Supplemental Material for
A Generic First-Order Algorithmic Framework for Bi-Level Programming Beyond Lower-Level Singleton

Risheng Liu [1 2] **Pan Mu** [1 2] **Xiaoming Yuan** [3] **Shangzhi Zeng** [3] **Jin Zhang** [4]

This supplemental material is organized as follows. In Section 1, we present all detailed proofs of our theories. Section 2 then proves some extended theoretical results, including the local convergence behaviors of BDA and the algorithmic scheme and convergence properties of BDA for BLPs with nonsmooth LL objective.

1. Detailed Proofs

1.1. Proof of Theorem 1

Basically, our proof method consists of two main steps:

(1) **LL solution set property:** For any $\epsilon > 0$, there exists $k(\epsilon) > 0$ such that whenever $K > k(\epsilon)$,

$$\sup_{\mathbf{x} \in \mathcal{X}} \mathtt{dist}(\mathbf{y}_K(\mathbf{x}), \mathcal{S}(\mathbf{x})) \le \epsilon.$$

(2) **UL objective convergence property:** $\varphi(\mathbf{x})$ is LSC on $\mathcal{X}$, and for each $\mathbf{x} \in \mathcal{X}$,

$$\lim_{K \to \infty} \varphi_K(\mathbf{x}) \to \varphi(\mathbf{x}).$$

Theorem 1. *Suppose both the above LL solution set and UL objective convergence properties hold. Then we have*

(1) *suppose $\mathbf{x}_K \in \arg\min_{\mathbf{x} \in \mathcal{X}} \varphi_K(\mathbf{x})$, then any limit point $\bar{\mathbf{x}}$ of the sequence $\{\mathbf{x}_K\}$ satisfies that $\bar{\mathbf{x}} \in \arg\min_{\mathbf{x} \in \mathcal{X}} \varphi(\mathbf{x})$.*

(2) *$\inf_{\mathbf{x} \in \mathcal{X}} \varphi_K(\mathbf{x}) \to \inf_{\mathbf{x} \in \mathcal{X}} \varphi(\mathbf{x})$ as $K \to \infty$.*

---
[1]DUT-RU International School of Information Science and Engineering, Dalian University of Technology. [2]Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province. [3]Department of Mathematics, The University of Hong Kong. [4]SUSTech International Center for Mathematics and Department of Mathematics, Southern University of Science and Technology. Correspondence to: Jin Zhang <zhangj9@sustech.edu.cn>.

Proof. Since $\mathcal{X}$ is compact, we can assume without loss of generality that $\mathbf{x}_K \to \bar{\mathbf{x}} \in \mathcal{X}$. For any $\epsilon > 0$, there exists $k(\epsilon) > 0$ such that whenever $K > k(\epsilon)$, we have

$$\sup_{\mathbf{x} \in \mathcal{X}} \mathtt{dist}(\mathbf{y}_K(\mathbf{x}), \mathcal{S}(\mathbf{x})) \le \frac{\epsilon}{2L_0}.$$

Thus, for any $\mathbf{x} \in \mathcal{X}$, there exists $\mathbf{y}^*(\mathbf{x}) \in \mathcal{S}(\mathbf{x})$ such that

$$\|\mathbf{y}_K(\mathbf{x}) - \mathbf{y}^*(\mathbf{x})\| \le \frac{\epsilon}{L_0}.$$

Therefore, for any $\mathbf{x} \in \mathcal{X}$, we have

$$\begin{aligned}
\varphi(\mathbf{x}) &= \inf_{\mathbf{y} \in \mathcal{S}(\mathbf{x})} F(\mathbf{x}, \mathbf{y}) \\
&\le F(\mathbf{x}, \mathbf{y}^*(\mathbf{x})) \\
&\le F(\mathbf{x}, \mathbf{y}_K(\mathbf{x})) + L_0 \|\mathbf{y}_K(\mathbf{x}) - \mathbf{y}^*(\mathbf{x})\| \\
&\le \varphi_K(\mathbf{x}) + \epsilon.
\end{aligned}$$

This implies that, for any $\epsilon > 0$, there exists $k(\epsilon) > 0$ such that whenever $K > k(\epsilon)$, it holds

$$\varphi(\mathbf{x}_K) \le \varphi_K(\mathbf{x}_K) + \epsilon \le \varphi_K(\mathbf{x}) + \epsilon, \quad \forall \mathbf{x} \in \mathcal{X}.$$

Taking $K \to \infty$ and by the LSC of φ, we have

$$\begin{aligned}
\varphi(\bar{\mathbf{x}}) &\le \liminf_{K \to \infty} \varphi(\mathbf{x}_K) \\
&\le \liminf_{K \to \infty} \varphi_K(\mathbf{x}_K) + \epsilon \\
&\le \lim_{K \to \infty} \varphi_K(\mathbf{x}) + \epsilon = \varphi(\mathbf{x}) + \epsilon, \quad \forall \mathbf{x} \in \mathcal{X}.
\end{aligned}$$

By taking $\epsilon \to 0$, we have

$$\varphi(\bar{\mathbf{x}}) \le \varphi(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X},$$

which implies $\bar{\mathbf{x}} \in \arg\min_{\mathbf{x} \in \mathcal{X}} \varphi(\mathbf{x})$.

