A Piecewise Convexification Method for Solving Bilevel Programs with A Nonconvex Follower’s Problem

Gaoxi Li

Received: date / Accepted: date

Abstract A new numerical method is presented for bilevel programs with a nonconvex follower’s problem. The basic idea is to piecewise construct convex relations of the follower’s problem, replace the relaxed follower’s problem equivalently by their Karush-Kuhn-Tucker conditions, and solve the resulting mathematical programs with equilibrium constraints. The convex relaxations and needed parameters are constructed with ideas of the $\alpha BB$ method of global optimization. Under mild conditions, we prove that every accumulated point of the solutions of the sequence approximate problems is an optimal solution of the original problem, and the convergence theorem of this algorithm is presented and proved. Numerical experiments show that the algorithm is efficient for solving this class of bilevel programs.

Keywords Bilevel programs · Equilibrium constraints · Global optimization · Piecewise convexification

Mathematics Subject Classification (2010) 90C26 · 90C30 · 90C33

1 Introduction

In this paper, we consider the following bilevel programming problem

$$\min_{x,y} F(x, y) \quad \text{s.t. } G(x) \leq 0, y \in \psi(x),$$

(1)
ψ(x) is the solution set of the following optimization problem parameterized in x:

$$\begin{align*}
\min_{y} f(x, y) \\
s.t. \ y \in Y := [\underline{Y}, \overline{Y}],
\end{align*}$$

(2)

here, \( F, f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \) are continuously differentiable functions, \( G : \mathbb{R}^n \to \mathbb{R}^p \) is also a continuously differentiable function, and \( f \) is twice continuously differentiable with respect to \( y \). \( Y \in \mathbb{R}^m \) is a box set, and \( \underline{Y} < \overline{Y} \in \mathbb{R}^m \). For convenience, we let \( X := \{ x \in \mathbb{R}^n : G(x) \leq 0 \} \). The reader is referred to [1, 4, 5, 8, 9, 26] for applications and developments of this bilevel program.

It is obvious that bilevel program (1)-(2) can be reformulated as the following single level programming problem via the optimal value function.

$$\begin{align*}
\min_{x, y} F(x, y) \\
s.t. \ G(x) &\leq 0, \\
&& f(x, y) - \varphi(x) &\leq 0, \\
&& \underline{Y} - y &\leq 0, \\
&& y - \overline{Y} &\leq 0,
\end{align*}$$

(3)

where \( \varphi(x) := \min_{y \in Y} f(x, y) \) is the optimal value function of the follower’s problem. This reformulation was first proposed by Outrata [25] for a numerical purpose. Ye and Zhu [34] gave the necessary optimality condition of bilevel programs by this reformulation. It is easy to verify that the optimal value function is in general a nonsmooth function even though the function \( f \) is continuously differentiable. In recently literature [21, 28–30], the following integral entropy function (4) is proposed to approximate the optimal value function.

$$\varphi_\rho(x) := -\rho^{-1} \ln \left( \int_{Y} \exp[-\rho f(x, y)] dy \right).$$

(4)

The classical Karush-Kuhn-Tucker (KKT) approach to solve bilevel program is to replace the solution set \( \psi(x) \) by the set of KKT points of the follower’s problem and consider the following single level programming problem:

$$\begin{align*}
\min_{x, y} F(x, y) \\
s.t. \ G(x) &\leq 0, \\
&& \nabla_y f(x, y) - \lambda + \mu &= 0, \\
&& 0 &\leq \lambda \perp (y - \underline{Y}) \geq 0, \\
&& 0 &\leq \mu \perp (\overline{Y} - y) \geq 0,
\end{align*}$$

(5)

here, \( \lambda, \mu \in \mathbb{R}^m \), and \( a \perp b \) means that vector \( a \) is perpendicular to vector \( b \). This is a mathematical program with equilibrium constraints (MPEC). For the case where the lower level objective function \( f \) is convex in variable \( y \), bilevel problem (1)-(2) is equal to the single level programming problem (5). In the case where \( f \) is not convex in variable \( y \), from [10] we know that the locally optimal solution of (5) may not be a locally optimal solution of bilevel problem (1)-(2). Hence using the KKT approach to solve bilevel problem (1)-(2) is not valid. Recent developments on KKT approach can be found in [3,
In addition, a vast literature on related MPEC problem is widely available [11, 14, 15, 22, 31, 32].

In this article we will construct a sequence of convexifications of the non-convex follower’s problem. Then the original problem is approximated with a bilevel multi-follower programming. After that, we will replace these convex follower’s programs equivalently by their KKT conditions, and solve the resulting MPECs. The convexifications will be produced using the ideas of the \( \alpha \)BB method [2, 12] which is used to solve semi-infinite programming [13, 27]. The bilevel programming and semi-infinite programming are two different problems, although the latter also has simple bilevel structure. Thus the convexification transformation in the paper is clearly different from those in [13, 27].

The article is organized as follows. In section 2, a brief review of upper and lower convergence properties of set-valued mapping sequence and the main idea of the \( \alpha \)BB method are given. In section 3, the original bilevel program is approximated by a bilevel multi-follower programming, and the property of accumulated point of the solutions of the sequence approximate problems is discussed here. In section 4, the algorithms are presented, and the convergence theorem is presented and proved. In section 5, some numerical experiments are given.

2 Preliminaries

2.1 Basic Property of Set-valued Mapping

In this subsection we give a short overview of the convergence properties of set-valued mapping sequences and the semicontinuous property of set-valued mapping. For convenience, we use \( \mathbb{R}_+ \) and \( \mathbb{Z}_+ \) to denote the set of positive real number and the set of positive integer. For \( U \subseteq \mathbb{R}^p, V \subseteq \mathbb{R}^q \), we let \( M, M_n : U \rightrightarrows V \) be set-valued functions, \( n = 1, 2, 3,... \). And we recall the definitions of the lower and upper convergence of the set-valued function sequence \( \{ M_n \} \).

**Definition 2.1** [19, 20] The sequence \( \{ M_n \} \) is lower convergent to \( M \) at \( x \in U \) iff for any \( \{ x_n \} \) converging to \( x \) in \( U \), and any \( y \in M(x) \), there exists \( \{ y_n \} \) converging to \( y \) in \( V \) such that \( y_n \in M_n(x_n) \) for \( n \) large.

**Definition 2.2** [20] The sequence \( \{ M_n \} \) is upper convergent to \( M \) at \( x \in U \) iff for any \( y \in V \), any \( \{ x_n \} \) converging to \( x \) in \( U \), and any \( \{ y_k \} \) converging to \( y \) in \( V \) such that \( y_k \in M_n(x_{n_k}) \) for a selection of integers \( \{ n_k \} \), we have \( y \in M(x) \).

**Definition 2.3** [17] We claim that \( M \) is lower semicontinuous at \( x \) iff for any \( \{ x_n \} \) converging to \( x \) as \( n \rightarrow \infty \) in \( U \), and for any \( y \in M(x) \), there exists \( \{ y_n \} \) converging to \( y \) such that \( y_n \in M(x_n) \) for \( n \) large.

**Remark 2.1** Definition 2.1 and Definition 2.3 are two different definitions, although they are similar. All the three definitions are used to discuss the property of accumulated point of the solutions of sequence approximate problems.
Remark 2.2 Definition 2.1 shows the convergence of a sequence of set-valued maps \( \{M_n\} \), but Definition 2.3 shows the continuity of a set-valued map \( M \). If \( M_n(x) = M(x) \) for all \( n \in \mathbb{Z}_+ \) and \( x \in U \), and if \( M \) is a compact set at \( x \). Then Definition 2.1 amounts to Definition 2.3.

