Inexact Proximal Point Framework

Lecturer: Jiaming Liang

October 5, 2023

1 Proximal point method

We are interested in solving

$$\min\{\phi(x) := f(x) + h(x)\}$$

- h is closed and convex;
- f is closed and convex, $\operatorname{dom} h \subseteq \operatorname{dom} f$;
- the optimal set X_* is nonempty.

Algorithm 1 Proximal point method

Input: Initial point $x_0 \in \operatorname{dom} h$ and constant stepsize $\lambda > 0$
for $k \geq 0$ **do**
  Solve $x_{k+1} = \operatorname{argmin}_{x\in\mathbb{R}^n}\{\phi(x) + \frac{1}{2\lambda}\|x - x_k\|^2\}$.
end for

Theorem 1.

$$\phi(x_k) - \phi_* \leq \frac{1}{2\lambda k}\|x_0 - x_*\|^2$$

Proof. It follows from the optimality of x_{k+1} that for every $x \in \operatorname{dom} h$,

$$\phi(x) + \frac{1}{2\lambda}\|x - x_k\|^2 \geq \phi(x_{k+1}) + \frac{1}{2\lambda}\|x_{k+1} - x_k\|^2 + \frac{1}{2\lambda}\|x - x_{k+1}\|^2.$$

Taking $x = x_k$, we have

$$\phi(x_k) \geq \phi(x_{k+1}) + \frac{1}{\lambda}\|x_{k+1} - x_k\|^2,$$

and hence this is a descent method. Taking $x = x_*$, we have

$$\phi_* + \frac{1}{2\lambda}\|x_k - x_*\|^2 \geq \phi(x_{k+1}) + \frac{1}{2\lambda}\|x_{k+1} - x_k\|^2 + \frac{1}{2\lambda}\|x_{k+1} - x_*\|^2.$$

Rearranging the above inequality, we obtain

$$\phi(x_{k+1}) - \phi_* \leq \frac{1}{2\lambda}\|x_k - x_*\|^2 - \frac{1}{2\lambda}\|x_{k+1} - x_*\|^2.$$

Summing the resulting inequality and using the descent property, we have

$$k[\phi(x_k) - \phi_*] \le \sum_{i=1}^{k}[\phi(x_i) - \phi_*] \le \frac{1}{2\lambda}\|x_0 - x_*\|^2 - \frac{1}{2\lambda}\|x_k - x_*\|^2 \le \frac{1}{2\lambda}\|x_0 - x_*\|^2.$$

Therefore, the conclusion of the theorem follows. □

2 Inexact proximal point framework

The proximal point method is more conceptual than practical. In practice, we usually design algorithms to approximate the solution x_{k+1} to the proximal subproblem. Algorithms solving the proximal subproblem approximately can be described and analyzed under the inexact proximal point (IPP) framework.

2.1 Algorithm

Algorithm 2 Inexact proximal point framework

Input: Initial point $x_0 \in \operatorname{dom} h$ and scalar $\sigma \in (0, 1]$

for $k \ge 1$ **do**

Step 1. Choose $\lambda_k > 0$ and $\delta_k \ge 0$.

Step 2. Compute $(x_k, \tilde{x}_k, \varepsilon_k)$ such that

$$\frac{x_{k-1} - x_k}{\lambda_k} \in \partial_{\varepsilon_k}\phi(\tilde{x}_k), \tag{1}$$

$$\|x_k - \tilde{x}_k\|^2 + 2\lambda_k\varepsilon_k \le \sigma\|\tilde{x}_k - x_{k-1}\|^2 + \delta_k. \tag{2}$$

end for

The inclusion (1) in the IPP framework means

$$\tilde{v}_k = \frac{\tilde{x}_k - x_k}{\lambda_k} \in \partial_{\varepsilon_k}\left(\phi(\cdot) + \frac{1}{2\lambda_k}\|\cdot - x_{k-1}\|^2\right)(\tilde{x}_k).$$

In contrast to the PPM, the above inclusion provides two relaxations $\tilde{v}_k$ and ε_k. If both $\tilde{v}_k = 0$ (i.e., $\tilde{x}_k = x_k$) and $\varepsilon_k = 0$, then

$$0 \in \partial\left(\phi(\cdot) + \frac{1}{2\lambda_k}\|\cdot - x_{k-1}\|^2\right)(x_k),$$

i.e., the proximal problem is solved exactly

$$x_k = \operatorname{argmin}_{x\in\mathbb{R}^n}\left\{\phi(x) + \frac{1}{2\lambda_k}\|x - x_{k-1}\|^2\right\}.$$

Moreover, the inequality (2) is automatically satisfied

$$\|x_k - \tilde{x}_k\|^2 + 2\lambda_k\varepsilon_k = \|\tilde{v}_k\|^2 + 2\lambda_k\varepsilon_k = 0 \le \sigma\|\tilde{x}_k - x_{k-1}\|^2 + \delta_k.$$

Hence, the IPP framework becomes the PPM.