We next show that $\inf_{\mathbf{x} \in \mathcal{X}} \varphi_K(\mathbf{x}) \to \inf_{\mathbf{x} \in \mathcal{X}} \varphi(\mathbf{x})$ as $K \to \infty$. If this is not true, then there exist $\delta > 0$ and sequence $\{l\} \subseteq \mathbb{N}$ such that

$$\left| \inf_{\mathbf{x} \in \mathcal{X}} \varphi_l(\mathbf{x}) - \inf_{\mathbf{x} \in \mathcal{X}} \varphi(\mathbf{x}) \right| > \delta, \quad \forall l. \tag{1}$$

For each l, there exists $\mathbf{x}_l \in \mathcal{X}$ such that

$$\varphi_l(\mathbf{x}_l) \le \inf_{\mathbf{x} \in \mathcal{X}} \varphi_l(\mathbf{x}) + \delta/2.$$

Since $\mathcal{X}$ is compact, we can assume without loss of generality that $\mathbf{x}_l \to \tilde{\mathbf{x}} \in \mathcal{X}$. For any $\epsilon > 0$, there exists $k(\epsilon) > 0$ such that whenever $l > k(\epsilon)$, the following holds

$$\varphi(\mathbf{x}_l) \leq \varphi_l(\mathbf{x}_l) + \epsilon$$
$$\leq \inf_{\mathbf{x} \in \mathcal{X}} \varphi_l(\mathbf{x}) + \delta/2 + \epsilon$$
$$\leq \varphi_l(\mathbf{x}) + \delta/2 + \epsilon, \quad \forall \mathbf{x} \in \mathcal{X}.$$

By taking $l \to \infty$ and with the LSC of φ, we have

$$\varphi(\tilde{\mathbf{x}}) \leq \liminf_{l \to \infty} \varphi(\mathbf{x}_l)$$
$$\leq \liminf_{l \to \infty} \left(\inf_{\mathbf{x} \in \mathcal{X}} \varphi_l(\mathbf{x}) \right) + \delta/2 + \epsilon$$
$$\leq \limsup_{l \to \infty} \left(\inf_{\mathbf{x} \in \mathcal{X}} \varphi_l(\mathbf{x}) \right) + \delta/2 + \epsilon$$
$$\leq \varphi(\mathbf{x}) + \delta/2 + \epsilon, \quad \forall \mathbf{x} \in \mathcal{X}.$$

Then, by taking $\epsilon \to 0$, we have

$$\inf_{\mathbf{x} \in \mathcal{X}} \varphi(\mathbf{x}) \leq \liminf_{l \to \infty} \left(\inf_{\mathbf{x} \in \mathcal{X}} \varphi_l(\mathbf{x}) \right) + \delta/2$$
$$\leq \limsup_{l \to \infty} \left(\inf_{\mathbf{x} \in \mathcal{X}} \varphi_l(\mathbf{x}) \right) + \delta/2$$
$$\leq \inf_{\mathbf{x} \in \mathcal{X}} \varphi(\mathbf{x}) + \delta/2,$$

which implies a contradiction to Eq. (1). Thus we have $\inf_{\mathbf{x} \in \mathcal{X}} \varphi_K(\mathbf{x}) \to \inf_{\mathbf{x} \in \mathcal{X}} \varphi(\mathbf{x})$ as $K \to \infty$. $\qquad \square$

1.2. Proof of Theorem 2

Lemma 1. *Suppose $F(\mathbf{x}, \mathbf{y})$ is level-bounded in $\mathbf{y}$ locally uniformly in $\mathbf{x} \in \mathcal{X}$. If $\mathcal{S}(\mathbf{x})$ is ISC on $\mathcal{X}$, then $\cup_{\mathbf{x} \in \mathcal{X}} \tilde{\mathcal{S}}(\mathbf{x})$ is bounded.*

Proof. We prove this result by providing a contradiction, that is, we have $\{\mathbf{x}^t\} \subseteq \mathcal{X}$ and $\mathbf{y}^t \in \tilde{\mathcal{S}}(\mathbf{x}^t)$ such that $\|\mathbf{y}^t\| \to +\infty$. As $\mathcal{X}$ is compact, we can assume without loss of generality that $\mathbf{x}^t \to \bar{\mathbf{x}} \in \mathcal{X}$. Since $F(\mathbf{x}, \mathbf{y})$ is level-bounded in $\mathbf{y}$ locally uniformly in $\mathbf{x} \in \mathcal{X}$, we must have $\varphi(\mathbf{x}^t) = F(\mathbf{x}^t, \mathbf{y}^t) \to +\infty$. On the other hand, for any $\epsilon > 0$, let $\bar{\mathbf{y}} \in \mathcal{S}(\bar{\mathbf{x}})$ satisfy $F(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \leq \varphi(\bar{\mathbf{x}}) + \epsilon$. As F is continuous at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, there exists $\delta_0 > 0$ such that

$$F(\mathbf{x}, \mathbf{y}) \leq F(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \epsilon, \ \forall (\mathbf{x}, \mathbf{y}) \in \mathbb{B}_{\delta_0}(\bar{\mathbf{x}}, \bar{\mathbf{y}}).$$

As $\mathcal{S}(\mathbf{x})$ is ISC at $\bar{\mathbf{x}}$ relative to $\mathcal{X}$, then it follows that there exists $\frac{\sqrt{2}}{2} \delta_0 \geq \delta > 0$ satisfying

$$\mathcal{S}(\mathbf{x}) \cap \mathbb{B}_{\frac{\sqrt{2}}{2} \delta_0}(\bar{\mathbf{y}}) \neq \varnothing, \ \forall \mathbf{x} \in \mathbb{B}_\delta(\bar{\mathbf{x}}) \cap \mathcal{X}.$$

Therefore, for any $\mathbf{x} \in \mathbb{B}_\delta(\bar{\mathbf{x}}) \cap \mathcal{X}$, there exists $\mathbf{y} \in \mathcal{S}(\mathbf{x})$ satisfying $(\mathbf{x}, \mathbf{y}) \in \mathbb{B}_{\delta_0}(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ and thus $F(\mathbf{x}, \mathbf{y}) \leq F(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \epsilon$. Consequently, for any $\mathbf{x} \in \mathbb{B}_\delta(\bar{\mathbf{x}}) \cap \mathcal{X}$, we have

$$\varphi(\mathbf{x}) = \min_{\mathbf{y} \in \mathcal{S}(\mathbf{x})} F(\mathbf{x}, \mathbf{y}) \leq F(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \epsilon = \varphi(\bar{\mathbf{x}}) + 2\epsilon,$$

which contradicts to $\varphi(\mathbf{x}^t) \to \infty$. $\qquad \square$

Thanks to the continuity of $f(\mathbf{x}, \mathbf{y})$, we further have the following result.

