

A novel approach for bilevel programs based on Wolfe duality*

Yuwei Li[†], Gui-Hua Lin[‡], Jin Zhang[§], Xide Zhu[¶]

February 16, 2021

Abstract. This paper considers a bilevel program, which has many applications in practice. To develop effective numerical algorithms, it is generally necessary to transform the bilevel program into a single-level optimization problem. The most popular approach is to replace the lower-level program by its KKT conditions and then the bilevel program can be transformed into a mathematical program with equilibrium constraints (MPEC for short). However, the MPEC does not satisfy the Mangasarian-Fromovitz constraint qualification at any feasible point so that the well-developed nonlinear programming theory may not be applied to MPECs directly. In this paper, we apply the Wolfe duality to show that, under very mild conditions, the bilevel program is equivalent to a new single-level reformulation (WDP for short) in the globally or locally optimal sense. We give an example to show that WDP may satisfy the Mangasarian-Fromovitz constraint qualification at its feasible points and hence WDP is no longer an MPEC. We give some properties of the WDP reformulation and the relations between the WDP and MPEC reformulations. Furthermore, we propose a relaxation method for solving WDP and investigate its limiting behavior. Comprehensive numerical experiments indicate that, although solving WDP directly did not perform very well in our tests, the relaxation method based on the WDP reformulation was quite efficient.

Key Words. Bilevel programming, Wolfe duality, Constraint qualification, Relaxation method, Global convergence.

2010 Mathematics Subject Classification. 90C30, 90C33, 90C46.

1 Introduction

Bilevel programs are hierarchical optimization problems which contain some parameterized optimization or equilibrium problems as constraints. This type of leader-follower game problems can be traced back to the model introduced firstly by von Stackelberg [40] and they have

*This work was supported in part by NSFC (Nos. 12071280, 11901380, 11971220).

[†]School of Management, Shanghai University, Shanghai 200444, China. E-mail: yuwei_li@shu.edu.cn.

[‡]Corresponding author. School of Management, Shanghai University, Shanghai 200444, China. E-mail: guihualin@shu.edu.cn.

[§]Department of Mathematics, SUSTech International Center for Mathematics, Southern University of Science and Technology, Shenzhen 518055, China. E-mail: zhangj9@sustech.edu.cn.

[¶]School of Management, Shanghai University, Shanghai 200444, China. E-mail: xidezhu@shu.edu.cn.

been applied to many fields such as game theory, industrial engineering, economic planning, management science, and so on [1, 6, 7, 13, 29]. For more details about their developments, we refer the reader to the monographs and survey papers [1, 3, 6, 7].

In this paper, we consider the optimistic bilevel program

$$\begin{aligned} \text{(BP)} \quad & \min && F(x, y) \\ & \text{s.t.} && x \in X, y \in S(x). \end{aligned}$$

Here, $F : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$ denotes the leader's objective function, $X \subseteq \mathbb{R}^n$ denotes the leader's strategy set, and $S(x) \subseteq \mathbb{R}^m$ denotes the follower's optimal reaction set, which is the optimal solution set of the parameterized program

$$\begin{aligned} \text{(P}_x\text{)} \quad & \min_y && f(x, y) \\ & \text{s.t.} && g(x, y) \leq 0, h(x, y) = 0, \end{aligned}$$

where $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$, $g : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^p$, and $h : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^q$. Throughout the paper, we denote by $Y(x) := \{y \in \mathbb{R}^m : g(x, y) \leq 0, h(x, y) = 0\}$ the feasible region of (P_x) and assume that F is continuously differentiable, $\{f, g, h\}$ are all twice continuously differentiable, and $S(x) \neq \emptyset$ for each $x \in X$. (BP) describes the case where the leader can anticipate how the follower makes its decisions. In practice, there are other cases where the leader and the follower are not cooperative and, in these cases, the leader may make decisions in pessimistic or eclectic manners. See [1, 7, 11, 19, 22, 24, 25, 30] for more details about optimistic and pessimistic bilevel models.

As is well-known to us, due to the hierarchical structure, bilevel programs are very difficult to solve. Actually, even for the linear cases, bilevel programs have been shown to be strong NP-hard. From the perspective of continuous optimization, to develop effective numerical algorithms, it is generally necessary to transform bilevel programs into single-level optimization problems. Along this way, there have been presented several reformulations for bilevel programs, which include

- ◇ the reformulation based on lower-level solution functions,
- ◇ the reformulation based on lower-level optimal value functions,
- ◇ the reformulation based on lower-level optimality conditions,
- ◇ the reformulation based on lower-level Lagrangian duality.

See Section 2 given below for details about these reformulations.

Among the reformulations for bilevel programs mentioned above, the first one requires the lower-level solution functions to be known and, in such cases, it is generally neither differentiable nor convex even if both the upper-level and lower-level programs are differentiable and convex. For the second reformulation, the lower-level optimal value functions are

usually difficult to evaluate because they do not have analytic expressions in general. The third reformulation is the most popular in the literature and it is known as the mathematical program with equilibrium constraints (MPEC). However, MPECs are highly nonconvex with combinatorial structure and, in theory, MPECs do not satisfy the Mangasarian-Fromovitz constraint qualification (MFCQ) at any feasible point so that the well-developed nonlinear programming theory cannot be applied to MPECs directly. The fourth reformulation uses the so-called regularized constrained Lagrangian dual function, but it has some faults as follows:

- The regularized constrained Lagrangian dual function used in the duality-based reformulation is defined by minimizing some strongly convex function and hence it generally has no analytic expressions.
- The lower-level feasible region is assumed to be uniformly contained within the interior of some compact and convex set, which is too stringent.
- If the lower-level objective function is convex but not strictly convex, an approximation parameter is introduced and required to tend to 0 in solving the original bilevel program.

In this paper, we present a new single-level reformulation for (BP) with better structure and properties than the existing ones. Our strategy is to use the Wolfe duality for constrained optimization problems to deal with the lower-level program. Compared with the existing reformulations introduced above, the new reformulation based on lower-level Wolfe duality does not involve any function without analytic expression and, more importantly, it may satisfy the MFCQ at its feasible points, which means that it is no longer an MPEC.

The paper is organized as follows. In Section 2, we first recall the existing single-level reformulations for (BP) and then introduce some Wolfe duality theorems for constrained optimization problems. In Section 3, we apply the Wolfe duality theorems to present our new single-level reformulation for (BP). We show their equivalence from both globally and locally optimal senses. In Section 4, we make a comparison between the new reformulation and the most popular MPEC reformulation. In particular, we first give an example to show that the new reformulation is no longer an MPEC and then investigate the relations between two reformulations. In Section 5, we present a relaxation scheme for solve the new reformulation and establish a convergence analysis. In Section 6, we report our numerical experiments on a number of linear bilevel programs generated randomly. In Section 7, we make some concluding remarks.

Throughout, for a given differentiable function $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$, we use $\nabla G(x) \in \mathbb{R}^{n \times m}$ to denote the transposed Jacobian of G at x . When $m = 1$, $\nabla G(x) \in \mathbb{R}^n$ stands for the gradient of G at x . Given two vectors $a, b \in \mathbb{R}^n$, $a \perp b$ means $a^T b = 0$.

2 Preliminaries

In this section, we introduce some existing reformulations for bilevel programs and the Wolfe duality for constrained optimization problems.

2.1 Existing reformulations for bilevel programs

As mentioned before, in order to develop effective numerical algorithms, we need to transform bilevel programs into single-level optimization problems. In this section, we briefly recall the existing reformulations for bilevel programs.

