Learning Bilevel Proximal Programs for Joint Feasibility and Optimality Pursuit (Supplemental Materials)

Paper ID: 920

Introduction

In this paper we consider a optimization problem for many vision and learning tasks which can be formulated as

$$\min_{\mathbf{x}\in\mathbb{R}^n} \Psi(\mathbf{x}) := g(\mathbf{x}) + \psi(\mathbf{x}), \tag{1}$$

where functions g and ψ typically capture the loss of data fitting and the regularization, respectively and g is convex and continuously differentiable, ψ is nonconvex and not necessarily differentiable. The proposed Joint Feasibility and Optimality Pursuit model (JFOP for short) that can be described as

$$\begin{aligned} \texttt{Leader (L)}: \quad & \min_{\mathbf{x}} \Psi(\mathbf{x}) := g(\mathbf{x}) + \psi(\mathbf{x}), \\ \texttt{Follower (F)}: \quad & s.t.\ \mathbf{x} \in \arg\min_{\mathbf{x}} F(\mathbf{x}) := f(\mathbf{x}) + \phi(\mathbf{x}). \end{aligned} \tag{2}$$

(L) in Eq. (2) is just the original objective in Eq. (1). Similar to that in (L) subproblem, f and ϕ in (F) that can be nonconvex also respectively represent the fidelity and prior for the pursuit of feasibility.

Preliminaries

In this section, we introduce necessary definitions, assumption and lemmas (see (Rockafellar and Wets 2009; Borwein and Lewis 2010; Bolte et al. 2010; Attouch et al. 2010)), which will be used in the following analyses.

Definition 1. *Let $\Psi : \mathbb{R}^D \to (-\infty, \infty]$ be a proper and lower semi-continuous function. The definition of subdifferential, proper, lower-semicontinuous and coercive of a function are described as follows:*

1. *The Frechét sub-differential (denoted as $\hat{\partial}\Psi$) of Ψ at point $\mathbf{x} \in \mathtt{dom}(\Psi)$ is the set of all vectors $\mathbf{z}$ which satisfies*
$$\lim_{\mathbf{y}\neq\mathbf{x}} \inf_{\mathbf{y}\to\mathbf{x}} \frac{\Psi(\mathbf{y}) - \Psi(\mathbf{x}) - \langle \mathbf{z}, \mathbf{y}-\mathbf{x}\rangle}{\|\mathbf{y}-\mathbf{x}\|} \geq 0, \tag{3}$$
where $\langle \cdot,\cdot \rangle$ denotes the inner product. Then the limiting Frechét sub-differential (denote as $\partial\Psi$) at $\mathbf{x} \in \mathtt{dom}(\Psi)$ is the following closure of $\hat{\partial}\Psi$:
$$\begin{aligned} \partial\Psi(\mathbf{x}) := \{\mathbf{z} \in \mathbb{R}^n : \exists \mathbf{x}^k \to \mathbf{x}, \Psi(\mathbf{x}^k) \to \Psi(\mathbf{x}), \\ \mathbf{z}^k \in \hat{\partial}\Psi(\mathbf{x}^k) \to \mathbf{z}, k \to \infty\}. \end{aligned}$$
Further, $\mathbf{x}$ is a critical point *of Ψ if $0 \in \partial\Psi(\mathbf{x})$.*
2. *A function $\Psi : \mathbb{R}^n \to (-\infty, +\infty]$ is said to be proper and lower semi-continuous if $\mathtt{dom}(\Psi) \neq \emptyset$, where $\mathtt{dom}(\Psi) := \{\mathbf{x} \in \mathbb{R}^n : \Psi(\mathbf{x}) < +\infty\}$ and $\liminf_{\mathbf{x}\to\mathbf{y}} \Psi(\mathbf{x}) \geq \Psi(\mathbf{y})$ at any point $\mathbf{y} \in \mathtt{dom}(\Psi)$.*
3. *A function Ψ is said to be coercive, if Ψ is bounded from below and $\Psi \to \infty$ as $\|\mathbf{x}\| \to \infty$.*
4. *A subset Ω of $\mathbb{R}^n$ is a real semi-algebraic set if there exist a finit number of real polynomial functions $g_{ij}, h_{ij} : \mathbb{R}^n \to \mathbb{R}$ such that*
$$\Omega = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{\mathbf{x} \in \mathbb{R}^n : g_{ij}(\mathbf{x}) = 0 \text{ and } h_{ij}(\mathbf{x}) < 0\}.$$
A function Ψ is called semi-algebraic if its graph $\{(\mathbf{x}, z) \in \mathbb{R}^{n+1} : \Psi(\mathbf{x}) = z\}$ is a semi-algebraic subset of $\mathbb{R}^{n+1}$.