Definition 2.4 For any \( \epsilon > 0 \), the open \( \epsilon \)-neighborhood of \( U \in \mathbb{R}^p \) is defined as \( U_{+\epsilon} = \{z \in \mathbb{R}^p : \text{there is } x \in U \text{ such that } \|z - x\| < \epsilon\} \).

2.2 The \( \alpha \) BB Method

The \( \alpha \) BB method operates within a branch-and-bound framework and is designed to solve nonconvex minimization problems [2, 13, 27]. In \( \alpha \) BB, a convex underestimator of a nonconvex function is constructed by decomposing it into a sum of nonconvex terms of special type. These nonconvex terms are then replaced by their convex envelop or very tight convex underestimators which are already known.

For \( E = [\underline{E}, \overline{E}] \), \( \underline{E} < \overline{E} \in \mathbb{R}^m \), we suppose that \( f : E \to \mathbb{R} \) is a real-valued twice continuously differentiable function. With the \( \alpha \) BB method of global optimization one can construct convex relaxation of \( f \) on \( E \) by adding a quadratic term. More precisely, we define a relaxation function \( \bar{f} : E \to \mathbb{R} \) of \( f \) by adding a negative function \( \phi(y; \rho, E) \) to \( f \), that is,

\[
\bar{f}(y, \rho, E) = f(y) + \phi(y; \rho, E),
\]

where \( \phi(y; \rho, E) \) is dependent on a parameter \( \rho \) and the interval \( E \). In this paper, the term \( \phi(y; \rho, E) \) is set as

\[
\phi(y; \rho, E) = \frac{\rho}{2} \langle \underline{E} - y, \overline{E} - y \rangle. \tag{6}
\]

In this case, the function \( \bar{f}(y, \rho, E) \) is twice continuously differentiable with respect to \( y \). Furthermore, the Hessian matrix of \( \bar{f} \) is

\[
D^2 \bar{f}(y, \rho, E) = D^2 f(y) + \rho I
\]

where \( I \) denotes the unit matrix. Let \( \lambda_{\min}(y) \) denote the minimum eigenvalue of \( D^2 f(y) \). From [27], it is easy to verify that \( \bar{f} \) is convex on \( E \), if we choose

\[
\rho \geq - \min_{y \in E} \lambda_{\min}(y).
\]

Note that determining the exact value of \( \min_{y \in E} \lambda_{\min}(y) \) would lead to a global optimization problem. It is easy to verify that \( \bar{f} \) is a convex overestimator of \( f \) on \( E \) if

\[
\rho \geq \max \left\{ 0, - \min_{y \in E} \lambda_{\min}(y) \right\}.
\]

The distance between \( \bar{f} \) and \( f \) is defined as

\[
d_{\alpha BB}(y; \rho, E) := \bar{f}(y, \rho, E) - f(y) = \phi(y; \rho, E).
\]
3 Reformulation and Relaxation

For \( Y = [\underline{Y}, \overline{Y}] \), \( Y < \overline{Y} \in \mathbb{R}^m \), \( K^n = \{1, \ldots, n\} \), we let \( \mu^m \) be the Lebesgue measure on \( \mathbb{R}^m \), and \( \Upsilon^n = \{Y^{k_n}, k_n \in K^n\} \) be a subdivision of \( Y \), here \( Y^{k_n} = [\underline{Y}^{k_n}, \overline{Y}^{k_n}] \). That is \( Y = \bigcup_{k_n=1}^n Y^{k_n} \), and \( \mu^m(Y^{k_n} \cap Y^{j_n}) = 0 \) hold for all \( k_n, j_n \in K^n \), \( k_n \neq j_n \). For any \( Y^{k_n} \), consider the following problem

\[
\begin{align*}
\min_y & \quad f(x, y) \\
\text{s.t.} & \quad y \in Y^{k_n}.
\end{align*}
\]  

Let \( \omega^{k_n}(x) \) be the optimal solution set of (7). A trivial but very useful observation is that, for any \( K^n \) the optimal solution set \( \psi(x) \) of original follower’s problem (2) can be formulated equivalently as the the following set

\[
\psi(x) = \psi_{K^n}(x) = \left\{ y \in Y : f(x, y) \leq f(x, v_{k_n}), \quad v_{k_n} \in \omega^{k_n}(x), \quad \forall k_n \in K^n \right\}.
\]  

That is, bilevel problem (1)-(2) is reformulated equivalently as the following bilevel multi-follower programming problem (9)-(7).

\[
\begin{align*}
\min_{(x, y, v_{k_n}, k_n \in K^n)} & \quad F(x, y) \\
\text{s.t.} & \quad G(x) \leq 0, \\
& \quad f(x, y) \leq f(x, v_{k_n}), \\
& \quad y \in \psi(x), \quad v_{k_n} \in \omega^{k_n}(x), \quad \forall k_n \in K^n,
\end{align*}
\]

here, \( \omega^{k_n}(x) \) is the solution set of problem (7). It is easy to verify the following Proposition.

**Proposition 3.1** If \( (x, y) \) is a global (local) optimal solution of bilevel problem (1)-(2), then there exists \( (v_{k_n}, k_n \in K^n) \) such that \( (x, y, v_{k_n}, k_n \in K^n) \) is a global (local) one of bilevel multi-follower programming problem (9)-(7). The reverse is also true, that is, if \( (\overline{x}, \overline{y}, v_{k_n}, k_n \in K^n) \) is a global (local) optimal solution of problem (9)-(7), then \( (\overline{x}, \overline{y}) \) is a global (local) one of problem (1)-(2).

**Proof** According to (8), it is obvious that problem (7) can be equivalent represented as following

\[
\begin{align*}
\min_{(x, y, v_{k_n}, k_n \in K^n)} & \quad F(x, y) \\
\text{s.t.} & \quad G(x) \leq 0, \\
& \quad y = \psi(x) = \psi_{K^n}(x).
\end{align*}
\]

Since problem (10) and problem (1) has the same objective function and constraint set, it is easy to obtain the result. \( \square \)

Based on such an observation, we will construct the convex relaxation problems of the follower’s problem on each constraint set \( Y^{k_n} \). Then we solve the resulting approximate problems and refine the subdivisions adaptively to obtain better approximate solutions of the original problem.
3.1 Approximation Method

Since \( X = \{ x \in \mathbb{R}^n : G(x) \leq 0 \} \), and \( Y^{k_n} \) is a subset of \( Y \), the convex overestimator with respect to \( y \) of \( f \) on \( X \times Y^{k_n} \) can be defined by

\[
f^{k_n}(x, y) = f(x, y) + \phi(y; \rho_{k_n}, Y^{k_n}),
\]

where, \( \phi(y; \rho_{k_n}, Y^{k_n}) = \frac{\rho_{k_n}}{2} \langle Y^{k_n} - y, Y^{k_n} - y \rangle \), the parameter \( \rho_{k_n} \) satisfies

\[
\rho_{k_n} \geq \max \left\{ 0, -\min_{(x,y) \in X \times Y^{k_n}} \lambda_{\min}(x, y) \right\}, \tag{11}
\]

here \( \lambda_{\min}(x, y) \) is the minimum eigenvalue of \( D^2_y f(x, y) \). It is easy to verify that, there exists a upper bound \( \bar{\rho} \) of \( \rho_{k_n} \), that is

\[
\bar{\rho} \geq \max \left\{ 0, -\min_{(x,y) \in X \times Y^{k_n}} \lambda_{\min}(x, y) \right\} \geq \max \left\{ 0, -\min_{(x,y) \in X \times Y^{k_n}} \lambda_{\min}(x, y) \right\} \tag{12}
\]

**Remark 3.1** For the case when \( Y \subseteq \mathbb{R} \), the parameter \( \rho_{k_n} \) satisfying

\[
\rho_{k_n} \geq \max \left\{ 0, -\min_{(x,y) \in X \times Y^{k_n}} \lambda_{\min}(x, y) \right\}
\]

will be chosen.