2.2 Convergence

Lemma 1. *Define* $v_k = (x_{k-1} - x_k)/\lambda_k$ *and*

$$\Gamma_k(x) = \phi(\tilde{x}_k) + \langle v_k, x - \tilde{x}_k\rangle - \varepsilon_k.$$

Then, the following statements hold:

(a) for every $x \in \operatorname{dom} h$*, we have*

$$\Gamma_k(x) \le \phi(x)$$

(b) the following inequality provides an upper bound for the optimality gap for the proximal problem

$$\phi(\tilde{x}_k) + \frac{1}{2\lambda_k}\|\tilde{x}_k - x_{k-1}\|^2 - \phi(x_k) - \frac{1}{2\lambda_k}\|x_k - x_{k-1}\|^2 \le \frac{\sigma}{2\lambda_k}\|\tilde{x}_k - x_{k-1}\|^2 + \frac{\delta_k}{\lambda_k};$$

(c)

$$x_k = \operatorname{argmin}\left\{\lambda_k\Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2\right\};$$

(d)

$$\min\left\{\lambda_k\Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2\right\} \ge \lambda_k\phi(\tilde{x}_k) + \frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 - \delta_k.$$

Proof. (a) It follows from (1) and the definition of Γ_k that

$$\phi(x) \ge \phi(\tilde{x}_k) + \langle v_k, x - \tilde{x}_k\rangle - \varepsilon_k = \Gamma_k(x), \quad \forall x \in \operatorname{dom} h.$$

(b) Taking $x = x_k$ in the inequality in the proof of (a) and rearranging the terms, we have

$$\phi(x) \ge \phi(\tilde{x}_k) + \langle v_k, x - \tilde{x}_k\rangle - \varepsilon_k, \quad \forall x \in \operatorname{dom} h.$$

Adding a quadratic function $\frac{1}{2\lambda_k}\|\tilde{x}_k - x_{k-1}\|^2$ and subtracting a quadratic fucntion $\frac{1}{2\lambda_k}\|x_k - x_{k-1}\|^2$, we have

$$\begin{aligned}
&\phi(\tilde{x}_k) + \frac{1}{2\lambda_k}\|\tilde{x}_k - x_{k-1}\|^2 - \phi(x_k) - \frac{1}{2\lambda_k}\|x_k - x_{k-1}\|^2 \\
\le&\frac{1}{2\lambda_k}\|\tilde{x}_k - x_{k-1}\|^2 - \frac{1}{2\lambda_k}\|x_k - x_{k-1}\|^2 + \frac{1}{\lambda_k}\langle x_{k-1} - x_k, \tilde{x}_k - x_k\rangle + \varepsilon_k \\
=&\frac{1}{2\lambda_k}\|\tilde{x}_k - x_k\|^2 + \varepsilon_k \\
\le&\frac{\sigma}{2\lambda_k}\|\tilde{x}_k - x_{k-1}\|^2 + \frac{\delta_k}{\lambda_k},
\end{aligned}$$

where the last inequality follows from (2).

(c) This is obvious in view of the definition of Γ_k.

(d) For any $x \in \operatorname{dom} h$, it is easy to see that

$$\begin{aligned}\min_x \left\{ \lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2 \right\} &= \lambda_k \Gamma_k(x_k) + \frac{1}{2}\|x_k - x_{k-1}\|^2 \\ &= \lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2 - \frac{1}{2}\|x - x_k\|^2.\end{aligned}$$

Taking $x = \tilde{x}_k$ in the above identity, we have

$$\begin{aligned}\min_x \left\{ \lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2 \right\} &= \lambda_k \Gamma_k(\tilde{x}_k) + \frac{1}{2}\|\tilde{x}_k - x_{k-1}\|^2 - \frac{1}{2}\|\tilde{x}_k - x_k\|^2 \\ &= \lambda_k \phi(\tilde{x}_k) + \frac{1}{2}\left(\|\tilde{x}_k - x_{k-1}\|^2 - \|\tilde{x}_k - x_k\|^2 - 2\lambda_k \varepsilon_k\right) \\ &\ge \lambda_k \phi(\tilde{x}_k) + \frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 - \delta_k,\end{aligned}$$

where the second identity is due to the definition of Γ_k and the inequality is due to (2). □

Lemma 2. *For $k \ge 1$, the IPP framework satisfies*

$$\lambda_k[\phi(\tilde{x}_k) - \phi_*] + \frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 \le \delta_k + \frac{1}{2}\|x_{k-1} - x_*\|^2 - \frac{1}{2}\|x_k - x_*\|^2.$$

Proof. Using Lemma 1(a) and (d), we have

$$\begin{aligned}\lambda_k \phi(x) + \frac{1}{2}\|x - x_{k-1}\|^2 &\ge \lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2 \\ &= \min \left\{ \lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2 \right\} + \frac{1}{2}\|x - x_k\|^2 \\ &\ge \lambda_k \phi(\tilde{x}_k) + \frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 - \delta_k + \frac{1}{2}\|x - x_k\|^2. \qquad (3)\end{aligned}$$

It is worth noting that the above inequality generalizes Lemma 1(b) since rearranging gives

$$\phi(\tilde{x}_k) + \frac{1}{2\lambda_k}\|\tilde{x}_k - x_{k-1}\|^2 - \phi(x) - \frac{1}{2\lambda_k}\|x - x_{k-1}\|^2 \le \frac{\sigma}{2\lambda_k}\|\tilde{x}_k - x_{k-1}\|^2 + \frac{\delta_k}{\lambda_k} - \frac{1}{2\lambda_k}\|x - x_k\|^2.$$

Take $x = x_*$ in (3), we have

$$\lambda_k \phi_* + \frac{1}{2}\|x_{k-1} - x_*\|^2 \ge \lambda_k \phi(\tilde{x}_k) + \frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 - \delta_k + \frac{1}{2}\|x_k - x_*\|^2.$$