Lemma 2. *Denote $f^*(\mathbf{x}) = \min_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})$. If $f(\mathbf{x}, \mathbf{y})$ is continuous on $\mathcal{X} \times \mathbb{R}^m$, then $f^*(\mathbf{x})$ is USC on $\mathcal{X}$.*

Proof. For any sequence $\{\mathbf{x}^t\} \subseteq \mathcal{X}$ satisfying $\mathbf{x}^t \to \bar{\mathbf{x}} \in \mathcal{X}$, given any $\epsilon > 0$, let $\bar{\mathbf{y}} \in \mathbb{R}^m$ satisfy $f(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \leq f^*(\bar{\mathbf{x}}) + \epsilon$. As f is continuous at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, there exists $T > 0$ such that

$$f^*(\mathbf{x}^t) \leq f(\mathbf{x}^t, \bar{\mathbf{y}}) \leq f(\bar{\mathbf{x}}, \bar{\mathbf{y}}) + \epsilon \leq f^*(\bar{\mathbf{x}}) + 2\epsilon, \quad \forall t > T,$$

and thus

$$\limsup_{t \to \infty} f^*(\mathbf{x}^t) \leq f^*(\bar{\mathbf{x}}) + 2\epsilon.$$

By taking $\epsilon \to 0$, we get $\limsup_{k \to \infty} f^*(\mathbf{x}^t) \leq f^*(\bar{\mathbf{x}})$. $\qquad \square$

In the following proposition, we derive properties on $\{\mathbf{y}_K(\mathbf{x})\}$ in the light of the general fact stated in (Sabach & Shtern, 2017).

Proposition 1. *Suppose Assumption 1 is satisfied and let $\{\mathbf{y}_K\}$ be the output generated as $\mathbf{y}_{k+1} = \mathbf{y}_k - \left(\alpha_k \mathbf{d}_k^F(\mathbf{x}) + (1 - \alpha_k) \mathbf{d}_k^f(\mathbf{x}) \right)$, $s_l \in (0, 1/L_f]$, $s_u \in (0, 2/(L_F + \sigma)]$,*

$$\alpha_k = \min \left\{ 2\gamma/k(1 - \beta), 1 - \varepsilon \right\},$$

with $k \geq 1$, $\varepsilon \in (0, 1)$, $\gamma \in (0, 1]$, and

$$\beta = \sqrt{1 - 2s_u \sigma L_F/(\sigma + L_F)}.$$

Denote $\tilde{\mathbf{y}}_K(\mathbf{x}) = \mathbf{y}_K(\mathbf{x}) - s_l \nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}_K(\mathbf{x}))$, and

$$C_{\mathbf{y}^*(\mathbf{x})} = \max \left\{ \|\mathbf{y}_0 - \mathbf{y}^*(\mathbf{x})\|, \frac{s_u}{1 - \beta} \|\nabla_{\mathbf{y}} F(\mathbf{x}, \mathbf{y}^*(\mathbf{x}))\| \right\},$$

with $\mathbf{y}^*(\mathbf{x}) \in \tilde{\mathcal{S}}(\mathbf{x})$ and $\mathbf{x} \in \mathcal{X}$. Then we have

$$\|\mathbf{y}_K(\mathbf{x}) - \mathbf{y}^*(\mathbf{x})\| \leq C_{\mathbf{y}^*(\mathbf{x})},$$
$$\|\mathbf{y}_K(\mathbf{x}) - \tilde{\mathbf{y}}_K(\mathbf{x})\| \leq \frac{2C_{\mathbf{y}^*(\mathbf{x})}(J + 2)}{K(1 - \beta)},$$
$$f(\mathbf{x}, \tilde{\mathbf{y}}_K(\mathbf{x})) - f^*(\mathbf{x}) \leq \frac{2C_{\mathbf{y}^*(\mathbf{x})}^2(J + 2)}{K(1 - \beta)s_l},$$

where $J = \lfloor 2/(1 - \beta) \rfloor$. Furthermore, for any $\mathbf{x} \in \mathcal{X}$, $\{\mathbf{y}_K(\mathbf{x})\}$ converges to $\tilde{\mathcal{S}}(\mathbf{x})$ as $K \to \infty$.

We next prove the uniform convergence of $\{\tilde{\mathbf{y}}_K(\mathbf{x})\}$ towards the solution set $\mathcal{S}(\mathbf{x})$ through the uniform convergence of $\{f(\mathbf{x}, \tilde{\mathbf{y}}_K(\mathbf{x}))\}$.

Proposition 2. *Let $\mathcal{Y} \subseteq \mathbb{R}^m$ be a bounded set and $\epsilon > 0$. If $\mathcal{S}(\mathbf{x})$ is ISC on $\mathcal{X}$, then there exists $\delta > 0$ such that for any $\mathbf{y} \in \mathcal{Y}$,*
$$\sup_{\mathbf{x} \in \mathcal{X}} \text{dist}(\mathbf{y}, \mathcal{S}(\mathbf{x})) \leq \epsilon,$$
in case $\sup_{\mathbf{x} \in \mathcal{X}} \{f(\mathbf{x}, \mathbf{y}) - f^*(\mathbf{x})\} \leq \delta$ is satisfied.