2.1.1 Reformulation based on lower-level solution functions

If the lower-level program (P_x) has a unique optimal solution $y(x)$ for any given $x \in X$, the bilevel program (BP) can be reformulated equivalently into the single-level optimization problem

$$\begin{aligned} \text{(SP)} \quad & \min \quad F(x, y(x)) \\ & \text{s.t.} \quad x \in X. \end{aligned}$$

Naturally, this reformulation requires stringent conditions to ensure the existence of lower-level optimal solutions and, under the condition that the solution function $y(x)$ is known, the objective function in the above problem is generally neither differentiable nor convex, even if both the upper-level and lower-level programs are differentiable and convex. Therefore, this approach is usually hard to apply in solving bilevel programs.

2.1.2 Reformulation based on lower-level optimal value functions

Another approach for bilevel programs is to use the lower-level optimal value function defined by $V(x) := \min_{y \in Y(x)} f(x, y)$. Then, the bilevel program (BP) can be reformulated equivalently into the single-level optimization problem

$$\begin{aligned} \text{(VP)} \quad & \min \quad F(x, y) \\ & \text{s.t.} \quad f(x, y) - V(x) \leq 0, \quad x \in X, \quad y \in Y(x). \end{aligned}$$

This reformulation was first proposed by Outrata in [34]. Subsequently, Ye et al. [46, 47] and Dempe et al. [9, 11, 12] used the lower-level optimal value functions to study various optimality conditions and constraint qualifications for bilevel programs.

On the other hand, although the reformulation (VP) looks as a single-level optimization problem, because the lower-level optimal value function $V(x)$ does not have analytic expression and may not be differentiable in general, it is obviously not a normal constrained

optimization problem. Moreover, it is easy to see that the inequality constraint in (VP) is actually an equality constraint and hence the nonsmooth MFCQ is never satisfied at each feasible point [46, Proposition 3.2]. In consequence, we cannot apply the well-developed optimization algorithms to solve it directly. Recently, motivated by the study of semi-infinite programming, Xu et al. [28, 44] considered a simple bilevel program with compact lower-level feasible region and proposed some approximation methods by using an integral entropy function to approximate the lower-level optimal value function.

2.1.3 Reformulation based on lower-level optimality conditions

Because of the existence of the lower-level optimal solution or optimal value functions, both reformulations introduced above are hard to apply in solving bilevel programs. The most popular approach so far is to use the lower-level optimality conditions to transform the bilevel program (BP) into an MPEC. If the lower-level feasible region $Y(x)$ is convex for each $x \in X$, by means of the classical first-order necessary conditions for (P_x) , the bilevel program (BP) can be reformulated as a mathematical program with parametric variational inequality constraints, which is essentially a semi-infinite programming problem and mainly useful in theoretical study. To develop numerical algorithms, we may use the KKT conditions for (P_x) , which usually requires some constraint qualifications, to reformulate the bilevel program (BP) as the following mathematical program with complementarity constraints:

$$\begin{aligned}
 \text{(MPEC)} \quad & \min && F(x, y) \\
 & \text{s.t.} && x \in X, \quad h(x, y) = 0, \\
 & && 0 \leq u \perp g(x, y) \leq 0, \\
 & && \nabla_y f(x, y) + \nabla_y g(x, y)u + \nabla_y h(x, y)v = 0.
 \end{aligned}$$

On the one hand, although the globally optimal solutions of the original bilevel program (BP) and the corresponding MPEC coincide with each other, their local optimal solutions may not be corresponded in the case that multiple multipliers exist for the lower-level program [8]. In addition, Ye gave an example to reveal that the optimal solutions of the original bilevel program may not even be a stationary point of the corresponding MPEC [45].

On the other hand, MPEC is not an ordinary constrained optimization problem because its feasible region has a combinatorial structure and, besides, it fails to satisfy the MFCQ at any feasible point, which means that the well-developed optimization algorithms in nonlinear programming may not be stable in solving MPEC. During the past decades, there have been proposed a number of approximation algorithms for solving MPEC, which include the relaxation and smoothing algorithms, penalty function algorithms, interior point algorithms, implicit programming algorithms, active-set identification algorithms, constrained equation algorithms, nonsmooth algorithms, etc. We refer the reader to [10, 14–18, 21, 26, 27, 31, 35, 38, 39] and the references therein for more details about the recent developments on theory and numerical algorithms for MPEC.

2.1.4 Reformulation based on lower-level Lagrangian duality

Recently, Ouattara and Aswani [33] studied another approach to deal with bilevel programs by using some kind of Lagrangian dual functions of the lower-level program (P_x) . In fact, since the classical Lagrangian dual function of (P_x) may not be differentiable and even not real-valued, under the condition that the lower-level feasible region is uniformly contained within the interior of some compact and convex set \bar{Y} , the authors defined a so-called constrained Lagrangian dual function by

$$\mathcal{V}(x, u, v) := \min_y \{f(x, y) + u^T g(x, y) + v^T h(x, y) : y \in \bar{Y}\}$$

and showed that, if $f(x, y)$ is strictly convex in y for each $x \in X$, the constrained Lagrangian dual function is differentiable when $u \geq 0$. In this case, they presented the following duality-based reformulation:

$$\begin{aligned} \text{(LDP)} \quad & \min && F(x, y) \\ & \text{s.t.} && f(x, y) \leq \mathcal{V}(x, u, v), \\ & && u \geq 0, \quad g(x, y) \leq 0, \\ & && x \in X, \quad h(x, y) = 0. \end{aligned}$$

If $f(x, y)$ is only convex in y for each $x \in X$, they introduced a regularized constrained Lagrangian dual function

$$\mathcal{V}_\delta(x, u, v) := \min_y \{\delta \|y\|^2 + f(x, y) + u^T g(x, y) + v^T h(x, y) : y \in \bar{Y}\},$$

where $\delta > 0$, and gave the following approximated duality-based reformulation:

$$\begin{aligned} \min && F(x, y) \\ \text{s.t.} && f(x, y) \leq \mathcal{V}_\delta(x, u, v), \\ && u \geq 0, \quad g(x, y) \leq 0, \\ && x \in X, \quad h(x, y) = 0. \end{aligned} \tag{2.1}$$

Note that both \mathcal{V} and \mathcal{V}_δ do not have analytic expressions in general and, in addition, the approximation parameter δ in (2.1) is required to tend to 0 decreasingly in solving the original bilevel program.

2.2 Wolfe duality for constrained optimization problems

This subsection is devoted to gathering some useful results related to the Wolfe duality for constrained optimization problems, which play a crucial role in our proposed approach to bilevel programs. For simplicity, in this subsection, we omit the upper-level variable x and rewrite the lower-level program (P_x) as

$$\begin{aligned} \text{(P)} \quad & \min && f(y) \\ & \text{s.t.} && g(y) \leq 0, \quad h(y) = 0, \end{aligned}$$

where $f : \mathbb{R}^m \rightarrow \mathbb{R}$, $g : \mathbb{R}^m \rightarrow \mathbb{R}^p$, and $h : \mathbb{R}^m \rightarrow \mathbb{R}^q$. Denote by Y and S the feasible region and solution set of (P) respectively. The Wolfe dual problem given in [43] for (P) is

$$\begin{aligned} \text{(WD)} \quad & \max \quad L(z, u, v) \\ & \text{s.t.} \quad \nabla_z L(z, u, v) = 0, \quad u \geq 0, \end{aligned}$$

where $L : \mathbb{R}^{m+p+q} \rightarrow \mathbb{R}$ represents the Lagrangian function

$$L(z, u, v) := f(z) + u^T g(z) + v^T h(z). \quad (2.2)$$

We denote by W the feasible region of (WD) and, for any fixed (u, v) , we let $Z(u, v) := \{z \in \mathbb{R}^m : \nabla_z L(z, u, v) = 0\}$.

Note that, since the Lagrangian function L may not be convex over W due to the existence of the terms $u^T g$ and $v^T h$, the Wolfe dual problem (WD) is usually a nonconvex optimization problem, even though the primal problem (P) is convex.