Definition 2. *(Kurdyka-Łojasiewicz property): Ψ is said to have the Kurdyka-Łojasiewicz property at $\bar{\mathbf{x}} \in \mathtt{dom}\partial\Psi := \{\mathbf{x} \in \mathbb{R}^n : \partial\Psi(\mathbf{x}) \neq \emptyset\}$ if there exist $\xi \in (0, \infty]$), a neighborhood $\mathcal{N}_{\bar{\mathbf{x}}}$ of $\bar{\mathbf{x}}$ and a desingularizing function $\varphi : [0, \xi) \to \mathbb{R}_+$ which satisfies (1) $\varphi(0) = 0$; (2) φ is C^1 on $(0, \xi)$ and continuous at 0; (3) for all $s \in (0, \xi) : \varphi'(s) > 0$, such that for all*

$$\mathbf{x} \in \mathcal{N}_{\bar{\mathbf{x}}} \cap [\Psi(\bar{\mathbf{x}}) < \Psi(\mathbf{x}) < \Psi(\bar{\mathbf{x}}) + \xi],$$

the following inequality holds

$$\varphi'(\Psi(\mathbf{x}) - \Psi(\bar{\mathbf{x}}))\mathtt{dist}(0, \partial\Psi(\mathbf{x})) \geq 1.$$

Assumption 1. *The object function in Eq. (1) should satisfy:*

1. *The function Ψ is proper, lower-semicontinuous and coercive function.*
2. *$g(\mathbf{x})$ is Lipschitz smooth, i.e., for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^D$, we have*
$$\|\nabla g(\mathbf{x}) - \nabla g(\mathbf{y})\| \leq L^g \|\mathbf{x} - \mathbf{y}\|,$$
where L^g is the Lipschitz constant for ∇g.

The Summary of Key Equations and Algorithms

BPP

In general, the leader subproblem in Eq. (2) is updated by the standard Proximal Gradient (PG) scheme

$$\mathcal{L}_s(\mathbf{x}) := \mathtt{prox}_{s\psi}(\mathbf{x} - s\nabla g(\mathbf{x})),$$

Copyright © 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

where $\texttt{prox}_{s\psi}(\mathrm{x}) = \arg\min_{\mathbf{y}} \left\{\psi(\mathbf{y}) + \frac{1}{2s}\|\mathbf{y}-\mathbf{x}\|^2\right\}$. Denote the updating of the follower subproblem at k-th iteration as $\mathcal{F}(\mathbf{x}^k)$. Then the formal updating rule in Eq. (2) is

$$\mathbf{x}^{k+1} = \alpha^k \mathcal{F}(\mathbf{x}^k) + (1-\alpha^k)\mathcal{L}_s(\mathbf{x}^k),$$

where $\{\alpha^k\}$ is a sequence of real numbers in the range $[0, 1)$ and will be analyzed and determined later. We summarize our complete iteration in Alg. 1.

Algorithm 1 Bilevel Proximal Programs

Require: The input $\mathbf{x}^0$, parameters $s \in (0, 1/L^g)$, $t \in (0, 1/L^f)$ and $\left\{\alpha^k | \alpha^k \in [0, 1)\right\}$.