Proof. We are going to prove this statement by a contradiction. We assume that there exist bounded set $\mathcal{Y} \subseteq \mathbb{R}^m$, $\epsilon > 0$, sequences $\{(\mathbf{x}^t, \mathbf{y}^t)\} \subseteq \mathcal{X} \times \mathcal{Y}$ and $\{\delta_k\}$ with $\delta_k \to 0$ satisfying
$$f(\mathbf{x}^t, \mathbf{y}^t) - f^*(\mathbf{x}^t) \leq \delta_k \text{ and } \texttt{dist}(\mathbf{y}^t, \mathcal{S}(\mathbf{x}^t)) > \epsilon.$$

Without loss of generality, we can assume that $\mathbf{x}^t \to \bar{\mathbf{x}} \in \mathcal{X}$ and $\mathbf{y}^t \to \bar{\mathbf{y}} \in \mathbb{R}^m$ as $t \to \infty$. According to the continuity of f and the USC of f^* from Lemma 2, we have
$$0 \leq f(\bar{\mathbf{x}}, \bar{\mathbf{y}}) - f^*(\bar{\mathbf{x}}) \leq \liminf_{t \to \infty} f(\mathbf{x}^t, \mathbf{y}^t) - f^*(\mathbf{x}^t) \leq 0,$$

which implies $\bar{\mathbf{y}} \in \mathcal{S}(\bar{\mathbf{x}})$. However, as $\texttt{dist}(\mathbf{y}^t, \mathcal{S}(\mathbf{x}^t)) > \epsilon$, following from the ISC of $\mathcal{S}(\mathbf{x})$ at $\bar{\mathbf{x}}$ and Proposition 5.11 of (Rockafellar & Wets, 2009), we have
$$\begin{aligned}
\texttt{dist}(\bar{\mathbf{y}}, \mathcal{S}(\bar{\mathbf{x}})) &\geq \limsup_{t \to \infty} \texttt{dist}(\bar{\mathbf{y}}, \mathcal{S}(\mathbf{x}^t)) \\
&= \limsup_{t \to \infty} \left(\texttt{dist}(\mathbf{y}^t, \mathcal{S}(\mathbf{x}^t)) + \|\mathbf{y}^t - \bar{\mathbf{y}}\| \right) \\
&\geq \liminf_{t \to \infty} \texttt{dist}(\mathbf{y}^t, \mathcal{S}(\mathbf{x}^t)) \geq \epsilon,
\end{aligned}$$

which contradicts to $\bar{\mathbf{y}} \in \mathcal{S}(\bar{\mathbf{x}})$. □

Combining Lemmas 1 and 2, together with Proposition 2, the *LL solution set* property required in Theorem 1 can be eventually derived. Let us now prove the LSC property of φ on $\mathcal{X}$ in the following proposition.

Proposition 3. *Suppose $F(\mathbf{x}, \mathbf{y})$ is level-bounded in $\mathbf{y}$ locally uniformly in $\mathbf{x} \in \mathcal{X}$. If $\mathcal{S}(\mathbf{x})$ is OSC at $\mathbf{x} \in \mathcal{X}$, then $\varphi(\mathbf{x})$ is LSC at $\mathbf{x} \in \mathcal{X}$.*

Proof. We assume that there exists $\bar{\mathbf{x}} \in \mathcal{X}$ satisfying $\mathbf{x}^t \to \bar{\mathbf{x}}$ as $t \to \infty$, then the following
$$\liminf_{\mathbf{x} \to \bar{\mathbf{x}}} \varphi(\mathbf{x}) < \varphi(\bar{\mathbf{x}}),$$

holds. Next, there exist $\epsilon > 0$ and sequences $\mathbf{x}^t \to \bar{\mathbf{x}} \in \mathcal{X}$ and $\mathbf{y}^t \in \mathcal{S}(\mathbf{x}^t)$ satisfying
$$F(\mathbf{x}^t, \mathbf{y}^t) \leq \varphi(\mathbf{x}^t) + \epsilon < \varphi(\bar{\mathbf{x}}) - \epsilon.$$

Furthermore, since $F(\mathbf{x}, \mathbf{y})$ is level-bounded in $\mathbf{y}$ locally uniformly in $\mathbf{x} \in \mathcal{X}$, we have that $\{\mathbf{y}^t\}$ is bounded. Take a subsequence $\{\mathbf{y}^\nu\}$ of $\{\mathbf{y}^t\}$ such that $\mathbf{y}^\nu \to \hat{\mathbf{y}}$ and it follows from the OSC of $\mathcal{S}$ that $\hat{\mathbf{y}} \in \mathcal{S}(\bar{\mathbf{x}})$. Then we have
$$\begin{aligned}
\varphi(\bar{\mathbf{x}}) \leq F(\bar{\mathbf{x}}, \hat{\mathbf{y}}) &\leq \limsup_{t \to \infty} F(\mathbf{x}^t, \mathbf{y}^t) = \limsup_{t \to \infty} \varphi(\mathbf{x}^t) \\
&\leq \varphi(\bar{\mathbf{x}}) - \epsilon,
\end{aligned}$$

which implies a contradiction. Thus
$$\varphi(\bar{\mathbf{x}}) \leq \liminf_{\mathbf{x} \to \bar{\mathbf{x}}} \varphi(\mathbf{x})$$

and we get the conclusion. □

Theorem 2. *Suppose Assumption 1 is satisfied and let $\{\mathbf{y}_K\}$ be the output generated as $\mathbf{y}_{k+1} = \mathbf{y}_k - \left(\alpha_k \mathbf{d}_k^F(\mathbf{x}) + (1 - \alpha_k) \mathbf{d}_k^f(\mathbf{x}) \right)$, $s_l \in (0, 1/L_f]$, $s_u \in (0, 2/(L_F + \sigma)]$,*
$$\alpha_k = \min \{2\gamma/k(1 - \beta), 1 - \varepsilon\},$$
with $k \geq 1$, $\varepsilon \in (0, 1)$, $\gamma \in (0, 1]$ and
$$\beta = \sqrt{1 - 2s_u \sigma L_F / (\sigma + L_F)}.$$

Assume further that $\mathcal{S}(\mathbf{x})$ is continuous on $\mathcal{X}$. Then we have that both the LL solution set and UL objective convergence properties hold.