For any \( k_n \in \mathcal{K}^n \), the follower’s problem (7) can be approximated by the following convex parametric programming problem.

\[
\min_z f^{k_n}(x, z) \quad \text{s.t. } z \in Y^{k_n}. \tag{13}
\]

Let \( \omega^{k_n}_{ap}(x) \) be the optimal solution set of (13), if we take

\[
\rho_{k_n} > \max \left\{ 0, -\min_{(x,y) \in X \times Y^{k_n}} \lambda_{\min}(x, y) \right\}
\]

then, the solution of problem (13) is unique, that is \( \omega^{k_n}_{ap}(x) \) is singleton for any \( k_n \in \mathcal{K}^n \). The first-order derivative of \( f^{k_n} \) is

\[
\nabla_y f^{k_n}(x,y) = \nabla_y f(x,y) + \rho_{k_n} y - \rho_{k_n} \frac{Y^{k_n} + Y^{k_n}}{2}.
\]

It is easy to verify that the Slater’s CQ is satisfied for (13) at every \( x \in X \). Combining this with the convexity of \( f^{k_n}(x, \cdot) \), it follows that problem (13) is equal to its KKT conditions

\[
A^{k_n}(x) = \left\{ (z_{k_n}, \lambda_{k_n}, \mu_{k_n}) : 0 \leq \lambda_{k_n} \perp (z_{k_n} - Y^{k_n}) \geq 0, 0 \leq \mu_{k_n} \perp (Y^{k_n} - z_{k_n}) \geq 0 \right\}. \tag{14}
\]
Now the bilevel multi-follower programming (9)-(7) is approximated by the following bilevel multi-follower programming (15)-(13)

$$\min_{(x,y,z_{kn}, k_n \in \mathcal{K}^n)} F(x, y)$$

s.t. $G(x) \leq 0$,

$$f(x, y) \leq f(x, z_{kn}),$$

$y \in Y, z_{kn} \in \omega_{ap}^k(x), \forall k_n \in \mathcal{K}^n,$

(15)

here $\omega_{ap}^k(x)$ is the optimal solution set of problem (13). Replace all the follower’s problems equivalently by their Karush-Kuhn-Tucker conditions, then bilevel multi-follower programming (15)-(13) is equivalently reformulated as the following MPEC problem.

$$(MPEC(\mathcal{K}^n))(\min_{(x, y, z_{kn}, \lambda_{kn}, \mu_{kn}, k_n \in \mathcal{K}^n)} F(x, y),$$

s.t. $G(x) \leq 0, y \in Y$,

$$f(x, y) \leq f(x, z_{kn}),$$

$$(z_{kn}, \lambda_{kn}, \mu_{kn}) \in A^k(x), \forall k_n \in \mathcal{K}^n.$$ (16)

Thus, one can obtain the approximately optimal solution of bilevel problem (1)-(2) by solve $MPEC(\mathcal{K}^n) (16)$. Precisely because of it, we need to discuss the convergence property of the solutions of the sequence bilevel multi-follower programming (15)-(13). Before proceeding further, we consider the following set-valued mapping:

$$\Omega_{\mathcal{K}^n}(x) = \left\{ y \in Y : \begin{array}{l}
 f(x, y) \leq f(x, z_{kn}), \\
 z_{kn} \in \omega_{ap}^k(x), \forall k_n \in \mathcal{K}^n,
\end{array} \right\}$$

(17)

Base on this mapping, for any $\mathcal{K}^n$, bilevel multi-follower programming (15)-(13) can be reformulated as the following bilevel programming problem (18)-(17).

$$\min_{x,y} F(x, y)$$

s.t. $G(x) \leq 0, y \in \Omega_{\mathcal{K}^n}(x).$ (18)

Similarly, it is easy to verify the following proposition which shows that this is an equivalent transformation.

**Proposition 3.2** If $(x, y)$ is a global (local) optimal solution of bilevel problem (18)-(17), then there exists $(z_{kn}, k_n \in \mathcal{K}^n)$ such that $(x, y, z_{kn}, k_n \in \mathcal{K}^n)$ is a global (local) one of bilevel multi-follower programming (15)-(13). The reverse is also true, that is if $(\tilde{x}, \tilde{y}, \tilde{z}_{kn}, k_n \in \mathcal{K}^n)$ is a global (local) optimal solution of problem (15)-(13), then $(\tilde{x}, \tilde{y})$ is a global (local) one of problem (18)-(17).
3.2 Perturbation Analysis of Bilevel Programs

For any $K^n$, we let $\gamma^n = \max\{\|\mathbf{Y}^k_n - \mathbf{Y}^k_n\| : k_n \in K^n\}$. Based on the following assumption, we discuss the relationship between $\psi_{K^n}(x)$ and $\Omega_{K^n}(x)$.

**Assumption 3.1** In this paper we assume that $\gamma^n \to 0$ as $n \to \infty$.

**Lemma 3.1** For any $x \in X$, any $K^n$, one has $\psi_{K^n}(x) \subseteq \Omega_{K^n}(x)$.

**Proof** For any $x \in X$ and $k_n \in K^n$, since $v_{k_n}$ is a minimizer of $f$ on $Y^k_n$ and $z_{k_n} \in Y^k_n$, it follows that
\[
 f(x, v_{k_n}) \leq f(x, z_{k_n}). \tag{19}
\]
From (8), it is easy to verify that for any $y \in \psi_{K^n}(x)$, one has
\[
 f(x, y) \leq f(x, v_{k_n}) \quad \forall k_n \in K^n. \tag{20}
\]
Combining (20) with (19) it follows that
\[
 f(x, y) \leq f(x, z_{k_n}) \quad \forall k_n \in K^n
\]
that is $y \in \Omega_{K^n}(x)$. \hfill \Box

Before discussing the convergence of the optimal solution set of sequence of bilevel programming problem (18)-(17), we first consider the lower and upper convergence of the sequence $\{\Omega_{K^n}\}$ as $n \to \infty$.

**Proposition 3.3** Suppose that $\psi$ is lower semicontinuous at $\bar{x} \in X$, and Assumption 3.1 holds, then the sequence $\{\Omega_{K^n}\}$ is lower convergent to $\psi$ at $\bar{x}$ as $n \to \infty$.

**Proof** For any sequence $\{x_n\}$ converging to $x$ in $X$, and any $y \in \psi(x)$, from definition 2.3, it follows that there exists sequence $\{y_n\}$ converging to $y$ such that $y_n \in \psi(x_n)$, for $n$ large. Combining this with Lemma 3.1, it is easy to verify that $y_n \in \Omega_{K^n}(x_n)$, that is the sequence $\{\Omega_{K^n}\}$ is lower convergent to $\psi$ at $\bar{x}$ \hfill \Box

**Remark 3.2** Proposition (3.3) needs the lower semicontinuity of $\psi$. If the optimal solution set $\psi(x)$ of the follower’s problem is a singleton at $\bar{x} \in X$, it is easy to verify the lower semicontinuity of $\psi$ at $\bar{x}$. For the case that $\psi(\bar{x})$ has more than one elements, in order to obtain the lower semicontinuity of $\psi$, Zhao [35] suppose that for any $\epsilon > 0$, there exist $\alpha > 0$ and $N \in \mathbb{Z}_+$ such that for any $n > N$, and any $y \in \psi(x_n) + \epsilon$, one has
\[
 f(x_n, y) \geq \varphi(x_n) + \alpha,
\]
here $\varphi(x)$ is the optimal value function of the follower’s problem.