1: **while** not converged **do**
2: $\quad \mathbf{f}^{k+1} = \mathcal{F}(\mathbf{x}^k)$ and $\mathbf{l}^{k+1} = \mathcal{L}_s(\mathbf{x}^k)$.
3: $\quad \mathbf{z}^{k+1} = \alpha^k \mathbf{f}^{k+1} + (1-\alpha^k)\mathbf{l}^{k+1}$.
4: $\quad$ **if** $\Psi(\mathbf{z}^{k+1}) \leq \Psi(\mathbf{l}^{k+1})$ **then**
5: $\quad\quad \mathbf{x}^{k+1} = \mathbf{z}^{k+1}$.
6: $\quad$ **else**
7: $\quad\quad \mathbf{x}^{k+1} = \mathbf{l}^{k+1}$.
8: $\quad$ **end if**
9: **end while**

LBPP

By further considering

$$F_\mu^k(\mathbf{x}) := F(\mathbf{x}) + \frac{\mu}{2}\|\mathbf{x} - \tilde{\mathbf{x}}^k\|^2, \tag{4}$$

as the proximal approximation of F (with parameter $\mu > 0$) and $\mathcal{D}(\cdot; \boldsymbol{\theta}^k)$ as the learning-based iterative building-block at k-th iteration, we actually update the optimization problem by

$$\mathcal{F}_t^{\mathcal{D}}(\tilde{\mathbf{x}}^k) = \texttt{prox}_{s\phi}\left(\tilde{\mathbf{x}}^k - s(\nabla f(\tilde{\mathbf{x}}^k) + \mu(\tilde{\mathbf{x}}^k - \mathbf{x}^k))\right), \tag{5}$$

where $\tilde{\mathbf{x}}^k = \mathcal{D}(\mathbf{x}^k; \boldsymbol{\theta}^k)$. Then we summarize the complete DBPP algorithm in Alg. 2.

Algorithm 2 Learning-based BPP

Require: The input $\mathbf{x}^0$, parameters $s \in (0, 1/L^g)$, $t \in (0, 1/L^f)$, $\{\alpha^k | \alpha^k \in [0, 1)\}$ and $C > 0$.

while not converged **do**
$\quad \mathbf{l}^{k+1} = \mathcal{L}_s(\mathbf{x}^k)$.
$\quad \mathbf{f}^{k+1} = \mathcal{F}_t^{\mathcal{D}}(\tilde{\mathbf{x}}^k)$ where $\tilde{\mathbf{x}}^k = \mathcal{D}(\mathbf{x}^k; \boldsymbol{\theta}^k)$.
$\quad$ **if** $\|\tilde{\mathbf{x}}^k - \mathbf{x}^k\| \leq C\|\mathbf{f}^{k+1} - \mathbf{x}^k\|$ **then**
$\quad\quad \mathbf{z}^{k+1} = \alpha^k \mathbf{f}^{k+1} + (1-\alpha^k)\mathbf{l}^{k+1}$.
$\quad$ **else**
$\quad\quad \mathbf{f}^{k+1} = \mathcal{F}_t(\mathbf{x}^k)$.
$\quad\quad \mathbf{z}^{k+1} = \alpha^k \mathbf{f}^{k+1} + (1-\alpha^k)\mathbf{l}^{k+1}$
$\quad$ **end if**
$\quad$ **if** $\Psi(\mathbf{z}^{k+1}) \leq \Psi(\mathbf{l}^{k+1})$ **then**
$\quad\quad \mathbf{x}^{k+1} = \mathbf{z}^{k+1}$.
$\quad$ **else**
$\quad\quad \mathbf{x}^{k+1} = \mathbf{l}^{k+1}$.
$\quad$ **end if**
end while

Theoretical Results for BPP

The following theorem summarizes the convergence properties of BPP.

Theorem 1. *Suppose that $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ be an iteration sequence generated by Alg. 1 and the Assumption 1 holds. Then, the iterations sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ is bounded and there exists subsequences $\{\mathbf{x}^{k_q}\}_{q\in\mathbb{N}}$, such that it converges to the critical point (denoted as $\mathbf{x}^*$) of the minimization problem* (1).