Proof. We first show that $F(\mathbf{x}, \mathbf{y})$ is level-bounded in $\mathbf{y}$ locally uniformly in $\mathbf{x} \in \mathcal{X}$. For any $\bar{\mathbf{x}} \in \mathcal{X}$, let $\{\mathbf{x}^t\} \subseteq \mathcal{X}$ with $\mathbf{x}^t \to \bar{\mathbf{x}}$ and $\{\mathbf{y}^t\} \in \mathbb{R}^m$ with $\|\mathbf{y}^t\| \to +\infty$. Then, with Assumption 1 we have
$$\begin{aligned}
F(\mathbf{x}^t, \mathbf{y}^t) \geq &F(\mathbf{x}^t, \mathbf{y}^1) + \langle \nabla_{\mathbf{y}} F(\mathbf{x}^t, \mathbf{y}^1), \mathbf{y}^t - \mathbf{y}^1 \rangle \\
&+ \frac{\sigma}{2} \|\mathbf{y}^t - \mathbf{y}^1\|^2.
\end{aligned}$$

As $F(\mathbf{x}, \cdot) : \mathbb{R}^m \to \mathbb{R}$ is Lipschitz continuous with uniform constant L_0 for any $\mathbf{x} \in \mathcal{X}$, we have $\|\nabla_{\mathbf{y}} F(\mathbf{x}^t, \mathbf{y}^1)\| \leq L_0$. Then, by the continuity of F, with $\mathbf{x}^t \to \bar{\mathbf{x}} \in \mathcal{X}$, and $\|\mathbf{y}^t\| \to +\infty$, we have $F(\mathbf{x}^t, \mathbf{y}^t) \to +\infty$. Thus $F(\mathbf{x}, \mathbf{y})$ is level-bounded in $\mathbf{y}$ locally uniformly in $\mathbf{x} \in \mathcal{X}$. Then with Proposition 3 and assumptions in Theorem 2, we get the LSC property of φ on $\mathcal{X}$. And according to Lemma 1, there exists $M > 0$ such that $C_{\mathbf{y}^*(\mathbf{x})} \leq M$ for any $\mathbf{y}^*(\mathbf{x}) \in \tilde{\mathcal{S}}(\mathbf{x})$ and $\mathbf{x} \in \mathcal{X}$. Following Proposition 1, there exists $C > 0$ such that for any $\mathbf{x} \in \mathcal{X}$ we have
$$\|\mathbf{y}_K(\mathbf{x})\| \leq C, \quad \forall K \geq 0,$$
$$\|\mathbf{y}_K(\mathbf{x}) - \tilde{\mathbf{y}}_K(\mathbf{x})\| \leq \frac{C}{K},$$
and
$$f(\tilde{\mathbf{y}}_K(\mathbf{x})) - f^*(\mathbf{x}) \leq \frac{C}{K}, \quad \forall K \geq 0.$$

Next, according to Proposition 2, for any $\epsilon > 0$, there exists $k(\epsilon) > 0$ such that whenever $K > \max\{2C/\epsilon, k(\epsilon)\}$ we have
$$\begin{aligned}
&\sup_{\mathbf{x} \in \mathcal{X}} \texttt{dist}(\mathbf{y}_K(\mathbf{x}), \mathcal{S}(\mathbf{x})) \\
&\leq \|\mathbf{y}_K(\mathbf{x}) - \tilde{\mathbf{y}}_K(\mathbf{x})\| + \sup_{\mathbf{x} \in \mathcal{X}} \texttt{dist}(\tilde{\mathbf{y}}_K(\mathbf{x}), \mathcal{S}(\mathbf{x})) \leq \epsilon.
\end{aligned}$$

Then it follows from Proposition 1 that $\varphi_K(\mathbf{x}) \to \varphi(\mathbf{x})$ when $K \to \infty$ for any $\mathbf{x} \in \mathcal{X}$. □

1.3. Proof of Theorem 3

Lemma 3. *Suppose $\mathcal{S}(\mathbf{x})$ is single-valued on $\mathcal{X}$ and Assumption 2 is satisfied. Then $\mathcal{S}(\mathbf{x})$ is continuous on $\mathcal{X}$.*

Proof. First, according to Proposition 4.4 of (Bonnans & Shapiro, 2013), we know that if $f(\mathbf{x}, \mathbf{y}) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is continuous on $\mathcal{X} \times \mathbb{R}^m$, level-bounded in $\mathbf{y}$ locally uniformly in $\mathbf{x} \in \mathcal{X}$, then $f^*(\mathbf{x})$ is continuous on $\mathcal{X}$, $\mathcal{S}(\mathbf{x})$ is OSC on $\mathcal{X}$ and locally bounded at $\bar{\mathbf{x}}$. Thus, for any $\bar{\mathbf{x}} \in \mathcal{X}$, $f^*(\mathbf{x})$ is locally bounded at $\bar{\mathbf{x}}$. As $\mathcal{S}(\mathbf{x})$ is a single-valued mapping on $\mathcal{X}$ and $\mathcal{S}(\mathbf{x})$ is OSC at $\bar{\mathbf{x}} \in \mathcal{X}$ and locally bounded at $\bar{\mathbf{x}}$, Upon Proposition 5.20 of (Rockafellar & Wets, 2009), we conclude that $\mathcal{S}(\mathbf{x})$ is ISC at $\bar{\mathbf{x}}$, and thus continuous at $\bar{\mathbf{x}}$. This completes the proof. $\square$

Theorem 3. *Suppose $\mathcal{S}(\mathbf{x})$ is single-valued on $\mathcal{X}$ and Assumption 2 is satisfied, $\{\mathbf{y}_K(\mathbf{x})\}$ is uniformly bounded on $\mathcal{X}$, and $\{f(\mathbf{x}, \mathbf{y}_K(\mathbf{x}))\}$ converges uniformly to $f^*(\mathbf{x})$ on $\mathcal{X}$ as $K \to \infty$. Then we have that both the LL solution set and UL objective convergence properties hold.*