Therefore, the lemma follows from rearranging the terms. □

Theorem 2. *Let*

$$\sum_{i=1}^{k} \lambda_i = \Lambda_k, \quad \hat{x}_k = \frac{\sum_{i=1}^{k} \lambda_i \tilde{x}_i}{\Lambda_k}.$$

Then, we have

$$\phi(\hat{x}_k) - \phi_* \le \frac{\sum_{i=1}^{k} \delta_i}{\Lambda_k} + \frac{\|x_0 - x_*\|^2}{2\Lambda_k}.$$

Proof. Summing the inequality in Lemma 2 from $k = 1$ to k and dividing by Λ_k, we have

$$\frac{\sum_{i=1}^{k} \lambda_i [\phi(\tilde{x}_i) - \phi_*]}{\Lambda_k} \le \frac{\sum_{i=1}^{k} \delta_i}{\Lambda_k} + \frac{1}{2\Lambda_k}\|x_0 - x_*\|^2.$$

It follows from the convexity of ϕ and the definition of $\hat{x}_k$ that

$$\phi(\hat{x}_k) - \phi_* \le \frac{\sum_{i=1}^{k} \lambda_i [\phi(\tilde{x}_i) - \phi_*]}{\Lambda_k}.$$

The conclusion immediately follows from the above two inequalities. □

Corollary 1. *If $\lambda_k = \lambda$ and $\delta_k = 0$ for every $k \ge 1$, then*

$$\min_{1 \le i \le k} \phi(\tilde{x}_i) - \phi_* \le \frac{\|x_0 - x_*\|^2}{2\lambda k}.$$

2.3 Proximal gradient method as an example

In this subsection, we assume f is L-smooth, then we show that the proximal gradient (PG) method with stepsize $\lambda_k \le \sigma/L$ for some $\sigma \in (0, 1]$ is an instance of the IPP framework. We begin with an iteration of the PG method

$$x_k = \operatorname*{argmin}_{x \in \mathbb{R}^n} \left\{ \ell_f(x; x_{k-1}) + h(x) + \frac{1}{2\lambda_k}\|x - x_{k-1}\|^2 \right\}. \tag{4}$$

The optimality condition is

$$\frac{x_{k-1} - x_k}{\lambda_k} \in \partial[\ell_f(\cdot; x_{k-1}) + h(\cdot)](x_k),$$

which means for every $x \in \operatorname{dom} h$,

$$\begin{aligned}
\phi(x) &\ge \ell_f(x; x_{k-1}) + h(x) \\
&\ge \ell_f(x_k; x_{k-1}) + h(x_k) + \frac{1}{\lambda_k}\langle x_{k-1} - x_k, x - x_k \rangle \\
&= \phi(x_k) + \frac{1}{\lambda_k}\langle x_{k-1} - x_k, x - x_k \rangle - \varepsilon_k
\end{aligned}$$

where the first inequality is due to the convexity of f and ε_k is defined as

$$\varepsilon_k := f(x_k) - \ell_f(x_k; x_{k-1}).$$

Hence, PG satisfies the inclusion (1) of IPP with

$$\tilde{x}_k = x_k, \quad \varepsilon_k = f(x_k) - \ell_f(x_k; x_{k-1}), \quad \delta_k = 0.$$

Moreover, it follows from the assumption that f is L-smooth that

$$\varepsilon_k = f(x_k) - \ell_f(x_k; x_{k-1}) \le \frac{L}{2}\|x_k - x_{k-1}\|^2 = \frac{L}{2}\|\tilde{x}_k - x_{k-1}\|^2 \le \frac{\sigma}{2\lambda_k}\|\tilde{x}_k - x_{k-1}\|^2.$$

Hence, the inequality (2) of IPP is also satisfied. Now, we have shown PG is an instance of IPP.

Next, let us show the convergence of PG using the general convergence guarantee of IPP. It follows from (4) and L-smoothness of f that

$$\begin{aligned}
\phi(x_{k-1}) &\ge \ell_f(x_k; x_{k-1}) + h(x_k) + \frac{1}{\lambda_k}\|x_k - x_{k-1}\|^2 \\
&\ge \phi(x_k) + \left(\frac{1}{\lambda_k} - \frac{L}{2}\right)\|x_k - x_{k-1}\|^2 \\
&\ge \phi(x_k) + \frac{2-\sigma}{2\lambda_k}\|x_k - x_{k-1}\|^2 \ge \phi(x_k) + \frac{1}{2\lambda_k}\|x_k - x_{k-1}\|^2,
\end{aligned}$$

where the second last inequality is due to $\lambda_k \le \sigma/L$ and the last inequality is due to $\sigma \le 1$. Suppose $\lambda_k = \lambda = \sigma/L$, then it follows from Corollary 1 that

$$\phi(x_k) - \phi_* = \min_{1 \le i \le k} \phi(\tilde{x}_i) - \phi_* \le \frac{\|x_0 - x_*\|^2}{2\lambda k} = \frac{L\|x_0 - x_*\|^2}{2\sigma k}.$$

3 Monotone operators and generalized IPP framework

Consider a point-to-set map T, and our goal is to find a point x_* such that zero belongs to the set $T(x_*)$, i.e.,

$$0 \in T(x_*).$$

This task generally includes optimization and other problems with convex structures.