Proof. Firstly, the boundness of iteration sequence $\{\mathbf{x}^k\}$ is obvious according to the Assumption 1. Step 2 in Alg. 1 shows that

$$\psi(\mathbf{l}^{k+1}) + \left\langle \mathbf{l}^{k+1} - \mathbf{x}^k, \nabla g(\mathbf{x}^k)\right\rangle + \frac{1}{2s}\|\mathbf{l}^{k+1} - \mathbf{x}^k\|^2 \leq \psi(\mathbf{x}^k),$$

and this with the following inequality

$$g(\mathbf{l}^{k+1}) \leq g(\mathbf{x}^k) + \left\langle \mathbf{l}^{k+1} - \mathbf{x}^k, \nabla g(\mathbf{x}^k)\right\rangle + \frac{L^g}{2}\|\mathbf{l}^{k+1} - \mathbf{x}^k\|^2,$$

yields

$$\Psi(\mathbf{l}^{k+1}) \leq \Psi(\mathbf{x}^k) - (\frac{1}{2s} - \frac{L^g}{2})\|\mathbf{l}^{k+1} - \mathbf{x}^k\|^2.$$

If $\Psi(\mathbf{z}^{k+1}) \leq \Psi(\mathbf{l}^{k+1})$ then we have $\mathbf{x}^{k+1} = \mathbf{z}^{k+1}$ and

$$\Psi(\mathbf{x}^{k+1}) = \Psi(\mathbf{z}^{k+1}) \leq \Psi(\mathbf{l}^{k+1}),$$

which implies that

$$\Psi(\mathbf{x}^{k+1}) \leq \Psi(\mathbf{x}^k) - (\frac{1}{2s} - \frac{L^g}{2})\|\mathbf{l}^{k+1} - \mathbf{x}^k\|^2, \tag{6}$$

else if $\Psi(\mathbf{z}^{k+1}) > \Psi(\mathbf{l}^{k+1})$, we have $\mathbf{x}^{k+1} = \mathbf{l}^{k+1}$. This yields

$$\Psi(\mathbf{x}^{k+1}) = \Psi(\mathbf{l}^{k+1}) \leq \Psi(\mathbf{x}^k) - (\frac{1}{2s} - \frac{L^g}{2})\|\mathbf{l}^{k+1} - \mathbf{x}^k\|^2. \tag{7}$$

Summing (6) and (7) yields

$$\sum_{k=1}^{\infty}\|\mathbf{l}^{k+1} - \mathbf{x}^k\|^2 < \infty,$$

which means that

$$\|\mathbf{l}^{k+1} - \mathbf{x}^k\| \to 0, \ k \to \infty, \tag{8}$$

i.e., there exist subsequence $\{\mathbf{l}^{k_p+1}\}$ and $\{\mathbf{x}^{k_p}\}$ convergence to a same point $\mathbf{x}^*$ as $p \to \infty$. Then, it is necessary to consider the optimal conditions,

$$0 \in \partial\psi(\mathbf{l}^{k+1}) + \nabla g(\mathbf{x}^k) + \frac{1}{s}(\mathbf{l}^{k+1} - \mathbf{x}^k), \tag{9}$$

i.e.,

$$-\frac{1}{s}(\mathbf{l}^{k+1} - \mathbf{x}^k) - \nabla g(\mathbf{x}^k) + \nabla g(\mathbf{l}^{k+1}) \in \partial\Psi(\mathbf{l}^{k+1}).$$

So we have

$$\begin{aligned} &\|\frac{1}{s}(\mathbf{l}^{k+1} - \mathbf{x}^k) + \nabla g(\mathbf{x}^k) - \nabla g(\mathbf{l}^{k+1})\| \\ \leq &(\frac{1}{s} + L^g)\|\mathbf{l}^{k+1} - \mathbf{x}^k\| \to 0, \text{ as } k \to \infty. \end{aligned} \tag{10}$$

Since Ψ is lower-semicontinuous, we obtain that $\liminf_{p\to\infty} \Psi(\mathbf{x}^{k_p}) \geq \Psi(\mathbf{x}^*)$. From the iterative step 2, we have that for all integer k

$$\begin{aligned}&\big\langle \mathbf{l}^{k+1}-\mathbf{x}^k, \nabla g(\mathbf{x}^k)\big\rangle + \frac{1}{2s}\|\mathbf{l}^{k+1}-\mathbf{x}^k\|^2 + \psi(\mathbf{l}^{k+1})\\ \leq&\big\langle \mathbf{x}^*-\mathbf{x}^k, \nabla g(\mathbf{x}^k)\big\rangle + \frac{1}{2s}\|\mathbf{x}^*-\mathbf{x}^k\|^2 + \psi(\mathbf{x}^*).\end{aligned}$$