Definition 2.1 We say f to be pseudoconvex on $\mathcal{Y} \subseteq \mathbb{R}^m$ if $(y_1 - y_2)^T \nabla f(y_2) \geq 0$ implies $f(y_1) \geq f(y_2)$ for any $y_1, y_2 \in \mathcal{Y}$.

We have the following duality theorems.

Theorem 2.1 (weak Wolfe duality) *Suppose that $L(\cdot, u, v)$ is pseudoconvex on $Y \cup Z(u, v)$ for any $u \in \mathbb{R}_+^p$ and $v \in \mathbb{R}^q$. Then, the optimal value of the primal problem (P) is no less than its Wolfe dual problem (WD), i.e.,*

$$\min_{y \in Y} f(y) \geq \max_{(z, u, v) \in W} L(z, u, v).$$

Proof. For any $y \in Y$ and $(z, u, v) \in W$, we have $\nabla_z L(z, u, v) = 0$ and hence $L(y, u, v) \geq L(z, u, v)$ by the pseudoconvexity of $L(\cdot, u, v)$. Since $u \in \mathbb{R}_+^p$ and $y \in Y$ satisfies $h(x, y) = 0$ and $g(x, y) \leq 0$, we have

$$f(y) \geq f(y) + u^T g(y) + v^T h(y) = L(y, u, v) \geq L(z, u, v).$$

By the arbitrariness of $y \in Y$ and $(z, u, v) \in W$, we got the conclusion. ■

Theorem 2.2 (strong Wolfe duality) *Suppose that $L(\cdot, u, v)$ is pseudoconvex on $Y \cup Z(u, v)$ for any $u \in \mathbb{R}_+^p$, $v \in \mathbb{R}^q$ and (P) satisfies the Guignard constraint qualification at some solution $\bar{y} \in S$. Then, there exists a solution $(\bar{z}, \bar{u}, \bar{v}) \in W$ to (WD) such that*

$$\min_{y \in Y} f(y) = f(\bar{y}) = L(\bar{z}, \bar{u}, \bar{v}) = \max_{(z, u, v) \in W} L(z, u, v).$$

Proof. Since (P) satisfies the Guignard constraint qualification at $\bar{y} \in S$, there exists (\bar{u}, \bar{v}) such that

$$\nabla_y L(\bar{y}, \bar{u}, \bar{v}) = 0, \quad \bar{u}^T g(\bar{y}) = 0, \quad \bar{u} \geq 0,$$

from which we have $(\bar{y}, \bar{u}, \bar{v}) \in W$. We can show that $(\bar{y}, \bar{u}, \bar{v})$ is an optimal solution to (WD). In fact, otherwise, there must exist some $(y, u, v) \in W$ such that

$$L(y, u, v) > L(\bar{y}, \bar{u}, \bar{v}) = f(\bar{y}),$$

which contradicts the weak Wolfe duality shown in Theorem 2.1. Therefore, $(\bar{y}, \bar{u}, \bar{v})$ is an optimal solution to (WD). As a result, we have

$$\min_{y \in Y} f(y) = f(\bar{y}) = L(\bar{y}, \bar{u}, \bar{v}) = \max_{(z, u, v) \in W} L(z, u, v).$$

This completes the proof. ■

Remark 2.1 Note that the pseudoconvexity assumption for the Lagrangian function in Theorems 2.1 and 2.2 can be replaced by some other weaker conditions such as the \mathcal{F} -pseudoconvexity or invexity assumptions. Since these conditions are not easy to understand by definitions, we use the well-known pseudoconvexity to describe our results. Actually, most results can be weakened by replacing the pseudoconvexity assumption by these weaker conditions. We refer the reader to [20, 32, 42, 43] for details about these conditions.

Remark 2.2 The above-mentioned \mathcal{F} -pseudoconvexity assumption, weaker than the pseudoconvexity assumption, was presented in [20]. Actually, in [20], the \mathcal{F} -pseudoconvexity was assumed over the feasible region Y only, rather than the set $Y \cup Z(u, v)$ assumed in Theorems 2.1 and 2.2. However, it can be seen from their proofs that the \mathcal{F} -pseudoconvexity over the feasible region only is not sufficient to derive their conclusions. This is why we retain the proofs for these theorems.

Remark 2.3 It should be noted that, in Theorem 2.2, the dual optimal solution \bar{z} may not solve the primal problem (P), unless there are some restrictions on the functions involved. See, e.g., the following converse duality theorem given in [4, 23].

Theorem 2.3 *Suppose that (P) is a convex program, \bar{y} solves (P), and $(\bar{z}, \bar{u}, \bar{v})$ solves (WD). Then, we have $\bar{z} = \bar{y}$ and $f(\bar{y}) = L(\bar{z}, \bar{u}, \bar{v})$ if one of the following conditions holds:*

- (i) *The Hessian matrix $\nabla_{zz}^2 L(\bar{z}, \bar{u}, \bar{v})$ is nonsingular.*
- (ii) *The Guignard constraint qualification holds at \bar{y} and the function f is strictly convex near \bar{z} .*

3 New reformulation for bilevel programs

In this section, by applying the Wolfe duality to the lower-level program, we present a novel single-level reformulation for the bilevel program (BP). Different from the existing reformulations mentioned in Subsection 2.1, the new reformulation does not involve any function without analytic expression and, in addition, an example is given in the next section to show that the new reformulation may satisfy the MFCQ at its feasible points, which means that it is no longer an MPEC.

As introduced in Subsection 2.2, for any given $x \in X$, the Wolfe dual problem for the lower-level program (P_x) is

$$\begin{aligned} (\text{WD}_x) \quad & \max_{z,u,v} && L(x, z, u, v) \\ & \text{s.t.} && \nabla_z L(x, z, u, v) = 0, \quad u \geq 0, \end{aligned}$$

where $L(x, z, u, v) := f(x, z) + u^T g(x, z) + v^T h(x, z)$. Similarly, we denote by $W(x)$ the feasible region of (WD_x) and, for any fixed (x, u, v) , we let $Z(x, u, v) := \{z \in \mathbb{R}^m : \nabla_z L(x, z, u, v) = 0\}$.

Theorem 3.1 *Suppose that (P_x) and (WD_x) satisfy the strong Wolfe duality for any $x \in X$. Then, in the globally optimal sense, solving the bilevel program (BP) is equivalent to solving the single-level optimization problem*

$$\begin{aligned} (\text{WDP}) \quad & \min && F(x, y) \\ & \text{s.t.} && x \in X, \quad y \in Y(x), \\ & && f(x, y) - L(x, z, u, v) \leq 0, \\ & && \nabla_z L(x, z, u, v) = 0, \quad u \geq 0, \end{aligned}$$

or equivalently,

$$\begin{aligned} \min & && F(x, y) \\ \text{s.t.} & && x \in X, \quad g(x, y) \leq 0, \quad h(x, y) = 0, \\ & && f(x, y) - f(x, z) - u^T g(x, z) - v^T h(x, z) \leq 0, \\ & && \nabla_z f(x, z) + \nabla_z g(x, z)u + \nabla_z h(x, z)v = 0, \quad u \geq 0. \end{aligned}$$

Proof. Since the strong Wolfe duality always holds for (P_x) and (WD_x) , the value function reformulation (VP) can be rewritten as

$$\begin{aligned} \min & && F(x, y) \\ \text{s.t.} & && x \in X, \quad y \in Y(x), \\ & && f(x, y) \leq \max_{(z,u,v) \in W(x)} L(x, z, u, v). \end{aligned} \tag{3.1}$$

Let $(\bar{z}_x, \bar{u}_x, \bar{v}_x) \in W(x)$ represent the optimal solution to (WD_x) mentioned as in Theorem 2.2, that is, $L(x, \bar{z}_x, \bar{u}_x, \bar{v}_x) = \max_{(z,u,v) \in W(x)} L(x, z, u, v)$.