Definition 1. *Consider a point-to-set map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, the graph of T is*

$$\operatorname{Gr} T = \{(x, v) : v \in T(x)\}.$$

Definition 2. *T is monotone if for every $(x, v) \in \operatorname{Gr} T$ and $(\tilde{x}, \tilde{v}) \in \operatorname{Gr} T$, we have*

$$\langle \tilde{x} - x, \tilde{v} - v \rangle \ge 0.$$

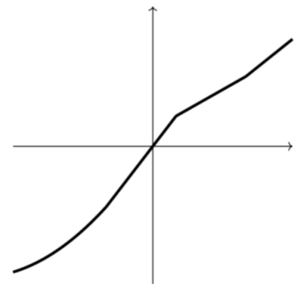

Figure 1: maximal monotone

Definition 3. *T is maximal monotone if T is monotone and for any monotone $\tilde{T}$ such that $\operatorname{Gr}\tilde{T} \subset \operatorname{Gr} T$, we have $T = \tilde{T}$ (i.e., $\operatorname{Gr} T = \operatorname{Gr}\tilde{T}$).*

Examples

(a) f is a closed convex function then $T = \partial f$ is maximal monotone. Also, x_* is an optimal solution to $\min\{f(x) : x \in \mathbb{R}^n\} \iff 0 \in \partial f(x_*) = T(x_*)$.

(b) $Q \subseteq \mathbb{R}^n$ is a closed convex set, then $T = N_Q = \partial I_Q$ is maximal monotone where

$$N_Q(x) = \{n : \langle n, \tilde{x} - x\rangle \leq 0, \forall \tilde{x} \in Q\}.$$

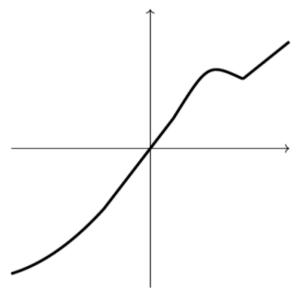

Figure 2: nonmonotone

3.1 Generalizd proximal point method

Our goal is find z_* such that

$$0 \in T(z_*),$$

which is equivalent to for some $\lambda > 0$

$$0 \in \lambda T(z_*).$$

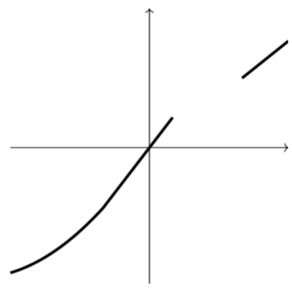

Figure 3: monotone-but-not-maximal

This is the same as

$$z_* \in z_* + \lambda T(z_*) = (I + \lambda T)(z_*),$$

and hence the fixed point problem

$$z_* = (I + \lambda T)^{-1}(z_*).$$

The following lemma indeed shows that the inverse map $(I + \lambda T)^{-1}$ is contracive and unique.

Lemma 3. *Assume T is monotone and $\lambda > 0$. Let*

$$z_1 \in (I + \lambda T)(w_1), \quad z_2 \in (I + \lambda T)(w_2), \tag{5}$$

then

$$\|w_1 - w_2\| \leq \|z_1 - z_2\|. \tag{6}$$

The mapping $(I + \lambda T)^{-1}$ is unique. As a result, there exists a unique z_* such that $0 \in T(z_*)$.

Proof. It follows from (5) that

$$z_1 - w_1 \in \lambda T(w_1), \quad z_2 - w_2 \in \lambda T(w_2).$$

Since λT is monotone, we have

$$\langle z_1 - w_1 - (z_2 - w_2), w_1 - w_2 \rangle \geq 0,$$

and hene

$$\langle z_1 - z_2, w_1 - w_2 \rangle \geq \|w_1 - w_2\|^2.$$

It follows from the Cauchy-Schwarz inequality that (6) holds. Assume for the contrary that $w_1 \neq w_2$ and

$$z \in (I + \lambda T)(w_1), \quad z \in (I + \lambda T)(w_2).$$

Clearly, it follows from (6) that

$$\|w_1 - w_2\| \leq \|z - z\| = 0.$$

So

$$w_1 = w_2.$$

Moreover, z_* is a unique fixed point

$$z_* = (I + \lambda T)^{-1}(z_*).$$

Following the argument before this lemma, we can show that $0 \in T(z_*)$. □

The fixed-point iteration

$$x_k = (I + \lambda_k T)^{-1}(x_{k-1}),$$

i.e.,

$$x_{k-1} \in (I + \lambda_k T)(x_k),$$

motivates us the following proximal point-type method

$$\frac{x_{k-1} - x_k}{\lambda_k} \in T(x_k).$$

Algorithm 3 Generalized proximal point framework

Input: Initial point $x_0 \in \operatorname{dom} h$

for $k \geq 1$ **do**

Step 1. Choose $\lambda_k > 0$.

Step 2. Compute x_k such that

$$\frac{x_{k-1} - x_k}{\lambda_k} \in T(x_k).$$

end for

Lemma 4. *For every $k \geq 1$, we have*

$$\frac{1}{2}\|x_k - x_{k-1}\|^2 \leq \frac{1}{2}\|x_{k-1} - x_*\|^2 - \frac{1}{2}\|x_k - x_*\|^2.$$

Proof. Clearly, we have

$$\frac{x_{k-1} - x_k}{\lambda_k} \in T(x_k), \quad 0 \in T(x_*).$$

It follows from Definition 2 that

$$\langle \frac{x_{k-1} - x_k}{\lambda_k}, x_k - x_* \rangle \geq 0,$$

and hence

$$\langle x_{k-1} - x_k, x_k - x_* \rangle \geq 0.$$

This is equivalent to

$$\frac{1}{2}\|x_k - x_{k-1}\|^2 \le \frac{1}{2}\|x_{k-1} - x_*\|^2 - \frac{1}{2}\|x_k - x_*\|^2.$$

□

Theorem 3. *There exists $i \le k$ such that*

$$\|v_i\|^2 \le \frac{\|x_0 - x_*\|^2}{\lambda_i^2 k},$$

where

$$v_i = \frac{x_{i-1} - x_i}{\lambda_i}.$$

3.2 Generalizd inexact proximal point framework

We introduce the ε-enlargement of a monotone operator, which can be viewed as a generalization of the ε-subdifferential.