Let $k=k_p$ and $p\to\infty$, then we obtain

$$\begin{aligned}&\limsup_{p\to\infty} \psi(\mathbf{l}^{k_p+1})\\ &\leq \limsup_{p\to\infty}\left(\big\langle \mathbf{x}^*-\mathbf{x}^{k_p}, \nabla g(\mathbf{x}^{k_p})\big\rangle + \tfrac{1}{2s}\|\mathbf{x}^*-\mathbf{x}^{k_p}\|^2\right)\\ &+\psi(\mathbf{x}^*).\end{aligned}$$

Hence, the above inequality reduces to

$$\limsup_{p\to\infty} \psi(\mathbf{x}^{k_p}) \leq \psi(\mathbf{x}^*).$$

Obviously, we have

$$\lim_{p\to\infty} \Psi(\mathbf{x}^{k_p}) = \Psi(\mathbf{x}^*).$$

Further, Eq.(10) implies that $0\in\partial\Psi(\mathbf{x}^*)$, i.e., $\mathbf{x}^*$ is a critical point. This complete the proof. □

The finite length sequence of $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ stated in Remark 2 will be proofed as follows.

Proof. With the KL property (see (Bolte, Sabach, and Teboulle 2014)), we have

$$\varphi'(\Psi(\mathbf{l}^{k+1})-\Psi(\mathbf{x}^*))\mathtt{dist}(0,\partial\Psi(\mathbf{l}^{k+1}))\geq 1,$$

where φ is the desingularizing function. Then, with Eq.(10), the above inequality can be described as

$$\varphi'(\Psi(\mathbf{l}^{k+1})-\Psi(\mathbf{x}^*)) \geq \frac{s}{1+sL^g}\|\mathbf{l}^{k+1}-\mathbf{x}^k\|^{-1}.$$

On the other hand, from the concavity of φ we obtain

$$\varphi'(\Psi(\mathbf{x}^{k+1})-\Psi(\mathbf{x}^*)) \geq \varphi'(\Psi(\mathbf{l}^{k+1})-\Psi(\mathbf{x}^*)),$$

and

$$\begin{aligned}&\varphi\big(\Psi(\mathbf{x}^{k+1})-\Psi(\mathbf{x}^*)\big)-\varphi\big(\Psi(\mathbf{x}^{k+2})-\Psi(\mathbf{x}^*)\big)\\ \geq&\varphi'(\Psi(\mathbf{x}^{k+1})-\Psi(\mathbf{x}^*))\big(\Psi(\mathbf{x}^{k+1}-\Psi(\mathbf{x}^{k+2}))\big)\\ \geq&\frac{s}{1+sL^g}\frac{1}{\|\mathbf{l}^{k+1}-\mathbf{x}^k\|}\cdot\frac{1-sL^g}{2s}\|\mathbf{l}^{k+2}-\mathbf{x}^{k+1}\|^2.\end{aligned}$$

If we denote

$$\hat{\triangle}_{k,k+1} := \varphi(\Psi(\mathbf{x}^k)-\varphi(\Psi(\mathbf{x}^*))-\varphi(\Psi(\mathbf{x}^{k+1})-\Psi(\mathbf{x}^*)),$$

the following inequality holds,

$$\|\mathbf{l}^{k+2}-\mathbf{x}^{k+1}\|^2 \leq \hat{C}\hat{\triangle}_{k+1,k+2}\|\mathbf{l}^{k+1}-\mathbf{x}^k\|, \tag{11}$$

where $\hat{C}=\frac{1-sL^g}{2(1+sL^g)}$. Inequality (11) implies that

$$2\|\mathbf{l}^{k+2}-\mathbf{x}^{k+1}\| \leq \|\mathbf{l}^{k+1}-\mathbf{x}^k\| + \hat{C}\hat{\triangle}_{k+1,k+2}.$$