If (x, y) is feasible to (3.1), we have

$$f(x, y) \leq \max_{(z,u,v) \in W(x)} L(x, z, u, v) = L(x, \bar{z}_x, \bar{u}_x, \bar{v}_x).$$

It is easy to see that $(x, y, \bar{z}_x, \bar{u}_x, \bar{v}_x)$ is also feasible to (WDP). Conversely, if (x', y', z', u', v') is feasible to (WDP), since $(z', u', v') \in W(x')$, we have

$$f(x', y') \leq L(x', z', u', v') \leq \max_{(z,u,v) \in W(x')} L(x', z, u, v).$$

This means that (x', y') is also feasible to (3.1).

The above analysis shows that (3.1) and (WDP) have the same feasible region with respect to the variables x and y . Since their objective functions are both $F(x, y)$, we can claim that solving (3.1), which is equivalent to the original bilevel program (BP), is equivalent to solving (WDP) in the globally optimal sense. This completes the proof. \blacksquare

Theorem 3.2 *Let $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ be feasible to (WDP) and $L(\bar{x}, \cdot, \bar{u}, \bar{v})$ be pseudoconvex on $Y(\bar{x}) \cup Z(\bar{x}, \bar{u}, \bar{v})$. Then, \bar{y} is globally optimal to $(P_{\bar{x}})$ and $(\bar{z}, \bar{u}, \bar{v})$ is globally optimal to $(WD_{\bar{x}})$.*

Proof. By the feasibility of $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ to (WDP), we have

$$\min_{y \in Y(\bar{x})} f(\bar{x}, y) \leq f(\bar{x}, \bar{y}) \leq L(\bar{x}, \bar{z}, \bar{u}, \bar{v}) \leq \max_{(z,u,v) \in W(\bar{x})} L(\bar{x}, z, u, v). \quad (3.2)$$

By Theorem 2.1, the weak Wolfe duality holds for $(P_{\bar{x}})$ and $(WD_{\bar{x}})$, that is, $\min_{y \in Y(\bar{x})} f(\bar{x}, y) \geq \max_{(z,u,v) \in W(\bar{x})} L(\bar{x}, z, u, v)$. Therefore, all inequalities in (3.2) should be equal, which means that \bar{y} is a globally optimal solution to $(P_{\bar{x}})$ and $(\bar{z}, \bar{u}, \bar{v})$ is a globally optimal solution to $(WD_{\bar{x}})$. This completes the proof. \blacksquare

Theorem 3.1 shows the equivalence between (BP) and (WDP) in the globally optimal sense. Since both of them are nonconvex optimization problems, it is necessary to investigate their equivalence in the locally optimal sense. Denote by $U(w)$ some neighborhood of a point w .

Theorem 3.3 (i) *Suppose that (\bar{x}, \bar{y}) is locally optimal to (BP) restricting on $U(\bar{x}) \times U(\bar{y})$ and $(P_{\bar{x}})$ satisfies the Guignard constraint qualification at some optimal solution \bar{z} with $\{\bar{u}, \bar{v}\}$ being the corresponding Lagrange multipliers. If $L(x, \cdot, u, v)$ is pseudoconvex on $Y(x) \cup Z(x, u, v)$ for any $(x, u, v) \in U(\bar{x}) \times U(\bar{u}) \times U(\bar{v})$, then $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ is locally optimal to (WDP).*

(ii) *Conversely, let $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ be locally optimal to (WDP) restricting on $U(\bar{x}) \times U(\bar{y}) \times U(\bar{z}) \times \mathbb{R}^p \times \mathbb{R}^q$ with $U(\bar{y}) \subseteq U(\bar{z})$. If (P_x) satisfies the Guignard constraint qualification at all points in $S(x)$ for any $x \in U(\bar{x})$, then (\bar{x}, \bar{y}) is locally optimal to (BP).*

Proof. (i) Since both \bar{y} and \bar{z} are optimal solutions to $(P_{\bar{x}})$ and $\{\bar{u}, \bar{v}\}$ are the Lagrange multipliers corresponding to \bar{z} , we have

$$f(\bar{x}, \bar{y}) = f(\bar{x}, \bar{z}) = L(\bar{x}, \bar{z}, \bar{u}, \bar{v})$$

and hence $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ is feasible to (WDP).

On the other hand, for any feasible point (x, y, z, u, v) to (WDP) in the neighborhood $U(\bar{x}) \times U(\bar{y}) \times U(\bar{z}) \times U(\bar{u}) \times U(\bar{v})$, by the assumptions, the Lagrangian function $L(x, \cdot, u, v)$ is pseudoconvex on $Y(x) \cup Z(x, u, v)$. By Theorem 3.2, there must hold $y \in S(x)$, which means that (x, y) is feasible to (BP) and belongs to $U(\bar{x}) \times U(\bar{y})$. Noting that (\bar{x}, \bar{y}) is locally optimal to (BP) restricting on $U(\bar{x}) \times U(\bar{y})$, we have

$$F(\bar{x}, \bar{y}) \leq F(x, y). \quad (3.3)$$

By the arbitrariness of (x, y, z, u, v) , we get that $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ is a locally optimal solution of (WDP).

(ii) Let $(x, y) \in U(\bar{x}) \times U(\bar{y})$ be feasible point to (BP). It follows that $y \in S(x)$. By the assumptions, (P_x) satisfies the Guignard constraint qualification at y and hence there exist multipliers $\{u, v\}$ satisfying

$$\nabla_y L(x, y, u, v) = 0, \quad u \geq 0, \quad u^T g(x, y) = 0.$$

It then follows from $h(x, y) = 0$ that $f(x, y) = L(x, y, u, v)$. In consequence, (x, y, y, u, v) is feasible to (WDP) and belongs to $U(\bar{x}) \times U(\bar{y}) \times U(\bar{y}) \times U(\bar{u}) \times U(\bar{v})$. Since $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ is locally optimal to (WDP) restricting on $U(\bar{x}) \times U(\bar{y}) \times U(\bar{z}) \times \mathbb{R}^p \times \mathbb{R}^q$ with $U(\bar{y}) \subseteq U(\bar{z})$, we have (3.3). By the arbitrariness of (x, y) , (\bar{x}, \bar{y}) is a locally optimal solution of (BP). This completes the proof. \blacksquare

4 Comparison between WDP and MPEC

In the previous section, it is shown that, under mild assumptions, the bilevel program (BP) is equivalent to the single-level optimization problem (WDP) in the globally or locally optimal sense. Note that, unlike the reformulations (SP), (VP), and (LDP) introduced in Section 2, (WDP) does not involve any function without analytic expression. In this section, our first purpose is to compare (WDP) with the most popular reformulation MPEC to reveal which is better and our second purpose is to investigate their relations.

4.1 WDP is no longer an MPEC

As stated before, in order to solve the bilevel program (BP) effectively, it generally requires to transform it into some single-level optimization problem, no matter whether it satisfies

those conditions to ensure their equivalence in theory. So, which is better between (WDP) and MPEC?

Recall that MPEC is not an ordinary constrained optimization problem. In fact, MPEC fails to satisfy the MFCQ at any feasible point. This means that the well-developed optimization algorithms in nonlinear programming may not be stable in solving MPEC. What about (WDP)?