Proof. First, we get the continuity of $\mathcal{S}(\mathbf{x})$ on $\mathcal{X}$ from Lemma 3. Then, by Proposition 3, we obtain the LSC of $\varphi(\mathbf{x})$ on $\mathcal{X}$. From Proposition 2 and Lemma 3, we have that for any $\epsilon > 0$, there exists $k(\epsilon) > 0$ such that whenever $K > k(\epsilon)$,

$$\sup_{\mathbf{x} \in \mathcal{X}} \mathtt{dist}(\mathbf{y}_K(\mathbf{x}), \mathcal{S}(\mathbf{x})) \leq \epsilon.$$

As $\mathcal{S}(\mathbf{x})$ is a single-valued mapping on $\mathcal{X}$, we have $\varphi_K(\mathbf{x}) \to \varphi(\mathbf{x})$ for any $\mathbf{x} \in \mathcal{X}$ as $K \to \infty$. $\square$

In the following two propositions, we assume that $f(\mathbf{x}, \cdot) : \mathbb{R}^m \to \mathbb{R}$ is L_f-smooth and convex, $s_l \leq 1/L_f$.

Proposition 4. *Let $\mathbf{y}_K$ be the output generated as $\mathbf{y}_{k+1} = \mathbf{y}_k - s_l \nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}_k)$. Then, it holds that*

$$\|\mathbf{y}_K(\mathbf{x}) - \mathbf{y}^*(\mathbf{x})\| \leq \|\mathbf{y}_0 - \mathbf{y}^*(\mathbf{x})\|,$$

and

$$f(\mathbf{y}_K(\mathbf{x})) - f^*(\mathbf{x}) \leq \frac{\|\mathbf{y}_0 - \mathbf{y}^*(\mathbf{x})\|^2}{2 s_l K},$$

with $\mathbf{y}^*(\mathbf{x}) \in \mathcal{S}(\mathbf{x})$ and $\mathbf{x} \in \mathcal{X}$.

Proof. This proposition can be directly obtained from Theorem 10.21 and Theorem 10.23 of (Beck, 2017). $\square$

Then in the following proposition we can immediately verify our required assumption on $\{f(\mathbf{x}, \mathbf{y}_K(\mathbf{x}))\}$ in the absence of the strong convexity property on the LL objective.

Proposition 5. *Suppose that $\mathcal{S}(\mathbf{x})$ is single-valued on $\mathcal{X}$ and Assumption 2 is satisfied. Let $\mathbf{y}_K$ be the output generated as $\mathbf{y}_{k+1} = \mathbf{y}_k - s_l \nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}_k)$. Then $\{\mathbf{y}_K(\mathbf{x})\}$ is uniformly bounded on $\mathcal{X}$ and $\{f(\mathbf{x}, \mathbf{y}_K(\mathbf{x}))\}$ converges uniformly to $f^*(\mathbf{x})$ on $\mathcal{X}$ as $K \to \infty$.*

Proof. By the same arguments given in proof of Lemma 3, we can show that $\mathcal{S}(\mathbf{x})$ is locally bounded at each point on $\mathcal{X}$ under Assumption 2. As $\mathcal{X}$ is compact, thus $\cup_{\mathbf{x} \in \mathcal{X}} \mathcal{S}(\mathbf{x})$ is bounded. Then the conclusion follows from Proposition 4 directly. $\square$

2. Extended Theoretical Results

2.1. Local Convergence Results

In this part, we analyze the local convergence behaviors of BDA. In fact, even if $\mathbf{x}_K$ is a local minimum of $\varphi_K(\mathbf{x})$ with uniform neighborhood modulus $\delta > 0$, we can still obtain similar convergence results as that in Theorem 1. Such properties are summarized in the following theorem.

Theorem 4. *Suppose both the LL solution set and UL objective convergence properties (stated in Section 1.1) hold and let $\mathbf{x}_K$ be a local minimum of $\varphi_K(\mathbf{x})$ with uniform neighborhood modulus $\delta > 0$. Then we have that any limit point $\bar{\mathbf{x}}$ of the sequence $\{\mathbf{x}_K\}$ is a local minimum of φ, i.e., there exists $\tilde{\delta} > 0$ such that*

$$\varphi(\bar{\mathbf{x}}) \leq \varphi(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{B}_{\tilde{\delta}}(\bar{\mathbf{x}}) \cap \mathcal{X}.$$

Proof. Since $\mathcal{X}$ is compact, we can assume without loss of generality that $\mathbf{x}_K \to \bar{\mathbf{x}} \in \mathcal{X}$ and $\mathbf{x}_K \in \mathbb{B}_{\delta/2}(\bar{\mathbf{x}})$ by considering a subsequence of $\{\mathbf{x}_K\}$. For any $\epsilon > 0$, there exists $k(\epsilon) > 0$ such that whenever $K > k(\epsilon)$, we have

$$\sup_{\mathbf{x} \in \mathcal{X}} \mathtt{dist}(\mathbf{y}_K(\mathbf{x}), \mathcal{S}(\mathbf{x})) \leq \frac{\epsilon}{2 L_0}.$$

Thus, for any $\mathbf{x} \in \mathcal{X}$, there exists $\mathbf{y}^*(\mathbf{x}) \in \mathcal{S}(\mathbf{x})$ such that

$$\|\mathbf{y}_K(\mathbf{x}) - \mathbf{y}^*(\mathbf{x})\| \leq \frac{\epsilon}{L_0}.$$

Therefore, for any $\mathbf{x} \in \mathcal{X}$, we have

$$
\begin{aligned}
\varphi(\mathbf{x}) &= \inf_{\mathbf{y} \in \mathcal{S}(\mathbf{x})} F(\mathbf{x}, \mathbf{y}) \\
&\leq F(\mathbf{x}, \mathbf{y}^*(\mathbf{x})) \\
&\leq F(\mathbf{x}, \mathbf{y}_K(\mathbf{x})) + L_0 \|\mathbf{y}_K(\mathbf{x}) - \mathbf{y}^*(\mathbf{x})\| \\
&\leq \varphi_K(\mathbf{x}) + \epsilon.
\end{aligned}
$$