Then, we have

$$\begin{aligned}2\sum_{i=l+1}^{k}\|\mathbf{l}^{i+2}-\mathbf{x}^{i+1}\| \leq &\sum_{i=l+1}^{k}\|\mathbf{l}^{i+2}-\mathbf{x}^{i+1}\|\\ &+\|\mathbf{l}^{i+2}-\mathbf{x}^{i+1}\| + \hat{C}\hat{\triangle}_{l+1,k+2}.\end{aligned}$$

Obviously, this shows that the sequence $\{\mathbf{l}^{k+1}-\mathbf{x}^k\}$ has finite length, i.e.,

$$\sum_{k=1}^{\infty}\|\mathbf{l}^{k+1}-\mathbf{x}^k\| = \sum_{k=1}^{\infty}\|\mathbf{l}^{k+2}-\mathbf{x}^{k+1}\| + \|\mathbf{l}^2-\mathbf{x}^1\| < \infty. \tag{12}$$

If $\mathbf{x}^{k+1}=\mathbf{l}^{k+1}$, it is easily to show that sequence $\{\mathbf{x}^{k+1}-\mathbf{x}^k\}$ has finite length. If $\mathbf{x}^{k+1}=\alpha^k\mathbf{f}^{k+1}+(1-\alpha^k)\mathbf{l}^{k+1}$, we have

$$\|\mathbf{x}^{k+1}-\mathbf{x}^k\| \leq \alpha^k\|\mathbf{f}^{k+1}-\mathbf{x}^k\| + (1-\alpha^k)\|\mathbf{l}^{k+1}-\mathbf{x}^k\|.$$

As F is proper, lower-semicontinuous and coercive function, the iterative sequence $\{\mathbf{f}^{k+1}\}$ is bounded. Then we have that

$$\begin{aligned}\sum_{k=1}^{\infty}\|\mathbf{x}^{k+1}-\mathbf{x}^k\| \leq &\sum_{k=1}^{\infty}\alpha^k\|\mathbf{f}^{k+1}-\mathbf{x}^k\|\\ &+\sum_{k=1}^{\infty}(1-\alpha^k)\|\mathbf{l}^{k+1}-\mathbf{x}^k\|\\ &<\infty,\end{aligned} \tag{13}$$

where $\alpha^k\in(0,1]$ and the second inequality follows form (12) and the assumption of α, i.e., $\sum_{k=1}^{\infty}\alpha^k<\infty$. This completes the proof. □

Theoretical Results for LBPP

Corollary 1. *Suppose that the Assumption 1 holds. Then the iterations sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ generated by Alg. 2 is bounded and there exists subsequences $\{\mathbf{x}^{k_p}\}_{p\in\mathbb{N}}$ convergence to the critical point of the problem* (1).

Proof. In this proof, we just need to verify the boundness of the iterations sequence $\{\mathbf{f}^{k+1}\}$ and the other steps are similarily to Theorem 1.

According to step 3 in Alg. 2, we have

$$\begin{aligned}\mathbf{f}^{k+1} =&\mathtt{prox}\left(\tilde{\mathbf{x}}^k-(t\nabla f(\tilde{\mathbf{x}}^k)+\mu(\tilde{\mathbf{x}}^k-\mathbf{x}^k))\right)\\ =&\arg\min_x\left\{\frac{1}{2t}\|\mathbf{x}-(\tilde{\mathbf{x}}^k-t\nabla f(\tilde{\mathbf{x}}^k)-t\mu(\tilde{\mathbf{x}}^k-\mathbf{x}^k))\|^2\right.\\ &\left.+\phi(\mathbf{x})\right\}\\ =&\arg\min_x\left\{\frac{1}{t}\big\langle\mathbf{x}-\mathbf{x}^k, t\nabla f(\tilde{\mathbf{x}}^k)-(1-t\mu)(\tilde{\mathbf{x}}^k-\mathbf{x}^k)\big\rangle\right.\\ &\left.+\frac{1}{2t}\|\mathbf{x}-\mathbf{x}^k\|^2+\phi(\mathbf{x})\right\}.\end{aligned}$$