Definition 4. *The ε-enlargement of T denoted by T^ε is defined as*

$$T^\varepsilon(\tilde{x}) = \{\tilde{v} : \langle \tilde{x} - x, \tilde{v} - v\rangle \ge -\varepsilon, \quad \forall (x, v) \in \operatorname{Gr} T\}.$$

Now, we are ready to present the generalized IPP framework.

Algorithm 4 Generalized inexact proximal point framework

Input: Initial point $x_0 \in \operatorname{dom} h$ and scalar $\sigma \in (0, 1)$
for $k \ge 1$ **do**
Step 1. Choose $\lambda_k > 0$ and $\delta_k \ge 0$.
Step 2. Compute $(x_k, \tilde{x}_k, \varepsilon_k)$ such that

$$\frac{x_{k-1} - x_k}{\lambda_k} \in T^{\varepsilon_k}(\tilde{x}_k),$$
$$\|x_k - \tilde{x}_k\|^2 + 2\lambda_k\varepsilon_k \le \sigma\|\tilde{x}_k - x_{k-1}\|^2 + \delta_k. \tag{7}$$

end for

3.3 Pointwise convergence

The following lemma is a generalization of Lemma 1.

Lemma 5. *Define $v_k = (x_{k-1} - x_k)/\lambda_k$ and*

$$\Gamma_k(x) = \langle v_k, x - \tilde{x}_k\rangle - \varepsilon_k.$$

Then, the following statements hold:

(a) for every $x_* \in T^{-1}(0)$, we have

$$\Gamma_k(x_*) \le 0;$$

(b)

$$x_k = \operatorname{argmin}\left\{\lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2\right\};$$

(c)

$$\min\left\{\lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2\right\} \ge \frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 - \delta_k.$$

Proof. (a) We know $v_k \in T^{\varepsilon_k}(\tilde{x}_k)$. For $(x_*, 0) \in \operatorname{Gr} T$, we have

$$\langle v_k - 0, \tilde{x}_k - x_* \rangle \ge -\varepsilon_k.$$

By definition, we have $\Gamma_k(x_*) \le 0$.

(b) This is obvious by the definition of Γ_k.

(c) For any $x \in \operatorname{dom} h$, we have

$$\begin{aligned}\min\left\{\lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2\right\} &= \lambda_k \Gamma_k(x_k) + \frac{1}{2}\|x_k - x_{k-1}\|^2 \\ &= \lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2 - \frac{1}{2}\|x - x_k\|^2.\end{aligned}$$

Taking $x = \tilde{x}_k$, we obtain

$$\begin{aligned}\min\left\{\lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2\right\} &= \lambda_k \Gamma_k(\tilde{x}_k) + \frac{1}{2}\|\tilde{x}_k - x_{k-1}\|^2 - \frac{1}{2}\|\tilde{x}_k - x_k\|^2 \\ &= \frac{1}{2}\left(\|\tilde{x}_k - x_{k-1}\|^2 - \|\tilde{x}_k - x_k\|^2 - 2\lambda_k \varepsilon_k\right) \\ &\ge \frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 - \delta_k,\end{aligned}$$

where the second identity follows from the definition of Γ_k and the inequality is due to (7). □

Lemma 6. *For every $k \ge 1$, we have*

$$-\lambda_k \Gamma_k(x_*) + \frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 \le \delta_k + \frac{1}{2}\|x_* - x_{k-1}\|^2 - \frac{1}{2}\|x_* - x_k\|^2.$$

Proof. Using Lemma 5 (c), we have

$$\begin{aligned}\lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2 &= \min\left\{\lambda_k \Gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2\right\} + \frac{1}{2}\|x - x_k\|^2 \\ &\ge \frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 - \delta_k + \frac{1}{2}\|x - x_k\|^2.\end{aligned}$$

Taking $x = x_*$, we have

$$\frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 \le \delta_k + \frac{1}{2}\|x_{k-1} - x_*\|^2 - \frac{1}{2}\|x_k - x_*\|^2 + \lambda_k \Gamma_k(x_*).$$

The conclusion of the lemma follows by rearranging the terms. $\square$

From now on, for simplicity, let us consider the cae $\delta_k = 0$ for every $k \ge 1$, i.e., the smooth case.