This implies that, for any $\epsilon > 0$, there exists $k(\epsilon) > 0$ such that whenever $K > k(\epsilon)$, we have

$$\varphi(\mathbf{x}_K) \leq \varphi_K(\mathbf{x}_K) + \epsilon \leq \varphi_K(\mathbf{x}) + \epsilon, \quad \forall \mathbf{x} \in \mathcal{X}.$$

Next, as $\mathbf{x}_K$ is a local minimum of $\varphi_K(\mathbf{x})$ with uniform neighborhood modulus δ, it follows

$$\varphi_K(\mathbf{x}_K) \leq \varphi_K(\mathbf{x}), \ \forall \mathbf{x} \in \mathbb{B}_\delta(\mathbf{x}_K) \cap \mathcal{X}.$$

Since $\mathbb{B}_{\delta/2}(\bar{\mathbf{x}}) \subseteq \mathbb{B}_{\delta/2 + \|\mathbf{x}_k - \bar{\mathbf{x}}\|}(\mathbf{x}_K) \subseteq \mathbb{B}_\delta(\mathbf{x}_K)$, we have that for any $\epsilon > 0$, $\forall \mathbf{x} \in \mathbb{B}_{\delta/2}(\bar{\mathbf{x}}) \cap \mathcal{X}$, there exists $k(\epsilon) > 0$ such that whenever $K > k(\epsilon)$,

$$\varphi(\mathbf{x}_K) \leq \varphi_K(\mathbf{x}_K) + \epsilon \leq \varphi_K(\mathbf{x}) + \epsilon.$$

Taking $K \to \infty$ and by the LSC of φ, $\forall \mathbf{x} \in \mathbb{B}_{\delta/2}(\bar{\mathbf{x}}) \cap \mathcal{X}$, we have

$$
\begin{aligned}
\varphi(\bar{\mathbf{x}}) &\leq \liminf_{K \to \infty} \varphi(\mathbf{x}_K) \\
&\leq \liminf_{K \to \infty} \varphi_K(\mathbf{x}_K) + \epsilon \\
&\leq \lim_{K \to \infty} \varphi_K(\mathbf{x}) + \epsilon = \varphi(\mathbf{x}) + \epsilon.
\end{aligned}
$$

By taking $\epsilon \to 0$, we have

$$
\varphi(\bar{\mathbf{x}}) \leq \varphi(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{B}_{\delta/2}(\bar{\mathbf{x}}) \cap \mathcal{X},
$$

which implies $\bar{\mathbf{x}} \in \arg\min_{\mathbf{x} \in \mathbb{B}_{\delta/2}(\bar{\mathbf{x}}) \cap \mathcal{X}} \varphi(\mathbf{x})$, i.e, $\bar{x}$ is a local minimum of φ. $\square$

2.2. Nonsmooth LL Objective

It is well-known that a variety of nonsmooth regularization techniques (e.g., ℓ_1-norm regularization) have been utilized in learning and vision areas. So in this section, we briefly discuss a potential extension of BDA for BLPs with the nonsmooth LL objective, e.g.,

$$
\mathcal{S}(\mathbf{x}) = \arg\min_{\mathbf{y}} h(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}, \mathbf{y}) + g(\mathbf{x}, \mathbf{y}). \quad (2)
$$

Here we consider f as a function with the same properties as that in our above analysis, while g is convex but not necessarily smooth, w.r.t. $\mathbf{y}$ and continuous w.r.t. $(\mathbf{x}, \mathbf{y})$. Since g is not necessarily differentiable w.r.t. $\mathbf{y}$, these existing gradient-based first-order BLP methods are not available for this problem. Fortunately, we demonstrate that by slightly modifying our inner updating rule, BDA can be directly extended to address BLPs with the nonsmooth LL objective in Eq. (2). Specifically, we first write the descent direction of the LL subproblem as

$$
\mathbf{d}_k^h(\mathbf{x}) = \mathbf{y}_k - \texttt{prox}_{s_l g(\mathbf{x}, \cdot)}(\mathbf{y}_k - s_l \nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}_k)),
$$

where $\texttt{prox}_{s_l g(\mathbf{x}, \cdot)}$ denotes the proximal operator w.r.t. the nonsmooth function $g(\mathbf{x}, \cdot)$ and step size s_l. Then by aggregating $\mathbf{d}_k^F(\mathbf{x})$ and $\mathbf{d}_k^h(\mathbf{x})$, we derive a new $\mathcal{T}_k$ to handle BLPs with the nonsmooth composite LL objective h, i.e.,

$$
\mathcal{T}_{k+1}(\mathbf{x}, \mathbf{y}_k(\mathbf{x})) = \mathbf{y}_k - \left(\alpha_k \mathbf{d}_k^h(\mathbf{x}) + (1 - \alpha_k) \mathbf{d}_k^F(\mathbf{x}) \right), \quad (3)
$$

where $\alpha_k \in (0, 1]$. In fact, since explicitly estimating the subgradient information of some proximal operators may be computationally infeasible in practice, one may apply automatic differentiation through the dynamical system with approximation techniques (Wang et al., 2016; Rajeswaran et al., 2019) to obtain $\frac{d\varphi_K}{d\mathbf{x}}$, where $\varphi_K(\mathbf{x}) = F(\mathbf{x}, \mathbf{y}_K(\mathbf{x}))$.

We are now in the position to extend the converge properties of BDA for BLPs in Eq. (3) from smooth LL case to nonsmooth LL case. Similar to the discussion in the smooth case, our analysis could follow the following roadmap:

Step 1: Denoting $\tilde{\mathcal{S}}(\mathbf{x}) = \arg\min_{\mathbf{y} \in \mathcal{S}(\mathbf{x})} F(\mathbf{x}, \mathbf{y})$ and further $h^*(\mathbf{x}) = \min_{\mathbf{y}} h(\mathbf{x}, \mathbf{y})$, as extensions to Lemma 1 and Lemma 2, we shall derive the boundedness of $\tilde{\mathcal{S}}(\mathbf{x})$ and the USC of $h^*(\mathbf{x})$ for the nonsmooth LL case, respectively. The proofs are indeed straightforward and purely technical, thus omitted here.