Example 4.1 Consider the bilevel program

$$\begin{aligned} \min \quad & (x - y - 8)^2 \\ \text{s.t.} \quad & x \geq 1, y \in S(x) = \arg \min_y \{y : y^3 \leq x, y \geq 0\}. \end{aligned} \quad (4.1)$$

Obviously, $S(x) = \{0\}$ for any $x \geq 1$ and so (4.1) has a unique minimizer $(8, 0)$. The (WDP) reformulation for (4.1) is

$$\begin{aligned} \min \quad & (x - y - 8)^2 \\ \text{s.t.} \quad & x \geq 1, y^3 - x \leq 0, y \geq 0, \\ & y - z - u_1(z^3 - x) + u_2z \leq 0, \\ & 1 + 3u_1z^2 - u_2 = 0, u_1 \geq 0, u_2 \geq 0. \end{aligned} \quad (4.2)$$

Consider the feasible point $(x^*, y^*, z^*, u_1^*, u_2^*) := (8, 0, -3, 0, 1)$, which is actually a globally optimal solution of (4.2). In what follows, we show that the MFCQ holds at this point.

- The gradient corresponding to the unique equality constraint at the point is $\mathbf{g}_0 := (0, 0, 0, 27, -1)^T$, which is a nonzero vector, and hence it is linearly independent.
- The active inequality constraints at the point include

$$y \geq 0, \quad y - z - u_1(z^3 - x) + u_2z \leq 0, \quad u_1 \geq 0.$$

The corresponding gradients at the point are respectively given by

$$\begin{aligned} \mathbf{g}_1 &:= (0, -1, 0, 0, 0)^T, \\ \mathbf{g}_2 &:= (0, 1, 0, 35, -3)^T, \\ \mathbf{g}_3 &:= (0, 0, 0, -1, 0)^T. \end{aligned}$$

Take a vector $d := (0, 1, 0, 1, 27)^T$. It is trivial to verify that $d^T \mathbf{g}_0 = 0$, $d^T \mathbf{g}_1 < 0$, $d^T \mathbf{g}_2 < 0$, and $d^T \mathbf{g}_3 < 0$.

The above analysis indicates that (4.2) satisfies the MFCQ at its feasible point $(8, 0, -3, 0, 1)$.

Example 4.1 shows that (WDP) may satisfy the MFCQ at its feasible points and hence it is no longer an MPEC. In consequence, compared with MPEC, (WDP) is relatively easy to deal with.

4.2 Relations between WDP and MPEC

In this subsection, for simplicity, we take the upper constraint $x \in X$ and the lower equality constraint $h(x, y) = 0$ away from the original program (BP). In this case, the corresponding MPEC becomes

$$\begin{aligned} \min \quad & F(x, y) \\ \text{s.t.} \quad & \nabla_y L(x, y, u) = 0, \\ & 0 \leq u \perp g(x, y) \leq 0, \end{aligned} \tag{4.3}$$

where $L(x, y, u) := f(x, y) + u^T g(x, y)$, while (WDP) becomes

$$\begin{aligned} \min \quad & F(x, y) \\ \text{s.t.} \quad & f(x, y) - L(x, z, u) \leq 0, \\ & \nabla_z L(x, z, u) = 0, \quad u \geq 0, \quad g(x, y) \leq 0. \end{aligned} \tag{4.4}$$

In what follows, we investigate the relations between the stationarity conditions of the above problems. There is no any difficulty to extend the subsequent analysis to the general MPEC and the general (WDP).

By simple deduction, we have the following result immediately.

Theorem 4.1 *The point (x, y, u) is feasible to (4.3) if and only if (x, y, y, u) is feasible to (4.4).*

Let $(\bar{x}, \bar{y}, \bar{u})$ be a feasible point of (4.3). We define the following index sets:

$$\begin{aligned} I_{0+} &:= \{i : g_i(\bar{x}, \bar{y}) = 0, \bar{u}_i > 0\}, \\ I_{-0} &:= \{i : g_i(\bar{x}, \bar{y}) < 0, \bar{u}_i = 0\}, \\ I_{00} &:= \{i : g_i(\bar{x}, \bar{y}) = 0, \bar{u}_i = 0\}. \end{aligned}$$

Definition 4.1 *We call $(\bar{x}, \bar{y}, \bar{u})$ a strongly stationary point (or S -stationary point) of (4.3) if there exists $(\lambda^g, \lambda^u, \gamma) \in \mathbb{R}^{p+p+m}$ such that*

$$\nabla_x F(\bar{x}, \bar{y}) + \nabla_{yx}^2 L(\bar{x}, \bar{y}, \bar{u})\gamma + \nabla_x g(\bar{x}, \bar{y})\lambda^g = 0, \tag{4.5}$$

$$\nabla_y F(\bar{x}, \bar{y}) + \nabla_{yy}^2 L(\bar{x}, \bar{y}, \bar{u})\gamma + \nabla_y g(\bar{x}, \bar{y})\lambda^g = 0, \tag{4.6}$$

$$\nabla_y g(\bar{x}, \bar{y})^T \gamma - \lambda^u = 0, \tag{4.7}$$

$$\nabla_y L(\bar{x}, \bar{y}, \bar{u}) = 0, \quad g(\bar{x}, \bar{y}) \leq 0, \quad \bar{u} \geq 0, \tag{4.8}$$

$$\lambda_i^g = 0, \quad i \in I_{-0}, \tag{4.9}$$

$$\lambda_i^u = 0, \quad i \in I_{0+}, \tag{4.10}$$

$$\lambda_i^g \geq 0, \quad \lambda_i^u \geq 0, \quad i \in I_{00}. \tag{4.11}$$

Remark 4.1 Other popular stationarity conditions for MPEC include the Mordukhovich stationarity (or M-stationarity) and the Clarke stationarity (or C-stationarity). Among these stationarities, the S-stationarity is the strongest and most favorable one in the study of MPEC.

Theorem 4.2 *Let $(\bar{x}, \bar{y}, \bar{y}, \bar{u})$ be a KKT point of (4.4). Then, $(\bar{x}, \bar{y}, \bar{u})$ is an S-stationary point of (4.3).*

Proof. If $(\bar{x}, \bar{y}, \bar{y}, \bar{u})$ is a KKT point of (4.4), there exists $(\eta^g, \eta^u, \alpha, \beta) \in \mathbb{R}^{p+p+1+m}$ such that

$$\nabla_x F(\bar{x}, \bar{y}) + \nabla_{yx}^2 L(\bar{x}, \bar{y}, \bar{u})\beta + \nabla_x g(\bar{x}, \bar{y})(\eta^g - \alpha\bar{u}) = 0, \quad (4.12)$$

$$\nabla_y F(\bar{x}, \bar{y}) + \alpha \nabla_y f(\bar{x}, \bar{y}) + \nabla_y g(\bar{x}, \bar{y})\eta^g = 0, \quad (4.13)$$

$$-\alpha \nabla_y L(\bar{x}, \bar{y}, \bar{u}) + \nabla_{yy}^2 L(\bar{x}, \bar{y}, \bar{u})\beta = 0, \quad (4.14)$$

$$-\alpha g(\bar{x}, \bar{y}) + \nabla_y g(\bar{x}, \bar{y})^T \beta - \eta^u = 0, \quad (4.15)$$

$$\nabla_y L(\bar{x}, \bar{y}, \bar{u}) = 0, \quad (4.16)$$

$$0 \leq \alpha \perp \bar{u}^T g(\bar{x}, \bar{y}) \geq 0, \quad (4.17)$$

$$0 \leq \eta^g \perp g(\bar{x}, \bar{y}) \leq 0, \quad (4.18)$$

$$0 \leq \eta^u \perp \bar{u} \geq 0. \quad (4.19)$$

Adding (4.13) and (4.14) together yields

$$\nabla_y F(\bar{x}, \bar{y}) + \nabla_{yy}^2 L(\bar{x}, \bar{y}, \bar{u})\beta + \nabla_y g(\bar{x}, \bar{y})(\eta^g - \alpha\bar{u}) = 0. \quad (4.20)$$

Set $\lambda^g := \eta^g - \alpha\bar{u}$, $\lambda^u := \eta^u + \alpha g(\bar{x}, \bar{y})$ and $\gamma := \beta$. We obtain (4.5)-(4.8) from (4.12) and (4.15)-(4.20) immediately. Next, we show (4.9)-(4.11).