Lemma 7. *Let*

$$v_k = \frac{x_{k-1} - x_k}{\lambda_k}, \quad \theta_k = \max\left\{\frac{2\lambda_k \varepsilon_k}{\sigma}, \frac{\lambda_k^2\|v_k\|^2}{(1+\sqrt{\sigma})^2}\right\}, \tag{8}$$

then

$$\theta_k \le \|\tilde{x}_k - x_{k-1}\|^2.$$

Proof. It follows from (7) with $\delta_k = 0$ that

$$\|x_k - \tilde{x}_k\|^2 + 2\lambda_k \varepsilon_k \le \sigma \|\tilde{x}_k - x_{k-1}\|^2. \tag{9}$$

Hence, we have

$$\|\tilde{x}_k - x_{k-1} + \lambda_k v_k\| = \|\tilde{x}_k - x_k\| \le \sqrt{\sigma}\|\tilde{x}_k - x_{k-1}\|.$$

Using the triangle inequality, we have

$$\lambda_k \|v_k\| \le (1+\sqrt{\sigma})\|\tilde{x}_k - x_{k-1}\|.$$

It follows from (9) that

$$2\lambda_k \varepsilon_k \le \sigma \|\tilde{x}_k - x_{k-1}\|^2.$$

Using the above two inequalities and the defition of θ_k, we conclude the lemma holds. $\square$

Given $(\bar{\rho}, \bar{\varepsilon}) \in \mathbb{R}^2_{++}$, using the generalized IPP framework, we want to find $(\tilde{x}, v, \varepsilon)$ satisfying

$$v \in T^{\varepsilon}(\tilde{x}), \quad \|v\| \le \bar{\rho}, \quad \varepsilon \le \bar{\varepsilon}.$$

Theorem 4. ***(Pointwise convergence)*** *The generalized IPP framework satisfies*

$$\min_{1 \le i \le k} \theta_i \le \frac{\|x_0 - x_*\|^2}{(1-\sigma)k}. \tag{10}$$

Then, there exists $i \le k$ such that

$$\varepsilon_i \le \frac{\sigma \|x_0 - x_*\|^2}{2\lambda_i (1-\sigma)k}, \quad \|v_i\|^2 \le \frac{(1+\sqrt{\sigma})^2 \|x_0 - x_*\|^2}{\lambda_i^2 (1-\sigma)k}. \tag{11}$$

Proof. It follows from Lemma 5(a) and Lemma 6 that

$$\frac{1-\sigma}{2}\|\tilde{x}_k - x_{k-1}\|^2 \le \frac{1}{2}\|x_* - x_{k-1}\|^2 - \frac{1}{2}\|x_* - x_k\|^2.$$

Summing the above inequality, we have

$$\frac{1-\sigma}{2}\sum_{i=1}^{k}\|\tilde{x}_i - x_{i-1}\|^2 \le \frac{1}{2}\|x_0 - x_*\|^2 - \frac{1}{2}\|x_k - x_*\|^2. \tag{12}$$

Using Lemma 7, we obtain

$$k\min_{1\le i\le k}\theta_i \le \sum_{i=1}^{k}\theta_i \le \sum_{i=1}^{k}\|\tilde{x}_i - x_{i-1}\|^2 \le \frac{\|x_0 - x_*\|^2}{1-\sigma}.$$

Finally, (11) follows from the definition of θ_k in (8) and (10). □

3.4 Ergodic convergence

In this subsection, we are interested in establishing the convergence of the generalized IPP framework based on the averge of iterates, i.e., the ergodic convergence.

Lemma 8. *Assume for* $i = 1, \dots, k$, $u_i \in T^{\varepsilon_i}(y_i)$, $\alpha_i \ge 0$, *and* $\sum_{i=1}^k \alpha_i = 1$. *Let*

$$u^a = \sum_{i=1}^{k}\alpha_i u_i, \quad y^a = \sum_{i=1}^{k}\alpha_i y_i$$

$$\varepsilon^a = \sum_{i=1}^{k}\alpha_i(\varepsilon_i + \langle u_i - u^a, y_i - y^a\rangle).$$

Then, we have

$$u^a \in T^{\varepsilon^a}(y^a), \quad \varepsilon^a \ge 0.$$

Lemma 9. *Define*

$$\Lambda_k = \sum_{i=1}^{k}\lambda_i, \quad \tilde{x}_k^a = \frac{\sum_{i=1}^k \lambda_i \tilde{x}_i}{\Lambda_k}, \quad v_k^a = \frac{\sum_{i=1}^k \lambda_i v_i}{\Lambda_k}, \tag{13}$$

and

$$\varepsilon_k^a = \frac{\sum_{i=1}^k \lambda_i(\varepsilon_i + \langle v_i, \tilde{x}_i - \tilde{x}_k^a\rangle)}{\Lambda_k} = \frac{-\sum_{i=1}^k \lambda_i \Gamma_i(\tilde{x}_k^a)}{\Lambda_k}. \tag{14}$$

Then, we have

$$v_k^a \in T^{\varepsilon_k^a}(\tilde{x}_k^a).$$

Theorem 5. *(**Ergodic convergence**) For every $k \geq 1$, we have*

$$\|v_k^a\| \leq \frac{2\|x_0 - x_*\|}{\Lambda_k}, \quad \varepsilon_k^a \leq \frac{\|\tilde{x}_k^a - x_0\|^2}{2\Lambda_k} \leq \frac{\left(2 + \frac{\sqrt{\sigma}}{\sqrt{1-\sigma}}\right)^2 \|x_0 - x_*\|^2}{2\Lambda_k}.$$

