \leq \|x_0 - u\|^2 - \|x_K - u\|^2 + 2\tau\Lambda_K;$$

b) if $\bar{x}_K \in \text{conv}\{x_1, \dots, x_K\}$ is a point such that

$$\phi(\bar{x}_K) \le \frac{1}{\Lambda_K} \sum_{k=1}^{K} \lambda_k \phi(x_k), \tag{15}$$

then

$$\phi(\bar{x}_K) - \phi(u) \le \frac{1}{2\Lambda_K} \left(\|x_0 - u\|^2 - \|x_K - u\|^2 \right) + \tau \tag{16}$$

Proof: a) This statement follows by summing the inequality in Lemma 3.3 from $k = 1$ to $k = K$.

b) This statement follows immediately from a) and assumption (15). ■

Corollary 3.5 *For any $K \ge 1$ and any point $\bar{x}_K \in \text{conv}\{x_1, \dots, x_K\}$ such that* (15) *holds, we have*

$$\phi(\bar{x}_K) - \phi_* \le \frac{d_0^2}{2\Lambda_K} + \tau \tag{17}$$

$$\|x_K - x_*\|^2 \le \|x_0 - x_*\|^2 + 2\Lambda_K \tau, \qquad \forall x_* \in X_* \tag{18}$$

Proof: Inequality (17) follows immediately from Proposition 3.4(b) with $u = \text{Proj}_{X_*}(x_0)$. Moreover, for any $x_* \in X_*$, (18) follows from Proposition 3.4(b) with $u = x_*$ and the fact that $\phi(\bar{x}_K) \ge \phi(x_*) = \phi_*$. ■

Corollary 3.6 *Let $\bar{x}_K \in \text{conv}\{x_1, \dots, x_K\}$ be a point such that* (15) *holds and define*

$$\bar{v}_K := \frac{x_0 - x_K}{\Lambda_K}$$

$$\bar{\varepsilon}_K := \frac{1}{2\Lambda_K} \left(\|x_0 - \bar{x}_K\|^2 - \|x_K - \bar{x}_K\|^2 \right) + \tau$$

Then, for every $K \ge 1$, we have

$$\bar{v}_K \in \partial_{\bar{\varepsilon}_K} \phi(\bar{x}_K) \tag{19}$$

and the following bounds hold:

$$\|\bar{v}_K\| \le \frac{2d_0}{\Lambda_K} + \sqrt{\frac{2\tau}{\Lambda_K}}, \qquad \bar{\varepsilon}_K \le \frac{2d_0^2}{\Lambda_K} + 3\tau \tag{20}$$

Proof: We first show that the inclusion (19) holds. Let $\mathcal{A}_K(u)$ denote the right hand side of (16) as a function of u. It is easy to see that $\mathcal{A}_K$ is an affine function whose gradient is $\nabla \mathcal{A}_K = (x_0 - x_k)/\Lambda_k = \bar{v}_K$ and whose value at $\bar{x}_K$ is $\bar{\varepsilon}_K$. Hence,

$$\mathcal{A}_K(u) = \mathcal{A}_K(\bar{x}_K) + \langle \nabla \mathcal{A}_K, u - \bar{x}_K \rangle = \bar{\varepsilon}_K + \langle \bar{v}_K, u - \bar{x}_K \rangle \quad \forall u \in \mathbb{R}^n.$$

It then follows from the above identity and (16) that

$$\phi(\bar{x}_K) - \phi(u) \le \bar{\varepsilon}_K + \langle \bar{v}_K, u - \bar{x}_K \rangle \quad \forall u \in \mathbb{R}^n.$$

This inequality together with the definition of the ε-subdifferential then imply that (19) holds.

We will now show that the two bounds in (20) hold. Letting $x_* = \mathrm{Proj}_{X_*}(x_0)$, and using the triangle inequality for norms and (18), we have that for every $k = 1, \dots, K$,

$$\begin{aligned} \|x_0 - x_k\| &\le \|x_0 - x_*\| + \|x_k - x_*\| \\ &\le \|x_0 - x_*\| + \sqrt{\|x_0 - x_*\|^2 + 2\Lambda_k \tau} \\ &\le d_0 + \sqrt{d_0^2 + 2\Lambda_k \tau} \le d_0 + \sqrt{d_0^2 + 2\Lambda_K \tau}. \end{aligned} \tag{21}$$

The above inequality with $k = K$ and the definition of $\bar{v}_K$ then imply that

$$\|\bar{v}_K\| = \frac{\|x_0 - x_K\|}{\Lambda_K} \le \frac{d_0 + \sqrt{d_0^2 + 2\Lambda_K \tau}}{\Lambda_K} \le \frac{2d_0}{\Lambda_K} + \sqrt{\frac{2\tau}{\Lambda_K}},$$

and hence that the first inequality in (20) holds. Now, the definition of $\bar{\varepsilon}_K$ implies that

$$\bar{\varepsilon}_K \le \frac{\|x_0 - \bar{x}_K\|^2}{2\Lambda_K} + \tau$$

Using the fact that $\bar{x}_K \in \mathrm{conv}\{x_1, \dots, x_K\}$ and (21), we easily see that

$$\|x_0 - \bar{x}_K\| \le \max\{\|x_0 - x_k\| : k = 1, \dots, K\} \le 2d_0 + \sqrt{2\Lambda_K \tau}.$$

The above two conclusions then imply that

$$\bar{\varepsilon}_K - \tau \le \frac{\|x_0 - \bar{x}_K\|^2}{2\Lambda_K} \le \frac{\left(d_0 + \sqrt{d_0^2 + 2\Lambda_K \tau}\right)^2}{2\Lambda_K} \le \frac{2d_0^2 + 2(d_0^2 + 2\Lambda_K \tau)}{2\Lambda_K},$$

and hence that the second inequality in (20) holds. ∎