- If $i \in I_{-0}$, which means $g_i(\bar{x}, \bar{y}) < 0$ and $\bar{u}_i = 0$, we have $\eta_i^g = 0$ by (4.18) and hence $\lambda_i^g = \eta_i^g - \alpha\bar{u}_i = 0$. Namely, (4.9) holds.
- If $i \in I_{0+}$, which means $g_i(\bar{x}, \bar{y}) = 0$ and $\bar{u}_i > 0$, we have $\eta_i^u = 0$ by (4.19) and hence $\lambda_i^u = \eta_i^u + \alpha g_i(\bar{x}, \bar{y}) = 0$. Namely, (4.10) holds.
- If $i \in I_{00}$, which means $g_i(\bar{x}, \bar{y}) = \bar{u}_i = 0$, we have $\lambda_i^g = \eta_i^g \geq 0$ and $\lambda_i^u = \eta_i^u \geq 0$ by (4.18) and (4.19) respectively. Namely, (4.11) holds.

This completes the proof. ■

The following example shows that the converse of Theorem 4.2 is not true.

Example 4.2 Consider the bilevel program

$$\begin{aligned} \min \quad & x^2 - (2y + 1)^2 \\ \text{s.t.} \quad & x \leq 0, \\ & y \in \arg \min_y \{(y - 1)^2 : 3x - y - 3 \leq 0, x + y - 1 \leq 0\}. \end{aligned} \quad (4.21)$$

It is easy to see that $(0, 1)$ is the unique globally optimal solution of (4.21). Since the lower-level program is convex and satisfies the linear constraint qualification everywhere, we can equivalently transform (4.21) into the MPEC

$$\begin{aligned}
\min \quad & x^2 - (2y + 1)^2 \\
\text{s.t.} \quad & x \leq 0, \quad 2(y - 1) - u_1 + u_2 = 0, \\
& 0 \leq u_1 \perp 3x - y - 3 \leq 0, \quad 0 \leq u_2 \perp x + y - 1 \leq 0,
\end{aligned} \tag{4.22}$$

or (WDP)

$$\begin{aligned}
\min \quad & x^2 - (2y + 1)^2 \\
\text{s.t.} \quad & x \leq 0, \quad 2(z - 1) - u_1 + u_2 = 0, \\
& (y - 1)^2 - (z - 1)^2 - u_1(3x - z - 3) - u_2(x + z - 1) \leq 0, \\
& u_1 \geq 0, \quad u_2 \geq 0, \quad 3x - y - 3 \leq 0, \quad x + y - 1 \leq 0.
\end{aligned} \tag{4.23}$$

Consider the feasible point $(0, 1, 0, 0)$ of the MPEC (4.22). The S-stationarity conditions at this point are

$$\begin{aligned}
\alpha + 3\lambda_1^g + \lambda_2^g &= 0, \quad -12 + 2\beta - \lambda_1^g + \lambda_2^g = 0, \\
-\beta - \lambda_1^u &= 0, \quad \beta - \lambda_2^u = 0, \quad \alpha \geq 0, \\
\lambda_1^g &= 0, \quad \lambda_2^g \geq 0, \quad \lambda_2^u \geq 0,
\end{aligned}$$

where $\{\alpha, \beta\}$ are multipliers corresponding to the first two constraints in (4.22) and $\{\lambda^u, \lambda^g\}$ are multipliers corresponding to the complementarity constraints. It is obvious that $\alpha = 0, \beta = 6, \lambda^u = (-6, 6)^T, \lambda^g = (0, 0)^T$ satisfy the above conditions and hence $(0, 1, 0, 0)$ is an S-stationary point of the MPEC (4.22).

Next, we consider the feasible point $(0, 1, 1, 0, 0)$ of problem (4.23). The KKT conditions at this point are

$$\begin{aligned}
\xi + 3\mu_1^g + \mu_2^g &= 0, \quad -12 - \mu_1^g + \mu_2^g = 0, \quad 2\eta = 0, \\
-\eta + 4\zeta - \mu_1^u &= 0, \quad \eta - \mu_2^u = 0, \\
\xi \geq 0, \quad \zeta \geq 0, \quad \mu_1^g &= 0, \quad \mu_2^g \geq 0, \quad \mu^u \geq 0,
\end{aligned}$$

where $\{\xi, \eta, \zeta\}$ are multipliers corresponding to the first three constraints in (4.23) and $\{\mu^u, \mu^g\}$ are multipliers corresponding to the others. Note that

$$\mu_1^g = 0 \quad \Rightarrow \quad \mu_2^g = 12 \quad \Rightarrow \quad \xi = -12,$$

which contradicts $\xi \geq 0$. Therefore, $(0, 1, 1, 0, 0)$ is not a KKT point of (4.23).

We further have the following result.

Theorem 4.3 *Let (x, y, z, u) be a feasible point of (4.4). Then, the MFCQ fails to hold at this point if $z = y$.*

Proof. The set of abnormal multipliers for (4.4) at (x, y, z, u) is given by

$$\mathcal{M} := \left\{ (\alpha, \beta, \eta^g, \eta^u) : \begin{array}{l} 0 = \alpha(\nabla_x f(x, y) - \nabla_x L(x, z, u)) \\ \quad + \nabla_{zx}^2 L(x, z, u)\beta + \nabla_x g(x, y)\eta^g \\ 0 = \alpha \nabla_y f(x, y) + \nabla_y g(x, y)\eta^g \\ 0 = -\alpha \nabla_z L(x, z, u) + \nabla_{zz}^2 L(x, z, u)\beta \\ 0 = -\alpha g(x, z) + \nabla_z g(x, z)^T \beta - \eta^u \\ \alpha \geq 0, \quad \alpha(f(x, y) - L(x, z, u)) = 0 \\ \eta^u \geq 0, \quad u^T \eta^u = 0, \quad \eta^g \geq 0, \quad g(x, y)^T \eta^g = 0 \end{array} \right\}.$$

If $z = y$, it is not difficult to verify by the feasibility of (x, y, y, u) to (4.4) that $(1, 0, u, -g(x, y)) \in \mathcal{M}$. Recall that, by [2, Chapter 6.3], a nonzero abnormal multiplier exists if and only if the MFCQ does not hold. Therefore, the MFCQ does not hold at (x, y, y, u) . This completes the proof. \blacksquare

5 Relaxation scheme for WDP

Theorem 4.3 indicates that, although (WDP) is no longer an MPEC, it is still not an ordinary constrained optimization problem in some cases. Moreover, our numerical experiments given in Subsection 6.2 show that the efficiency was not very high if we solved the bilevel program (BP) by solving (WDP) directly. Based on these reasons, we present a relaxation scheme to deal with (WDP) in this section.

Our strategy is to use the following problem to approximate (WDP):

$$\begin{array}{ll} \text{WDP}(t) & \min \quad F(x, y) \\ & \text{s.t.} \quad x \in X, \quad g(x, y) \leq 0, \quad h(x, y) = 0, \\ & \quad \quad \quad f(x, y) - L(x, z, u, v) \leq t, \\ & \quad \quad \quad \nabla_z L(x, z, u, v) = 0, \quad u \geq 0, \end{array}$$

where t is a positive parameter. Obviously, all feasible points of (WDP) are also feasible to the above problem and, in addition, the second inequality constraint is not active any more at these points so that the popular constraint qualifications may be easily satisfied. Therefore, we may treat $\text{WDP}(t)$ as an ordinary constrained optimization problem.

We first give some results related to constraint qualifications. To this end, we introduce the following definition.