We now turn to consider the upper convergence of $\{\Omega_{K^n}\}$. Before approaching this problem we first prove the following preparation lemmas.
Lemma 3.2 Suppose that $f$ is a Lipschitz continuous function on $X \times Y$ with Lipschitz constant $L$, then for any sequence $\{x_n\}$ converging to $\bar{x}$ and any $\delta > 0$, there exists a $N_1 \in \mathbb{Z}_+$ such that for any $n > N_1$, and $k_n \in K^n$ one has

$$|f(x_n, z_{k_n}) - f(\bar{x}, v_{k_n})| \leq \frac{\delta}{8}. $$

Proof For any $k_n \in K^n$, from Assumption 3.1 it is easy to verify that $\|z_{k_n} - v_{k_n}\| \to 0$ as $n \to \infty$. Combining this with the convergence of sequence $\{x_n\}$, it can easily be seen that for any $\delta > 0$ there exists $N_1 \in \mathbb{Z}_+$ such that, for any $n > N_1$ one has

$$\|x_n - \bar{x}\| \leq \frac{\delta}{16L}, \text{ and } \|z_{k_n} - v_{k_n}\| \leq \frac{\delta}{16L}. \tag{21}$$

According to the Lipschitz continuous property of $f$, it follows that

$$|f(x_n, z_{k_n}) - f(\bar{x}, v_{k_n})| \leq L \left(\|x_n - \bar{x}\| + \|z_{k_n} - v_{k_n}\|\right) \tag{22}$$

From (21) and (22) it is easy to verify the following result.

$$|f(x_n, z_{k_n}) - f(\bar{x}, v_{k_n})| \leq L(\|x_n - \bar{x}\| + \|z_{k_n} - v_{k_n}\|) \leq L\left(\frac{\delta}{16L} + \frac{\delta}{16L}\right) = \frac{\delta}{8}. \tag{23}$$

Lemma 3.3 Suppose that $f$ is a Lipschitz continuous function on $X \times Y$ with Lipschitz constant $L$, and $\bar{y} \in \psi(\bar{x}) = \psi_{K^n}(\bar{x})$. For any $\delta > 0$, any sequence $\{x_n\}$ converging to $\bar{x}$, there exists $N_2 \in \mathbb{Z}_+$ such that for any $n > N_2$ one has

$$|\min_{k_n \in K^n} f(x_n, z_{k_n}) - f(\bar{x}, \bar{y})| < \frac{\delta}{8}. \tag{24}$$

Proof For any $k_n \in K^n$, suppose that $v_{k_n}$ is a global optimal solution of $\min_{v \in Y_{K^n}} f(\bar{x}, v)$, and $z_{k_n}$ is a global one of $\min_{z \in Y_{K^n}} f^k(\bar{x}, z)$. Moreover we suppose that $k^n_*$ is a global optimal index of the $\min_{k_n \in K^n} f(\bar{x}, v_{k_n})$, and $k^{**}_n$ is also a global one of $\min_{k_n \in K^n} f(x_n, z_{k_n})$. According to the definition of $k^*$ and $k^{**}$, it follows that

$$f(x_n, z_{k^n_*}) \geq f(x_n, z_{k^{**}_n}), \tag{25}$$

$$f(\bar{x}, v_{k^n_*}) \leq f(\bar{x}, v_{k^{**}_n}). \tag{26}$$

In order to prove this theorem, we need compare $f(x_n, z_{k^n_*})$ with $f(\bar{x}, v_{k^n_*})$, and $f(x_n, z_{k^{**}_n})$ with $f(\bar{x}, v_{k^{**}_n})$. Obviously, there are four different cases:

Case (1) if

$$f(x_n, z_{k^n_*}) \geq f(\bar{x}, v_{k^n_*}), \tag{27}$$

$$f(x_n, z_{k^{**}_n}) \geq f(\bar{x}, v_{k^{**}_n}). \tag{28}$$
Combining (25)(26)(27) and (28), it is easy to verify the following inequation.
\[
f(x, v_{k_n^*}) \leq f(\bar{x}, v_{k_n^*}) \leq f(x, z_{k_n^*}).
\] (29)

Since \( f \) is a Lipschitz continuous function on \( X \times Y \), according to Lemma 3.2 it follows that, for \( k_n^* \) one has
\[
|f(\bar{x}, v_{k_n^*}) - f(x, z_{k_n^*})| \leq L(\|x_n - \bar{x}\| + \|v_{k_n^*} - z_{k_n^*}\|).
\] (30)

Now that \( \|x_n - \bar{x}\| \to 0 \) and \( \|v_{k_n^*} - z_{k_n^*}\| \to 0 \) as \( n \to \infty \), combining this with (30) and (29), it is easy to see that there exists \( N_2 \in \mathbb{Z}_+ \) such that for any \( n > N_2 \), we have
\[
|\min_{k_n \in K^n} f(x_n, z_{k_n}) - f(\bar{x}, \bar{y})| = |f(x_n, z_{k_n^*}) - f(\bar{x}, v_{k_n^*})| \\
\leq |f(\bar{x}, v_{k_n^*}) - f(x, z_{k_n^*})| \leq \frac{\delta}{8}.
\]

Case (II), if
\[
f(x_n, z_{k_n^*}) \leq f(\bar{x}, v_{k_n^*}),
\] (31)
\[
f(x_n, z_{k_n^*}) \geq f(\bar{x}, v_{k_n^*}).
\] (32)

According to (25)(26)(31) and (32), it is easy to verify the following inequation,
\[
f(\bar{x}, v_{k_n^*}) \leq f(x_n, z_{k_n^*}) \leq f(x_n, z_{k_n^*}) \leq f(\bar{x}, v_{k_n^*}).
\] (33)

The following equation can be easily proved by (33):
\[
f(x_n, z_{k_n^*}) = f(\bar{x}, v_{k_n^*}).
\]

That is
\[
|\min_{k_n \in K^n} f(x_n, z_{k_n}) - f(\bar{x}, \bar{y})| = |f(\bar{x}, v_{k_n^*}) - f(x_n, z_{k_n^*})| = 0.
\]

Case (III), if
\[
f(x_n, z_{k_n^*}) \leq f(\bar{x}, v_{k_n^*}),
\] (34)
\[
f(x_n, z_{k_n^*}) \leq f(\bar{x}, v_{k_n^*}).
\] (35)

According to (25)(26)(34) and (35), it is easy to verify that
\[
f(x_n, z_{k_n^*}) \leq f(x_n, z_{k_n^*}) \leq f(\bar{x}, v_{k_n^*}).
\] (36)

Since \( f \) is a Lipschitz continuous function on \( X \times Y \), according to Lemma 3.2 it can easily be checked that
\[
|f(\bar{x}, v_{k_n^*}) - f(x_n, z_{k_n^*})| \leq L(\|x_n - \bar{x}\| + \|v_{k_n^*} - z_{k_n^*}\|).
\] (37)

Duo to \( \|x_n - \bar{x}\| \to 0 \) and \( \|v_{k_n^*} - z_{k_n^*}\| \to 0 \) as \( n \to \infty \), combining this with (37) and (36), it is easy to verify that there exists \( N_2 \in \mathbb{Z}_+ \) such that for any \( n > N_2 \), one has
\[
|\min_{k_n \in K^n} f(x_n, z_{k_n}) - f(\bar{x}, \bar{y})| = |f(\bar{x}, v_{k_n^*}) - f(x_n, z_{k_n^*})| \\
\leq |f(\bar{x}, v_{k_n^*}) - f(x_n, z_{k_n^*})| \leq \frac{\delta}{8}.
\]
Case (IV), if
\[ f(\bar{x}, z_{k^*_n}) \geq f(\bar{x}, v_{k^*_n}), \tag{38} \]
\[ f(\bar{x}, z_{k^*_n}) \leq f(\bar{x}, v_{k^*_n}). \tag{39} \]