Proof. First, it follows from Lemma 6 with $\delta_k = 0$ and $x = x_*$ that $\|x_k - x_*\| \leq \|x_0 - x_*\|$. By the definition of v_k^a in (13), we have

$$\Lambda_k v_k^a = \sum_{i=1}^{k} \lambda_i v_i = \sum_{i=1}^{k} (x_{i-1} - x_i) = x_0 - x_k,$$

and hence

$$\Lambda_k \|v_k^a\| = \|x_0 - x_k\| \leq \|x_0 - x_*\| + \|x_* - x_k\| \leq 2\|x_0 - x_*\|,$$

which proves the first inequality in the theorem. Using (14) and Lemma 6 with $\delta_k = 0$ and $x = \tilde{x}_k^a$ (after summation), we have

$$\Lambda_k \varepsilon_k^a = -\sum_{i=1}^{k} \lambda_i \Gamma_i(\tilde{x}_k^a) \leq \frac{1}{2}\|\tilde{x}_k^a - x_0\|^2,$$

which proves the first inequality for ε_k^a in the theorem. By (12), we also have

$$\max_{1 \leq i \leq k} \|\tilde{x}_i - x_{i-1}\| \leq \frac{\|x_0 - x_*\|}{\sqrt{1-\sigma}}.$$

Finally, we have

$$\begin{aligned}
\|\tilde{x}_k^a - x_0\| &\leq \max_{1 \leq i \leq k} \|\tilde{x}_i - x_0\| \\
&\leq \max_{1 \leq i \leq k} (\|\tilde{x}_i - x_i\| + \|x_0 - x_i\|) \\
&\leq \max_{1 \leq i \leq k} \|\tilde{x}_i - x_i\| + 2\|x_0 - x_*\| \\
&\leq \max_{1 \leq i \leq k} \sqrt{\sigma}\|\tilde{x}_i - x_{i-1}\| + 2\|x_0 - x_*\| \\
&\leq \frac{\sqrt{\sigma}\|x_0 - x_*\|}{\sqrt{1-\sigma}} + 2\|x_0 - x_*\|,
\end{aligned}$$

where the second last inequality is due to (7). With this inequality, we complete the proof of the inequality of ε_k^a. □

4 Saddle point, Chambolle-Pock

Given $\hat{K} : C \times D \mapsto \mathbb{R}$ closed convex-concave and $C \times D \subseteq \mathbb{R}^n \times \mathbb{R}^m$ convex, define

$$p(x) = \max_{y \in D} \hat{K}(x, y), \quad d(y) = \min_{x \in C} \hat{K}(x, y).$$

Note that we have $p(x) \geq d(y)$ for all $(x,y) \in C \times D$.
Saddle point: $(x_*, y_*) \in C \times D$ s.t. $p(x_*) = d(y_*)$. Finding a saddle point (x_*, y_*) is equivalent to $\min_{x,y} F(x,y) = p(x) - d(y)$. Note that $F(x,y) \geq 0$ and $F(x_*, y_*) = 0$.
ε-saddle point: $(\bar{x}, \bar{y}) \in C \times D$ s.t. $p(\bar{x}) - d(\bar{y}) \leq \varepsilon$ or equivalently

$$0 \in \partial_\varepsilon [\hat{K}(\cdot, y) - \hat{K}(x, \cdot)](\bar{x}, \bar{y}).$$

For $\varepsilon = 0$, this is the problem $0 \in T(z)$ where

$$T(z) = T(x,y) := \partial[\hat{K}(\cdot, y) - \hat{K}(x, \cdot)](x,y).$$

4.1 Smooth composite structure

Assume

$$\hat{K}(x,y) = K(x,y) + g_1(x) - g_2(y).$$

where g_1 is closed and convex over C, g_2 is closed and convex over D, and K is a real-valued function differentiable on $C \times D$ and ∇K is L-Lipschitz. Here,

$$T(z) = \underbrace{\begin{pmatrix} \nabla_x K(x,y) \\ -\nabla_y K(x,y) \end{pmatrix}}_{=F(z)} + \underbrace{\begin{pmatrix} \partial g_1(x) \\ \partial g_2(x) \end{pmatrix}}_{=\partial g(z)}$$

where $g(z) = g(x,y) = g_1(x) + g_2(y)$.

4.2 Chambolle-Pock's method, a.k.a., primal-dual hybrid gradient

Consider the problem

$$\text{(P)} \ \min_x \max_y \langle Kx, y \rangle + G(x) - F^*(y)$$

where G is closed and convex over $\mathbb{R}^n$, F is closed and convex over $\mathbb{R}^m$, and $K : \mathbb{R}^n \mapsto \mathbb{R}^m$ is linear. The problem (P) is equivalent to

$$\min_x F(Kx) + G(x),$$

and has the dual formulation

$$\max_x \min_y \langle Kx, y \rangle + G(x) - F^*(y) = \psi(x,y),$$

or equivalently

$$\max_y -G^*(-K^* y) - F^*(y).$$

Furthermore, let us assume that $\exists\, (x^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$-Kx^* + \partial F^*(y^*) \ni 0, \quad K^* y^* + \partial G(x^*) \ni 0$$

or equivalently

$$(0,0) \in \partial\left[\psi\left(\cdot, y^*\right) - \psi\left(x^*, \cdot\right)\right]\left(x^*, y^*\right).$$

Algorithm 5 Chambolle-Pock method

Input: Initial point $\left(x^0, y^0\right) \in \mathbb{R}^n \times \mathbb{R}^m$, choose τ_1, $\tau_2 > 0$, and $\theta = 1$, and set $\bar{x}^0 = x^0$;

for $k \geq 0$ **do**

Compute

$$y^{k+1} = \left(I + \tau_2 \partial F^*\right)^{-1}\left(y^k + \tau_2 K \bar{x}^k\right),$$
$$x^{k+1} = \left(I + \tau_1 \partial G\right)^{-1}\left(x^k - \tau_1 K^* y^{k+1}\right),$$
$$\bar{x}^{k+1} = x^{k+1} + \theta\left(x^{k+1} - x^k\right).$$

end for

We have

$$\begin{aligned}
&\frac{x^{k+1} - x^k}{\tau_1} + K^* y^{k+1} + \partial G\left(x^{k+1}\right) \ni 0, \\
&\frac{y^{k+1} - y^k}{\tau_2} - K \bar{x}^k + \partial F^*\left(y^{k+1}\right) \ni 0, \\
&\bar{x}^{k+1} = 2x^{k+1} - x^k
\end{aligned} \tag{15}$$

The above algorithm is an instance of the following framework.