Definition 5.1 [5] The vectors $\{a^1, \dots, a^l\} \cup \{b^1, \dots, b^r\}$ is called to be positive-linearly dependent if there exist $\alpha \in \mathbb{R}_+^l$ and $\beta \in \mathbb{R}^r$ with $(\alpha, \beta) \neq 0$ such that

$$\sum_{i=1}^l \alpha_i a^i + \sum_{i=1}^r \beta_i b^i = 0.$$

Otherwise, the set is called to be positive-linearly independent.

By [36, Proposition 1.8.2], the MFCQ is satisfied at a feasible point if and only if the gradients of all active constraints at the point are positive-linearly independent. By [37, Proposition 2.2], for given continuous functions $\{G_1(w), \dots, G_l(w)\} \cup \{H_1(w), \dots, H_r(w)\}$, if they are positive-linearly independent at some point \bar{w} , there exists a neighborhood $U(\bar{w})$ such that they are positive-linearly independent at any point in $U(\bar{w})$.

From the above introduction, we have the following result immediately.

Theorem 5.1 *If WDP(t) satisfies the MFCQ at a feasible point $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$, there exists a neighborhood \bar{U} of $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ such that WDP(t) satisfies the MFCQ at any feasible point in \bar{U} .*

The above theorem can be extended to the following constraint qualification, which is weaker than the MFCQ and attracts lots of attention recently.

Definition 5.2 [37] Let $\bar{y} \in Y$ be a feasible point of the constrained optimization problem (P) in Subsection 2.2. We say that (P) satisfies the constant positive linear dependence (CPLD) at \bar{y} if, for any $I_0 \subseteq \{i : g_i(\bar{y}) = 0\}$ and $J_0 \subseteq \{1, \dots, q\}$ such that $\{\nabla g_i(\bar{y}), i \in I_0\} \cup \{\nabla h_j(\bar{y}), j \in J_0\}$ is positive-linearly dependent, there exists a neighborhood $U(\bar{y})$ such that $\{\nabla g_i(y), i \in I_0; \nabla h_j(y), j \in J_0\}$ is linearly dependent for any $y \in U(\bar{y})$.

Theorem 5.2 *If WDP(t) satisfies the CPLD at a feasible point $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$, there exists a neighborhood \bar{U} of $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ such that WDP(t) satisfies the CPLD at any feasible point in \bar{U} .*

Proof. For simplicity, we take the upper constraint $x \in X$ away from WDP(t) and denote by $\bar{w} := (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$, $w := (x, y, z, u, v)$. We further denote by $G_i(w)$ ($i = 0, 1, \dots, 2p$) all inequality constraint functions and by $H_i(w)$ ($i = 1, \dots, m + q$) all equality constraint functions.

Suppose by contradiction that the conclusion is wrong. That is, there is a sequence $\{w^k\}$ of feasible points tending to \bar{w} such that WDP(t) does not satisfy the CPLD at each w^k . Violation of CPLD means that there exist $I_0^k \subseteq \{i : G_i(w^k) = 0\}$ and $J_0^k \subseteq \{1, \dots, m + q\}$ such that $\{\nabla G_i(w^k), i \in I_0^k\} \cup \{\nabla H_j(w^k), j \in J_0^k\}$ is positive-linearly dependent, but linearly independent in some points arbitrarily close to w^k . Note that the number of subsets of the

index sets $\{i : G_i(w^k) = 0\}$ and $\{1, \dots, m + q\}$ is finite. Taking a subsequence if necessary, we may assume $I_0^k \equiv I_0$ and $J_0^k \equiv J_0$. Therefore, there is another sequence $\{\tilde{w}^k\}$ tending to \bar{w} , together with $\{w^k\}$, satisfying

- (i) $\{\nabla G_i(w^k), i \in I_0\} \cup \{\nabla H_j(w^k), j \in J_0\}$ is positive-linearly dependent;
- (ii) $\{\nabla G_i(\tilde{w}^k), i \in I_0; \nabla H_j(\tilde{w}^k), j \in J_0\}$ is linearly independent.

If $\{\nabla G_i(\bar{w}), i \in I_0\} \cup \{\nabla H_j(\bar{w}), j \in J_0\}$ is positive-linearly dependent, by the CPLD assumption, there exists a neighbourhood which remains the linear dependence, which contradicts the existence of $\{\tilde{w}^k\}$ in (ii). If $\{\nabla G_i(\bar{w}), i \in I_0\} \cup \{\nabla H_j(\bar{w}), j \in J_0\}$ is positive-linearly independent, by [37, Proposition 2.2], this property should remain in a neighbourhood, which contradicts the existence of $\{w^k\}$ in (i). As a result, the conclusion is true. This completes the proof. \blacksquare

Remark 5.1 It is easy to see from the proofs that the above results related to constraint qualifications remain true if $\text{WDP}(t)$ is replaced by (WDP).

In what follows, we consider the limiting behavior of $\text{WDP}(t)$ as the approximation parameter t tends to 0. The following convergence related to globally optimal solutions is evident.

Theorem 5.3 *Let $\{t_k\} \downarrow 0$ and $(x^k, y^k, z^k, u^k, v^k)$ be a globally optimal solution of $\text{WDP}(t_k)$ for each k . Then, every accumulation point $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ of the sequence $\{(x^k, y^k, z^k, u^k, v^k)\}$ is globally optimal to (WDP).*

Since $\text{WDP}(t)$ is usually a nonconvex optimization problem, it is necessary to investigate the limiting properties of its stationary points.

Theorem 5.4 *Let $\{t_k\} \downarrow 0$ and $(x^k, y^k, z^k, u^k, v^k)$ be a KKT point of $\text{WDP}(t_k)$ for each k . Assume that $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ is an accumulation point of the sequence $\{(x^k, y^k, z^k, u^k, v^k)\}$. If (WDP) satisfies the CPLD at $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$, then $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v})$ is a KKT point of (WDP).*

Proof. For simplicity, we take the upper constraint $x \in X$ away from (WDP) and $\text{WDP}(t)$. In addition, we denote by

$$\begin{aligned} \bar{w} &:= (\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}), & w^k &:= (x^k, y^k, z^k, u^k, v^k), & w &:= (x, y, z, u, v), \\ F_0(w) &:= F(x, y), & G_0(w) &:= f(x, y) - L(x, z, u, v), \\ G_i(w) &:= g_i(x, y) \quad (1 \leq i \leq p), & G_i(w) &:= -u_{i-p} \quad (p+1 \leq i \leq 2p), \\ H_i(w) &:= h_i(x, y) \quad (1 \leq i \leq q), & H_i(w) &:= \nabla_{z_{i-q}} L(x, z, u, v) \quad (q+1 \leq i \leq q+m). \end{aligned}$$

Then, (WDP) and $\text{WDP}(t_k)$ become

$$\begin{aligned} \min \quad & F_0(w) \\ \text{s.t.} \quad & G_i(w) \leq 0, \quad i = 0, 1, \dots, 2p, \\ & H_i(w) = 0, \quad i = 1, \dots, q + m, \end{aligned} \quad (5.1)$$

and

$$\begin{aligned} \min \quad & F_0(w) \\ \text{s.t.} \quad & G_0(w) \leq t_k, \\ & G_i(w) \leq 0, \quad i = 1, \dots, 2p, \\ & H_i(w) = 0, \quad i = 1, \dots, q + m. \end{aligned} \quad (5.2)$$

Obviously, \bar{w} is a feasible point of (WDP) and, without loss of generality, we may assume that the whole sequence $\{w^k\}$ converges to \bar{w} .