According to the fact that \( f \) is a Lipschitz continuous function on \( X \times Y \), and (38) and (39), it is easy to verify that for any \( \delta > 0 \) there exist \( N_2 \in \mathbb{Z}_+ \) such that for any \( n > N_2 \), one has
\[ 0 \leq f(x_n, z_{k^*_n}) - f(\bar{x}, v_{k^*_n}) \leq \frac{\delta}{8}. \tag{40} \]
\[ 0 \leq f(\bar{x}, v_{k^*_n}) - f(x_n, z_{k^*_n}) \leq \frac{\delta}{8}. \tag{41} \]

From (25) it follows
\[ f(x_n, z_{k^*_n}) - f(x_n, z_{k_n}) \leq 0. \tag{42} \]

Combining (42) with (40) it can easily be seen that
\[ f(x_n, z_{k^*_n}) - f(\bar{x}, v_{k^*_n}) \leq \frac{\delta}{8}. \tag{43} \]

In a similar way, from (26) it follows
\[ f(\bar{x}, v_{k^*_n}) - f(\bar{x}, v_{k_n}) \geq 0. \tag{44} \]

Combining (44) with (41) it is easy to show that
\[ f(x_n, z_{k^*_n}) - f(\bar{x}, v_{k_n}) \geq -\frac{\delta}{8}. \tag{45} \]

Finally, form (43) and (45), it is easy to verify that
\[ |f(x_n, z_{k^*_n}) - f(\bar{x}, v_{k_n})| \leq \frac{\delta}{8}. \]

The Lemma is now evident from what we have proved. \( \square \)

**Lemma 3.4** Suppose \( f \) is Lipschitz continuous with respect to \( y \) on \( Y \). For any \( x \in X \), any \( \delta > 0 \), there exist \( N_3 \in \mathbb{Z}_+ \) such that for any \( n > N_3 \), and any \( y \in \Omega_{K^n}(x) \) one has
\[ |f(x, y) - \min_{k_n \in K^n} f(x, z_{k_n})| \leq \frac{\delta}{8}. \]

**Proof** For any \( k_n \in K^n \), we suppose that \( v_{k_n} \) is a global optimal solution of \( \min_{v \in Y \times x} f(x, v) \), and \( z_{k_n} \) is a global one of \( \min_{z \in Y \times x} f^k(x, z) \). Moreover, we suppose that \( k^*_n \) is an optimal solution of \( \min_{k_n \in K^n} f(x, v_{k_n}) \), and \( k^{**}_n \) is an optimal one of \( \min_{k_n \in K^n} f(x, z_{k_n}) \). According to the fact that \( f(x, v_{k_n}) \) is the minimum value of \( f(x, \cdot) \) on \( Y \), it follows that
\[ f(x, v_{k_n}) \leq f(x, z_{k_n}). \tag{46} \]
Combining this with the definition of $k_n^*$, it is straightforward to show that

$$f(x, v_k^*) \leq f(x, z_k^*) \leq f(x, z_k^*).$$  \hspace{1cm} (47)

Since $f$ is Lipschitz continuous with respect to $y$ on $Y$, it follows that there exists a constant $L_2 > 0$ such that

$$|f(x, v_k^*) - f(x, z_k^*)| \leq L_2\|v_k^* - z_k^*\|.$$ \hspace{1cm} (48)

Due to $\|v_k^* - z_k^*\| \to 0$ as $n \to \infty$, it is easy to verify that for any $\delta > 0$, there exists $N_3 \in \mathbb{Z}_+$ such that for any $N > N_3$ one has

$$\|v_k^* - z_k^*\| \leq \frac{\delta}{8L_2}.$$ \hspace{1cm} (49)

From $y \in \Omega_{k_n}(x)$ it follows that

$$f(x, v_k^*) \leq f(x, y) \leq f(x, z_k^*).$$ \hspace{1cm} (50)

Combining (47) (49) and (50) it can easily be checked that

$$|f(x, y) - \min_{k_n \in K_n} f(x, z_{k_n})| = |f(x, y) - f(x, z_k^*)| \leq |f(x, v_k^*) - f(x, z_k^*)| \leq \frac{\delta}{8}.$$ \hspace{1cm} \(\Box\)

In this paper Lemma 3.4 need to be consistent held, that is the following assumption holds.

**Assumption 3.2** For any fixed $\bar{x} \in X$ and $\delta > 0$, there exist $\epsilon > 0$ and $N_4 \in \mathbb{Z}_+$ such that, for any $x \in B(\bar{x}, \epsilon)$, $y \in \Omega_{k_n}(x)$ and any $n > N_4$ one has

$$|f(x, y) - \min_{k_n \in K_n} f(x, z_{k_n})| \leq \frac{\delta}{8}.$$  \hspace{1cm} (51)

With the help of the preceding lemmas and assumption we can now prove the following upper convergence proposition.

**Proposition 3.4** Suppose that $f$ is a Lipschitz continuous function on $X \times Y$ and Assumption 3.2 holds, then for any $\bar{x} \in X$, the sequence $\{\Omega_{k_n}\}$ is upper convergent to $\psi$ at $\bar{x}$ as $n \to \infty$.

**Proof** On the contrary, suppose that there exists sequence $\{x_n\}$ converging to $\bar{x}$, and $\{y_n\}$ converging to $\bar{y}$ with $y_n \in \Omega_{k_n}(x_n)$ such that $\bar{y} \notin \psi(\bar{x})$. Then for any $\bar{y} \in \psi(\bar{x})$, there exists $\delta > 0$ such that

$$f(\bar{x}, \bar{y}) = f(\bar{x}, \bar{y}) = \delta.$$  \hspace{1cm} (51)

According to Assumption 3.1, it is obvious that $\|z_{k_n} - v_{k_n}\| \to 0$, $\|x_n - \bar{x}\| \to 0$, $\|y_n - \bar{y}\| \to 0$ as $n \to \infty$. According to the fact that $f$ is a continuous function on $X \times Y$, it follows that there exist $N_5 \in \mathbb{Z}_+$ such that for any $n > N_5$ one has

$$|f(x_n, y_n) - f(\bar{x}, \bar{y})| < \frac{\delta}{3}.$$
that is
\[ |f(x_n, y_n) - f(\bar{x}, \bar{y})| > \frac{2\delta}{3}, \] (52)

On the other hand, from the triangle inequality, it follows that
\[ |f(x_n, y_n) - f(\bar{x}, \bar{y})| \leq |f(x_n, y_n) - \min_{k_n \in K_n} f(x_n, z_{k_n})| + |\min_{k_n \in K_n} f(x_n, z_{k_n}) - f(\bar{x}, \bar{y})| \] (53)

From Assumption 3.2 and Lemma 3.3, it follows that there exists a positive integer \( \bar{N} = \max\{N_2, N_4\} \) such that inequation (24) and the following (54) hold.
\[ |f(x_n, y_n) - \min_{k_n \in K_n} f(x_n, z_{k_n})| \leq \frac{\delta}{8}. \] (54)

Combine this with (53), one has
\[ |f(x_n, y_n) - f(\bar{x}, \bar{y})| \leq \frac{\delta}{8} \times 2 = \frac{\delta}{4}. \]

This is contrary to (52). \( \square \)

With the help of the preceding propositions and assumption we can now prove the main theorem of this subsection. For convenience, we define the following notations.
\[ \eta_n(x) := \min_{y \in \Omega_{K_n}(x)} F(x, y), \quad \eta(x) := \min_{y \in \psi(x)} F(x, y). \]
\[ \zeta_n := \min_{x \in X} \eta_n(x), \quad \zeta := \min_{x \in X} \eta(x). \]
\[ M_n := \{ x \in X : \eta_n(x) \leq \zeta_n \}, \quad M := \{ x \in X : \eta(x) \leq \zeta \}. \]

The solution sets of BP (18)-(17) and BP (1)-(2) are defined as \( S_n \) and \( S \) respectively. It is clear that \( S_n \) and \( S \) also can be defined by:
\[ S_n := \{ (x, y) : x \in M_n, y \in \Omega_{K_n}(x), \eta_n(x) = F(x, y) \}. \]
\[ S := \{ (x, y) : x \in M, y \in \psi(x), \eta(x) = F(x, y) \}. \]

Lignola and Morgan [18] gave some results for the stability of bilevel programs. Based on them, we obtain the following Theorem 3.1 which shows that, the solutions of a sequence of the bilevel programs (18)-(17) converge to a solution of BP (1)-(2) under some appropriate conditions.