Proposition 1. *The Chambolle-Pock method is an instance of the following IPP framework as long as $\|K\|^2 \tau_1 \tau_2 \leq \sigma$. Given (x_k, y_k) and $\lambda_{k+1} > 0$, find (x_{k+1}, y_{k+1}), $(\tilde{x}_{k+1}, \tilde{y}_{k+1})$, and ε_{k+1} s.t.*

$$\frac{x_{k+1} - x_k}{\lambda_{k+1}} + \tau_1\left[K^* \tilde{y}_{k+1} + \partial G\left(\tilde{x}_{k+1}\right)\right] \ni 0, \tag{16}$$

$$\frac{y_{k+1} - y_k}{\lambda_{k+1}} + \tau_2\left[-K \tilde{x}_{k+1} + \partial F^*\left(\tilde{y}_{k+1}\right)\right] \ni 0. \tag{17}$$

We also have, for some $\sigma \in (0,1)$, the inequality

$$\frac{1}{\tau_1}\left\|x_{k+1} - \tilde{x}_{k+1}\right\|^2 + \frac{1}{\tau_2}\left\|y_{k+1} - \tilde{y}_{k+1}\right\|^2 + 2\lambda_{k+1}\varepsilon_{k+1} \leq \sigma\left[\frac{1}{\tau_1}\left\|\tilde{x}_{k+1} - x_k\right\|^2 + \frac{1}{\tau_2}\left\|\tilde{y}_{k+1} - y_k\right\|^2\right]. \tag{18}$$

Note:

$$\langle z_1, z_2 \rangle = \frac{1}{\tau_1}\langle x_1, x_2 \rangle_n + \frac{1}{\tau_2}\langle y_1, y_2 \rangle_m, \quad \|z\|^2 = \frac{1}{\tau_1}\|x\|_n^2 + \frac{1}{\tau_2}\|y\|_m^2.$$

Proof. Superscript: Chambolle-Pock, subscript: IPP framework.

We will show that C-P method is an instance of the IPP framework with $\lambda_{k+1} = 1, \varepsilon_{k+1} = 0$, and

$$
\begin{aligned}
x_{k+1} &= \tilde{x}_{k+1} = x^{k+1}, \quad \tilde{y}_{k+1} = y^{k+1}, \\
y_{k+1} &= y^{k+1} + \tau_2 K \left(\bar{x}^{k+1} - x^{k+1}\right).
\end{aligned}
$$

The proof of (16) is straightforward. The left-hand side of (17) is

$$
\begin{aligned}
& y^{k+1} + \tau_2 K \left(\bar{x}^{k+1} - x^{k+1}\right) - y^k - \tau_2 K \left(\bar{x}^k - x^k\right) \\
& + \tau_2 \left[-K x^{k+1} + \partial F^* \left(y^{k+1}\right)\right] \\
& = \tau_2 \left[\frac{y^{k+1} - y^k}{\tau_2} + K \left(\bar{x}^{k+1} - x^{k+1} - \bar{x}^k + x^k - x^{k+1}\right) + \partial F^* \left(y^{k+1}\right)\right] \\
& = \tau_2 \left[\frac{y^{k+1} - y^k}{\tau_2} - K \bar{x}^k + \partial F^* \left(y^{k+1}\right)\right],
\end{aligned}
$$

which contains 0 in view of (15). We now prove the inequality (18). We observe that

$$
\begin{aligned}
(18) & \iff \frac{1}{\tau_2} \left\|\tau_2 K \left(\bar{x}^{k+1} - x^{k+1}\right)\right\|^2 \leq \sigma \left[\frac{1}{\tau_1} \left\|x^{k+1} - x^k\right\|^2 + \frac{1}{\tau_2} \left\|y^{k+1} - y_k\right\|^2\right] \\
& \iff \left\|K \left(\bar{x}^{k+1} - x^{k+1}\right)\right\|^2 \leq \sigma \left[\frac{1}{\tau_1 \tau_2} \left\|x^{k+1} - x^k\right\|^2 + \frac{1}{\tau_2^2} \left\|y^{k+1} - y_k\right\|^2\right] \\
& \Longleftarrow \|K\|^2 \left\|\bar{x}^{k+1} - x^{k+1}\right\|^2 \leq \frac{\sigma}{\tau_1 \tau_2} \left\|x^{k+1} - x^k\right\|^2 \\
& \iff \|K\|^2 \left\|x^{k+1} - x^k\right\|^2 \leq \frac{\sigma}{\tau_1 \tau_2} \left\|x^{k+1} - x^k\right\|^2 \\
& \iff \|K\|^2 \tau_1 \tau_2 \leq \sigma.
\end{aligned}
$$

□

So the convergence of Chambolle-Pock follows from the convergence (pointwise/ergodic) of the IPP framework.