The above implies that

$$\begin{aligned}\phi(\mathbf{f}^{k+1})+\frac{1}{t}\left\langle\mathbf{f}^{k+1}-\mathbf{x}^k, t\nabla f(\tilde{\mathbf{x}}^k)-(1-t\mu)(\tilde{\mathbf{x}}^k-\mathbf{x}^k)\right\rangle&\\ +\frac{1}{2t}\|\mathbf{f}^{k+1}-\mathbf{x}^k\|^2\leq\phi(\mathbf{x}^k),&\end{aligned}$$

i.e.,

$$\begin{aligned}\phi(\mathbf{f}^{k+1}) \leq& \phi(\mathbf{x}^k) - \frac{1}{2t}\|\mathbf{f}^{k+1} - \mathbf{x}^k\|^2 \\ & - \big\langle \mathbf{f}^{k+1} - \mathbf{x}^k, \nabla f(\tilde{\mathbf{x}}^k) - \frac{1-s\mu}{t}(\tilde{\mathbf{x}}^k - \mathbf{x}^k)\big\rangle.\end{aligned}$$

With Assumption 1, we have

$$f(\mathbf{f}^{k+1}) \leq f(\mathbf{x}^k) + \big\langle \mathbf{f}^{k+1} - \mathbf{x}^k, \nabla f(\mathbf{x}^k)\big\rangle + \frac{L^f}{2}\|\mathbf{f}^{k+1} - \mathbf{x}^k\|^2.$$

Combining the above two inequalities, we obtain

$$\begin{aligned}& F(\mathbf{f}^{k+1}) \\ \leq& F(\mathbf{x}^k) - \left(\frac{1}{2s} - \frac{L^f}{2}\right)\|\mathbf{f}^{k+1} - \mathbf{x}^k\|^2 \\ & + \left\langle \mathbf{f}^{k+1} - \mathbf{x}^k, \nabla f(\mathbf{x}^k) - \nabla f(\tilde{\mathbf{x}}^k) + \frac{1-t\mu}{t}(\tilde{\mathbf{x}}^k - \mathbf{x}^k)\right\rangle \\ \leq& F(\mathbf{x}^k) - \left(\frac{1}{2t} - \frac{L^f}{2}\right)\|\mathbf{f}^{k+1} - \mathbf{x}^k\|^2 + \varepsilon^k,\end{aligned}$$

where $\varepsilon^k = \frac{1-t\mu+tL^f}{t}\|\tilde{\mathbf{x}}^k - \mathbf{x}^k\|\|\mathbf{f}^{k+1} - \mathbf{x}^k\|$. This means that

$$\|\mathbf{f}^{k+1} - \mathbf{x}^k\|^2 \leq \frac{1}{M}\left(F(\mathbf{x}^k) - F(\mathbf{f}^{k+1}) + \varepsilon^k\right),$$

where $M = \frac{1}{2t} - \frac{L^f}{2}$. Then with the condition $\|\tilde{\mathbf{x}}^k - \mathbf{x}^k\| \leq C\|\mathbf{f}^{k+1} - \mathbf{x}^k\|$, we have $\|\mathbf{f}^{k+1} - \mathbf{x}^k\|^2 \leq \bar{C}\left(F(\mathbf{x}^k) - F(\mathbf{l}^{k+1})\right)$, where $\bar{C} = t/\left(t(1 - L^f + \mu) - 1\right)$. Obviously, it is bounded. This completes the proof. $\square$

References

[Attouch et al. 2010] Attouch, H.; Bolte, J.; Redont, P.; and Soubeyran, A. 2010. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality. *Mathematics of Operations Research* 35(2):438–457.

[Bolte et al. 2010] Bolte, J.; Daniilidis, A.; Ley, O.; and Mazet, L. 2010. Characterizations of łojasiewicz inequalities: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society* 362(6):3319–3363.

[Bolte, Sabach, and Teboulle 2014] Bolte, J.; Sabach, S.; and Teboulle, M. 2014. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming* 146(1-2):459–494.

[Borwein and Lewis 2010] Borwein, J., and Lewis, A. S. 2010. *Convex analysis and nonlinear optimization: theory and examples*. Springer Science & Business Media.

[Rockafellar and Wets 2009] Rockafellar, R. T., and Wets, R. J.-B. 2009. *Variational analysis*, volume 317. Springer Science & Business Media.