**Theorem 3.1** For any sequence \( \{(x_n, y_n)\} \) such that \( (x_n, y_n) \in S_n \), and \( (x_n, y_n) \to (\bar{x}, \bar{y}) \) as \( n \to +\infty \), we suppose that \( \psi(x) \) is lower semicontinuous at \( \bar{x} \in X \). Assumption 3.1 and Assumption 3.2 hold. Moreover we suppose that \( f \) is a Lipschitz continuous function on \( X \times Y \). Then we have \( (\bar{x}, \bar{y}) \in S \).
Proof Since $\psi(x)$ is lower semicontinuous at $\bar{x} \in X$, and Assumption 3.1 holds, from Proposition 3.3 it follows that the sequence $\{\Omega_{k_n}\}$ is lower convergent to $\psi$ at $\bar{x}$ as $n \to \infty$. According to the fact that $f$ is a Lipschitz continuous function on $X \times Y$, Assumption 3.2 and Proposition 3.4, we know that $\{\Omega_{k_n}\}$ is also upper convergent to $\psi$ at $\bar{x}$. Due to $F$ is a continuous function on $\mathbb{R}^n \times \mathbb{R}^m$, and the sequence $\{\Omega_{k_n}\}$ is lower and upper convergent to $\psi$ at $\bar{x}$, from Proposition 4.4 in [20], it is easy to verify that $\eta_n$ sequentially continuous converges to $\eta$ at $\bar{x}$, and $\bar{x} \in M$. Because $y_n \in \Omega_{k_n}(x_n)$ and $\{\Omega_{k_n}\}$ is upper convergent to $\psi$ at $\bar{x}$, according to Definition 2.2, it follows that $\bar{y} \in \psi(\bar{x})$. Since $\eta_n$ sequentially continuous converges to $\eta$ at $\bar{x}$, it follows from $\eta_n(x_n) = F(x_n, y_n)$ that $\eta(\bar{x}) = F(\bar{x}, \bar{y})$. Thus $(\bar{x}, \bar{y}) \in S$. $\square$

4 Algorithm and Convergence

For the sake of clearness in the narrate, in this section, we simplify the index $k_n$ down to $k$. That is, for $Y = \{Y, \mathcal{Y}, \mathcal{Y} < \mathcal{Y} \in \mathbb{R}^m$, and $\mathcal{K}^n = \{1, \ldots, n\}$, we let $\mathcal{Y}^n = \{Y^k, k \in \mathcal{K}^n\}$ be a subdivision of $Y$, and $Y^k = [\mathcal{Y}^k, \mathcal{Y}^k]$, obviously, $Y = \bigcup_{k=1}^n Y^k$. Therefore the first-order derivative of $f^{k_n}$ is reduced to $f^k$ which is defined as

$$\nabla_y f^k(x, y) = \nabla_y f(x, y) + \rho_k y - \rho_k \frac{\lambda_k + \mathcal{Y}^k}{2}.$$

The MPEC($\mathcal{K}^n$) (16) is reduced to the following MPEC($\mathcal{K}^n$) (55)

$$(\text{MPEC}({\mathcal{K}^n})) \min_{(x,y,z_k,\lambda_k,\mu_k,k \in \mathcal{K}^n)} F(x, y) \text{ s.t. } G(x) \leq 0,$$

$$f(x, y) \leq f(x, z_k),$$

$$y \in Y, (z_k, \lambda_k, \mu_k) \in A^k(x), \forall k \in \mathcal{K}^n,$$

(55)

here $A^k(x)$ is the KKT point set of follower’s problem, that is

$$A^k(x) = \left\{(z_k, \lambda_k, \mu_k) : \begin{aligned}
\nabla_y f^k(x, z_k) - \lambda_k + \mu_k &= 0, \\
0 \leq \lambda_k \perp (z_k - \mathcal{Y}^k) &\geq 0, \\
0 \leq \mu_k \perp (\mathcal{Y}^k - z_k) &\geq 0
\end{aligned} \right\}. \quad (56)$$

Next we consider the refinement algorithm for the lower level constraint set $Y$. Before proceeding further, we give the following definition which is proposed in [27].

Definition 4.1 For a tessellation $\mathcal{Y}^n$, let $e \in Y^k$ with some $k \in \mathcal{K}^n$, and let $\epsilon_{\text{split}} > 0$ be given. Define the index set of coordinate directions along which the distance of $e$ from the boundary $\partial Y^k$ is sufficiently large by

$$P^k := P^k(e) = \left\{j \in \{1, \ldots, m\} : \min\{e_j - \mathcal{Y}^k_j, \mathcal{Y}^k_j - e_j\} > \epsilon_{\text{split}}(\mathcal{Y}^k_j - \mathcal{Y}^k_j) \right\}.$$
In the case \( P^k \neq \emptyset \) choose the coordinate direction \( l = \arg \max_{j \in P^k} (\vec{Y}_j^k - \vec{Y}_j^k) \). Choose the hyperplane normal to the coordinate direction \( l \) through \( e \) to split \( Y^k \) into two boxes \( Y^{k,1} \) and \( Y^{k,2} \). For \( P^k = \emptyset \) the box \( Y^k \) is not split. In this way, the following splitting operator is defined as

\[
S(Y^k, e) := \begin{cases} 
(Y^{k,1}, Y^{k,2}), & \text{if } P^k(e) \neq \emptyset, \\
Y^k, & \text{if } P^k(e) = \emptyset.
\end{cases}
\]

In order to avoid \( P^k(e) = \emptyset \) for any \( e \in Y^k \), we will always assume \( \epsilon_{\text{split}} < \frac{1}{2} \), this assumption is also proposed by Stein and Steuermann in [27]. The choice of the barycenter \( S^k \) of \( Y^k \) as a splitting point, always entails \( P^k = \{1, \ldots, m\} \).

As the index of the coordinate direction corresponding the longest edge of a box \( Y^k \) may never be contained in \( P^k \), we adding the following technique to avoid this degeneration case. Let \( e^k \) be given, \( \varepsilon > 0 \) and

\[
Q^k = \min_{j \in P^k} (\vec{Y}_j^k - \vec{Y}_j^k) / \| \vec{Y}_j^k - \vec{Y}_j^k \|_\infty
\]

be the relation of the shortest edge length with coordinate index in \( P^k \) to the longest edge length of \( Y^k \). In the case \( Q^k < \varepsilon \) a degeneration of a box \( Y^k \) may occur. By setting \( e^k = S^k \) the coordinate index of the longest edges is always contained in \( P^k \) and a degeneration can be avoided.