Step 2: As an extension to Proposition 1 which focuses on the smooth case, we may derive the following convergence results regarding $\{\mathbf{y}_K(\mathbf{x})\}$ in the light of the general fact stated in (Sabach & Shtern, 2017).

Proposition 6. *Suppose Assumption 1 is satisfied, g is continuous and convex w.r.t. $\mathbf{y}$, and let $\{\mathbf{y}_K\}$ be defined as in Eq. (3), $s_l \in (0, 1/L_f]$, $s_u \in (0, 2/(L_F + \sigma)]$,*

$$
\alpha_k = \min \{2\gamma/k(1 - \beta), 1 - \varepsilon\},
$$

with $k \geq 1$, $\varepsilon \in (0, 1)$, $\gamma \in (0, 1]$ and

$$
\beta = \sqrt{1 - 2s_u \sigma L_F / (\sigma + L_F)}.
$$

Denoting

$$
\tilde{\mathbf{y}}_K(\mathbf{x}) = \texttt{prox}_{s_l g(\mathbf{x}, \cdot)}(\mathbf{y}_K(\mathbf{x}) - s_l \nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}_K(\mathbf{x}))),
$$

and

$$
C_{\mathbf{y}^*(\mathbf{x})} = \max \left\{ \|\mathbf{y}_0 - \mathbf{y}^*(\mathbf{x})\|, \frac{s_u}{1 - \beta} \|\nabla_{\mathbf{y}} F(\mathbf{x}, \mathbf{y}^*(\mathbf{x}))\| \right\},
$$

with $\mathbf{y}^*(\mathbf{x}) \in \tilde{\mathcal{S}}(\mathbf{x})$ and $\mathbf{x} \in \mathcal{X}$. Then it holds that

$$
\begin{aligned}
\|\mathbf{y}_K(\mathbf{x}) - \mathbf{y}^*(\mathbf{x})\| &\leq C_{\mathbf{y}^*(\mathbf{x})}, \\
\|\mathbf{y}_K(\mathbf{x}) - \tilde{\mathbf{y}}_K(\mathbf{x})\| &\leq \frac{2C_{\mathbf{y}^*(\mathbf{x})}(J + 2)}{K(1 - \beta)}, \\
h(\mathbf{x}, \tilde{\mathbf{y}}_K(\mathbf{x})) - h^*(\mathbf{x}) &\leq \frac{2C_{\mathbf{y}^*(\mathbf{x})}^2(J + 2)}{K(1 - \beta)s_l},
\end{aligned}
$$

where $J = \lfloor 2/(1 - \beta) \rfloor$. Further, $\mathbf{y}_K(\mathbf{x})$ converges to $\tilde{\mathcal{S}}(\mathbf{x})$ as $K \to \infty$ for any $\mathbf{x} \in \mathcal{X}$.

Step 3: Taking a closer look at the proofs for Proposition 2 and Proposition 3, we observe that the techniques we used barely rely on the smoothness of the LL objective. Therefore, straightforward extensions of Proposition 2 and Proposition 3 to the nonsmooth case can yield the desired uniform convergence of $\tilde{\mathbf{y}}_K(\mathbf{x})$ and the UL objective convergence, respectively.

Step 4: Similar to the arguments in the proof of Theorem 2, by combining Step 1 and Step 2, we eventually meet the *LL solution set* and *UL objective convergence* properties, and hence the analysis framework in Theorem 1 has been activated. Therefore, the same convergence results concerning $\{\mathbf{x}_K\}_{K \in \mathbb{N}}$ and $\{\varphi_K(\mathbf{x})\}$ can be achieved as following.

Theorem 5. *Suppose Assumption 1 is satisfied, g is continuous and convex w.r.t.* $\mathbf{y}$*, and let* $\{\mathbf{y}_K\}$ *be defined as in Eq.* (3)*,* $s_l \in (0, 1/L_f]$*,* $s_u \in (0, 2/(L_F + \sigma)]$*,*

$$\alpha_k = \min\left\{2\gamma/k(1-\beta), 1-\varepsilon\right\},$$

with $k \geq 1$*,* $\varepsilon \in (0,1)$*,* $\gamma \in (0,1]$ *and*

$$\beta = \sqrt{1 - 2s_u\sigma L_F/(\sigma + L_F)}.$$

Assume further that $\mathcal{S}(\mathbf{x})$ *is nonempty for any* $\mathbf{x} \in \mathcal{X}$ *and* $\mathcal{S}(\mathbf{x})$ *is continuous on* $\mathcal{X}$*. Then*

(1) if $\mathbf{x}_K \in \arg\min_{\mathbf{x} \in \mathcal{X}} \varphi_K(\mathbf{x})$*, we have the same results as in Theorem 1;*

(2) if $\mathbf{x}_K$ *is a local minimum of* $\varphi_K(\mathbf{x})$ *with uniform neighborhood modulus* $\delta > 0$*, we have that any limit point* $\bar{\mathbf{x}}$ *of the sequence* $\{\mathbf{x}_K\}$ *is a local minimum of* φ*.*

References

Beck, A. *First-order methods in optimization*. SIAM, 2017.

Bonnans, J. F. and Shapiro, A. *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2013.

Rajeswaran, A., Finn, C., Kakade, S. M., and Levine, S. Meta-learning with implicit gradients. In *NeurIPS*, pp. 113–124, 2019.

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

Sabach, S. and Shtern, S. A first order method for solving convex bilevel optimization problems. *SIAM Journal on Optimization*, 27(2):640–660, 2017.

Wang, S., Fidler, S., and Urtasun, R. Proximal deep structured models. In *NeurIPS*, pp. 865–873, 2016.