With the help of the preceding discussion we can now give the refinement algorithm:

**Algorithm 4.1** Let \( e \in Y^k \), \( k^* \in K^n = \{1, \ldots, n\} \),

- **Step1.** Compute the barycenter \( S^{k^*} \) of \( Y^{k^*} \). Compute

\[
Q^{k^*} = \min_{j \in P^{k^*}} (\vec{Y}_j^{k^*} - \vec{Y}_j^{k^*}) / \| \vec{Y}_j^{k^*} - \vec{Y}_j^{k^*} \|_\infty
\]

- **Step2.** If \( Q^{k^*} < \varepsilon \), then set \( e = S^{k^*} \). If \( P^{k^*} \neq \emptyset \), then compute

\[
(Y^{k^*,1}, Y^{k^*,2}) = S(Y^{k^*}, e),
\]

compute \( \rho^{k^*,1}, \rho^{k^*,2} \leq \rho^{k^*} \) on \( Y^{k^*,1}, Y^{k^*,2} \) and set

\[
\begin{align*}
&f^{k^*,1}(x, y) = f(x, y) + \rho^{k^*,1} (\vec{Y}^{k^*,1} - y, \vec{Y}^{k^*,1} - y), \\
&f^{k^*,2}(x, y) = f(x, y) + \rho^{k^*,2} (\vec{Y}^{k^*,2} - y, \vec{Y}^{k^*,2} - y).
\end{align*}
\]

Generate variable \((z_{k^*,i}, \lambda_{k^*,i}, \mu_{k^*,i}), i = 1, 2\), then the KKT conditions of parametric optimization problem

\[
\min_{y \in Y^{k^*}} f^{k^*,i}(x, y) \text{ is}
\]

\[
A^{k^*,i}(x) = \left\{ (z_{k^*,i}, \lambda_{k^*,i}, \mu_{k^*,i}) : \begin{array}{l}
\nabla_y f^{k^*,i}(x, z_{k^*,i}) - \lambda_{k^*,i} \perp (z_{k^*,i} - \vec{Y}^{k^*,i}) \geq 0, \\
0 \leq \mu_{k^*,i} \perp (\vec{Y}^{k^*,i} - z_{k^*,i}) \geq 0,
\end{array} \right\}
\]

(57)
Step3. Replace \((z_{k^*}, \lambda_{k^*}, \mu_{k^*})\) by \((z_{k^*,-1}, \lambda_{k^*,-1}, \mu_{k^*,-1})\) and \((z_{k^*,-2}, \lambda_{k^*,-2}, \mu_{k^*,-2})\); Replace \(f(x, y) \leq f(x, z_k)\) by \(f(x, y) \leq f(x, z_{k^*,-1})\) and \(f(x, y) \leq f(x, z_{k^*,-2})\); Replace \(A^{k^*}(x)\) by \(A^{k^*-1}(x)\) and \(A^{k^*-2}(x)\). Set \(n = n + 1\).

In what follows, we give the adaptive convexification algorithm based on the refinement algorithm.

**Algorithm 4.2** Determine a tessellation \(T_n^*\) of \(Y\) with some \(K^* = \{1, ..., n\}\) as well as \(\rho_k \leq \rho\) on \(Y_k\), \(k \in K^*\). And set \(N=1\);

Step1. By solving \(MPEC(K_n^*)\) (55), compute an optimal solution \((x_N, y_N)\) of bilevel programming problem (18)-(17), with \((z_k, k \in K^*)\) and multipliers \((\lambda_k, \mu_k, k \in K^*)\).

Step2. For every \(k \in K^*\) refine \(z_k\) by Algorithm 4.1. Let \(N=N+1\), go to Step1.

**Remark 4.1** We obtain an optimal solution \((x_N, y_N, z_k, \lambda_k, \mu_k, k \in K^*)\) of \(MPEC(K_n^*)\) (55) by Step 1 of Algorithm 4.2. Since all the follower problems (13) are convex, it can easily be verified that \((x_N, y_N)\) is an optimal solution of bilevel programming problem (18)-(17).

Next, we consider the convergence of the upper algorithm. With the help of the discussion in Section 3, it is not difficult to obtain the following convergence theorem.

**Theorem 4.1** Suppose that \(\psi(x)\) is lower semicontinuous at \(x \in X\), Assumption 3.1 and Assumption 3.2 hold. Moreover we suppose that \(f\) is a Lipschitz continuous function on \(X \times Y\). If \((x_N, y_N)\) generated by Step 1 of Algorithm 4.2 is an optimal solution of \(MPEC(K_n^*)\) (55), then the sequence \(\{(x_N, y_N)\}\) is convergent to an optimal solution of bilevel problem (1)-(2) as \(N \to \infty\).

**Proof** According to Remark 4.1 it follows that \((x_N, y_N)\) is an optimal solution of bilevel programming problem (18)-(17). Since \(\psi(x)\) is lower semicontinuous at \(x \in X\), and Assumption 3.1 and Assumption 3.2 hold, and \(f\) is a Lipschitz continuous function on \(X \times Y\), from Theorem 3.1, it is easy to verify that sequence \(\{(x_N, y_N)\}\) is convergent to an optimal solution of bilevel problem (1)-(2) as \(N \to \infty\). \(\square\)

5 Numerical Experiments

For the numerical illustrations in this section we implement Algorithm 4.2 in MATLAB 2016b and use fmincon of Optimization Toolbox with default tolerance to solve the subproblem in step 1. All the following experiments were run on 2000 MHz Inter(R) Core i5 processor.

**Example 5.1** Considering the following bilevel programming problem with a nonconvex inner objective parameterized in \(x\).

\[
\begin{align*}
\min_{x_1, x_2, y_1, y_2} & \quad (x_2 - 2)^2 + (x_1 - 1)^2 + y_2 x_1^2 + y_1 x_2 \\
\text{s.t.} \quad & (x_1, x_2) \in [-2, 2] \times [-3, 3], (y_1, y_2) \in \psi_1(x)
\end{align*}
\] (58)
here, $\psi_1(x)$ is the global optimal solution set of the following problem.

$$\min_y x_1 \frac{y^2}{3} + \frac{y^3}{3}$$

s.t. $(y_1, y_2) \in [-2, 2] \times [-2, 2].$ \hfill (59)

This problem has an unique optimal solution $(\bar{x}_1, \bar{x}_2, \bar{y}_1, \bar{y}_2) = (2, 3, -2, -2)$ with an objective value of $-12$. We give a initial tessellation $T^4$ of $Y = [-2, 2] \times [-2, 2]$ as follows

$$T^4 = \{ [-2, 0] \times [-2, 0], [-2, 0] \times [0, 2], [0, 2] \times [-2, 0], [0, 2] \times [0, 2]\}.$$

Table 1 Numerical results for Example 5.1 by using Algorithm 4.2

<table>
<thead>
<tr>
<th>Iterations</th>
<th>N</th>
<th>Solution $(x_1, x_2, y_1, y_2)$</th>
<th>Real solution</th>
<th>fval</th>
<th>Cumulative CPU(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.00-2.9652, -2.00, 2.00</td>
<td>-</td>
<td>-11.9894</td>
<td>1.670</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.00-2.9921, -2.00, 2.00</td>
<td>-</td>
<td>-12.0000</td>
<td>4.467</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.00-2.9970, -2.00, 2.00</td>
<td>-</td>
<td>-12.0000</td>
<td>8.718</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2.00-2.9997, -2.00, 2.00</td>
<td>(2.3, -2, -2)</td>
<td>-12.0000</td>
<td>17.181</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows the numerical results for Example 5.1 by using Algorithm 4.2. In this table, Iterations $N$ represents the number of iterations, fval is the objective function value in each iteration. Solution $(x_1, x_2, y_1, y_2)$ stand for the optimal solution obtained by Step1 of Algorithm 4.2. Cumulative CPU is the cumulative CPU times, that is the interval of time between the application starts and the N times-iteration.

Example 5.2 (Example 3.9 [24]) Considering the following bilevel problem with a nonconvex monomial inner objective

$$\min_{x, y} x$$

s.t. $-x + y \leq 0,$

$$x \in [-10, 10], y \in \psi_2(x)$$ \hfill (60)

here, $\psi_2(x)$ is the global optimal solution set of the following problem.

$$\min_{y \in [-1, 1]} y^3.$$ \hfill (61)

It has the unique optimal solution $x = -1, y = -1$ with an objective value of $-1$. For this example the outer variable $x$ does not add much complication to the analysis, since the inner program is not parameterized by it.

Example 5.3 (Example 3.14 [24]) Considering the following bilevel problem

$$\min_{x, y} (x - \frac{1}{4})^2 + y^2$$

s.t. $x \in [-1, 1], y \in \psi_3(x)$ \hfill (62)
here, $\psi_3(x)$ is the global optimal solution set of the following problem.

$$\min_{y \in [-1,1]} \frac{y^3}{3} - xy.$$  \hspace{1cm} (63)

This problem has the unique optimal solution $x = \frac{1}{4}$, $y = \frac{1}{2}$ with an objective value of $\frac{1}{4}$. By the Slater’s CQ the KKT conditions are necessary. But due to nonconvexity of the objective function they are not sufficient for a local/global minimum.

**Example 5.4** (Example 3.15 [24]) Considering the following bilevel problem

$$\min_{x,y} x + y \quad \text{s.t. } x \in [-1,1], y \in \psi_4(x)$$  \hspace{1cm} (64)

here, $\psi_4(x)$ is the global optimal solution set of the following problem.

$$\min_{y \in [-1,1]} \frac{xy^2}{3} - \frac{y^3}{3}.$$  \hspace{1cm} (65)

It has the unique optimal solution $x = -1$, $y = 1$ with an objective value of 0. Also by the Slater’s CQ the KKT conditions are necessary. But due to nonconvexity of the objective function they are not sufficient for a local/global minimum.

**Example 5.5** (Example 3.11 [24]) Considering the following bilevel problem

with a nonconvex inner objective parameterized in $x$.

$$\min_{x,y} y \quad \text{s.t. } x \in [-1,1], y \in \psi_5(x)$$  \hspace{1cm} (66)

here, $\psi_5(x)$ is the global optimal solution set of the following problem.

$$\min_{y \in [-1,1]} 16y^4 + 2y^3 - 8y^2 - \frac{3}{2}y + \frac{1}{2}.$$  \hspace{1cm} (67)

It has the unique optimal solution $x = 0$, $y = 0.5$ with an objective value of 0.5.

### Table 2

<table>
<thead>
<tr>
<th>Example</th>
<th>Iterations</th>
<th>fval</th>
<th>Numerical result</th>
<th>Real solution</th>
<th>CPU(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>2</td>
<td>-1.0000</td>
<td>(-1.0000,-1.0000)</td>
<td>(-1.0000,-1.0000)</td>
<td>2.558</td>
</tr>
<tr>
<td>5.3</td>
<td>3</td>
<td>0.2504</td>
<td>(0.2504,0.5004)</td>
<td>(0.2500, 0.5000)</td>
<td>5.377</td>
</tr>
<tr>
<td>5.4</td>
<td>2</td>
<td>0.0000</td>
<td>(1.0000,-1.0000)</td>
<td>(1.0000,-1.0000)</td>
<td>3.501</td>
</tr>
<tr>
<td>5.5</td>
<td>1</td>
<td>0.5000</td>
<td>(0.0000,0.5000)</td>
<td>(0.0000, 0.5000)</td>
<td>2.510</td>
</tr>
</tbody>
</table>

Table 2 shows the numerical results for Example 5.2, 5.3, 5.4, 5.5 by using Algorithm 4.2. In this tables, Iterations $N$ represents the total number of
iterations, \( fval \) is the objective function value when the algorithms terminate. \textit{Numerical result} stands for the optimal solution obtained by Algorithm 4.2. \textit{Real solution} stands for the accurate optimum solutions of the four examples. \textit{CPU(s)} is the total CPU time.

\textbf{Example 5.6} (Example 4.1 [21]) Considering the following Mirrlees’ problem which has a nonconvex inner objective parameterized in \( x \).

\[
\min_{x,y} (x - 2)^2 + (y - 1)^2 \\
\text{s.t. } y \in \psi_6(x)
\]  

(68)

here, \( \psi_6(x) \) is the global optimal solution set of the following problem.

\[
\min_y -\exp[-(y + 1)^2] - \exp[-(y - 1)^2].
\]  

(69)

As shown by Mirrlees [23], at \( \bar{x} = 1 \), both \( \bar{y}_1 \approx 0.9575 \) and \( \bar{y}_2 \approx -0.9575 \) are optimal solutions of the lower level program (69). And the optimal solution of Mirrlees’ problem is \( (\bar{x}, \bar{y}) = (1, 0.9575) \). Note that the solution of Mirrlees’ problem does not change if we add the constraint \( x \in [-1, 2], y \in [-2, 2] \) into the problem. Hence, \( (\bar{x}, \bar{y}) = (1, 0.9575) \) is the optimal solution to the bilevel program

\[
\min_{x,y} (x - 2)^2 + (y - 1)^2 \\
\text{s.t. } x \in [-1, 2], y \in \psi_7(x)
\]  

(70)

here, \( \psi_7(x) \) is the global optimal solution set of the following problem.

\[
\min_{y \in [-2, 2]} -x \exp[-(y + 1)^2] - \exp[-(y - 1)^2]
\]  

(71)

\begin{table}[h]
\begin{center}
\begin{tabular}{llllll}
\hline
Iterations & N & Solution (\( x, y \)) & Real solution & fval & Cumulative CPU(s) \\
\hline
9 & 9 & (1.9265,-0.8888) & - & 3.5729 & 1.570 \\
10 & 10 & (1.9636,-0.9448) & - & 3.7836 & 1.871 \\
11 & 11 & (0.9990,0.9576) & - & 1.0038 & 2.510 \\
12 & 12 & (0.9845,0.9583) & - & 1.0330 & 3.370 \\
13 & 13 & (0.9878,0.9581) & - & 1.0264 & 4.489 \\
14 & 14 & (0.9799,0.9585) & - & 1.0424 & 5.697 \\
15 & 15 & (0.9984,0.9576) & - & 1.0050 & 6.918 \\
16 & 16 & (0.9993,0.9575) & (1,0.9575) & 1.0032 & 8.271 \\
\hline
\end{tabular}
\end{center}
\end{table}

\textbf{Table 3} Numerical results for Example 5.6 by using Algorithm 4.2

Table 3 shows the numerical results for Example 5.6 by using Algorithm 4.2. \((x, y)\) stand for the optimal solution obtained by Step1 of Algorithm 4.2. From this table we can see that the approximate solution \((0.9993,0.9575)\) is obtained after sixteen iterations.
6 Conclusions

We have presented a numerical algorithm for bilevel programs with a nonconvex lower level program. The key idea of the algorithm is to find an approximate function of the follower objective function on every $Y^k$ of tessellation $\Upsilon^n$. The convergence theorems of the solutions of the sequence approximate problems are obtained here. In the presented algorithm we used the optimal solution of the approximate follower’s problems in every $Y^k$ or the barycenter of $Y^k$ as splitting points. Maybe it is not the optimum strategy. We leave these questions for future research.

Acknowledgements This work was supported by the Natural Science Foundation of China (No. 11871383), High Level Introduction of Talent Research Start-up Fund (No.1856009) and Scientific and Technological Research Program of Chongqing Municipal Education Commission (No.KJQN201800810). The author is grateful to the two anonymous referees for their helpful comments and suggestions.

References