Since w^k is a KKT point of (5.2), it follows from [41, Lemma A.1] that there exists $(\lambda_0^k, \lambda^k, \mu^k) \in \mathbb{R} \times \mathbb{R}^{2p} \times \mathbb{R}^{q+m}$ such that

$$\begin{aligned} \nabla F_0(w^k) + \lambda_0^k \nabla G_0(w^k) + \nabla G(w^k) \lambda^k + \nabla H(w^k) \mu^k &= 0, \\ \lambda_0^k (G_0(w^k) - t_k) &= 0, \quad \lambda_0^k \geq 0, \\ G(w^k)^T \lambda^k &= 0, \quad \lambda^k \geq 0, \end{aligned} \quad (5.3)$$

and the gradients

$$\{\nabla G_i(w^k) : \lambda_i^k > 0, 0 \leq i \leq 2p\} \cup \{\nabla H_i(w^k) : \mu_i^k \neq 0\} \quad (5.4)$$

are linearly independent. To show that \bar{w} is a KKT point of (5.1), it is sufficient to show the boundedness of the sequence of multipliers $\{\lambda_0^k, \lambda^k, \mu^k\}$.

Assume by contradiction that $\{\lambda_0^k, \lambda^k, \mu^k\}$ is unbounded. Taking a subsequence if necessary, we may assume

$$\lim_{k \rightarrow \infty} \frac{(\lambda_0^k, \lambda^k, \mu^k)}{\|(\lambda_0^k, \lambda^k, \mu^k)\|} = (\lambda_0^*, \lambda^*, \mu^*). \quad (5.5)$$

In particular, for every k sufficiently large, we have

$$\lambda_i^* > 0 \Rightarrow \lambda_i^k > 0, \quad \mu_i^* \neq 0 \Rightarrow \mu_i^k \neq 0. \quad (5.6)$$

Dividing by $\|(\lambda_0^k, \lambda^k, \mu^k)\|$ in (5.3) and taking a limit, we have

$$\lambda_0^* \nabla G_0(\bar{w}) + \sum_{\lambda_i^* > 0} \lambda_i^* \nabla G_i(\bar{w}) + \sum_{\mu_i^* \neq 0} \mu_i^* \nabla H_i(\bar{w}) = 0, \quad \lambda_0^* \geq 0. \quad (5.7)$$

Since $\|(\lambda_0^*, \lambda^*, \mu^*)\| = 1$ by (5.5), (5.7) implies that the gradients

$$\{\nabla G_i(\bar{w}) : \lambda_i^* > 0, 0 \leq i \leq 2p\} \cup \{\nabla H_i(\bar{w}) : \mu_i^* \neq 0\}$$

are positive-linearly dependent. Since problem (5.1) satisfies the CPLD at \bar{w} , the gradients

$$\{\nabla G_i(w) : \lambda_i^* > 0, 0 \leq i \leq 2p\} \cup \{\nabla H_i(w) : \mu_i^* \neq 0\}$$

are linearly dependent in a neighborhood. This together with (5.6) contradicts the linear independence of the vectors in (5.4). Consequently, $\{\lambda_0^k, \lambda^k, \mu^k\}$ is bounded. This completes the proof. \blacksquare

6 Numerical experiments

In this section, we report our numerical experience in applying the proposed approach to solving linear bilevel programs. In our experiments, we generated the tested problems randomly and we made numerical comparison of the proposed (WDP) approach with the popular MPEC approach by solving them as nonlinear optimization problems and solving them through relaxation methods. In particular, all experiments were implemented in MATLAB 9.8.0 and run on the same computer with Windows Server 2019 Standard, Intel(R) Xeon(R) Platinum 8268, CPU @ 2.90 GHz 2.89 GHz (4 processors) and 2.00 TB memory.

6.1 Test problems

We designed a procedure to generate the following linear bilevel program:

$$\begin{aligned} \min \quad & c_1^T x + c_2^T y \\ \text{s.t.} \quad & A_1 x \leq b_1, \\ & y \in \arg \min_y \{d_2^T y : A_2 x + B_2 y \leq b_2, l_b \leq y \leq u_b\}, \end{aligned} \tag{6.1}$$

where all coefficients $A_1 \in \mathbb{R}^{p \times n}$, $A_2 \in \mathbb{R}^{q \times n}$, $B_2 \in \mathbb{R}^{q \times m}$, $b_1 \in \mathbb{R}^p$, $b_2 \in \mathbb{R}^q$, $c_1 \in \mathbb{R}^n$, and $\{c_2, d_2\} \subset \mathbb{R}^m$ were generated sparsely by *sprand* in $[0, 1]$ with density to be 0.1, the lower and upper bounds were given by $l_b = -10 * \text{ones}(m, 1)$ and $u_b = 10 * \text{ones}(m, 1)$ respectively. Since the lower-level program is linear, (6.1) can be equivalently transformed as the MPEC

$$\begin{aligned} \min \quad & c_1^T x + c_2^T y \\ \text{s.t.} \quad & A_1 x \leq b_1, \quad d_2 + B_2^T u_1 + u_2 - u_3 = 0, \\ & 0 \leq u_1 \perp A_2 x + B_2 y - b_2 \leq 0, \\ & 0 \leq u_2 \perp y - u_b \leq 0, \quad 0 \leq u_3 \perp y - l_b \geq 0, \end{aligned} \tag{6.2}$$

or (WDP)

$$\begin{aligned} \min \quad & c_1^T x + c_2^T y \\ \text{s.t.} \quad & A_1 x \leq b_1, \quad d_2 + B_2^T u_1 + u_2 - u_3 = 0, \\ & d_2^T y \leq d_2^T z + u_1^T (A_2 x + B_2 z - b_2) + u_2^T (z - u_b) + u_3^T (l_b - z), \\ & u_1 \geq 0, \quad u_2 \geq 0, \quad u_3 \geq 0, \quad A_2 x + B_2 y \leq b_2, \quad l_b \leq y \leq u_b. \end{aligned}$$

In our experiments, we generated 20 problems, which contain two groups: Problems 1-10 were generated with $\{n = 50, p = 40, m = 60, q = 50\}$ and problems 11-20 were generated with $\{n = 50, p = 40, m = 100, q = 80\}$. We adjusted the numbers m and q only because the lower-level variables and constraints are the essential influence factors in solving bilevel programs.

Moreover, in evaluating the numerical results, we mainly focused on three aspects: The first was the feasibility of the approximation solutions, the second was the objective values, and the last was the CPU times. In particular, we adopted the following unified criterion to measure the infeasibility for both MPEC and (WDP) in our tests:

$$\begin{aligned} \text{Infeasibility} \quad := \quad & \| \max(0, A_1 x^k - b_1) \| + \| \max(0, A_2 x^k + B_2 y^k - b_2) \| \\ & + \| \max(0, y^k - u_b) \| + \| \max(0, l_b - y^k) \| + |d_2^T y^k - h^*(x^k)|, \end{aligned}$$

where (x^k, y^k, u^k) or (x^k, y^k, z^k, u^k) is the current approximation solution to the MPEC or (WDP), $h^*(x^k) := \min\{d_2^T y : A_2 x^k + B_2 y \leq b_2, l_b \leq y \leq u_b\}$. The last term is used to replace the distance from y^k to the solution set of the lower-level program, which is actually difficult to calculate in practice.

6.2 Solving WDP and MPEC as nonlinear programs

In the first part of our experiments, we treated both MPEC and (WDP) as ordinary constrained optimization problems and solved them directly by invoking the SQP algorithm of the solver *fmincon*. We chose all starting points to be the origins and selected the required parameters, together with termination criterion, to default settings as set in the package. See Table 1 for the numerical results on the tested problems.

There are two aspects that are different from what we predicted before the experiments. The first is related to the feasibility of the approximation solutions. Since all lower-level programs are equivalent to their KKT conditions for the tested problems, the MPEC approach should have greater advantages in feasibility than (WDP), but the numerical results indicate that this is not always the case, at least in our tests. In fact, from Table 1, we can see that, if we set the tolerance for infeasibility to be 10^{-4} , there were 10 and 8 satisfactory cases among 20 problems for MPEC and (WDP) respectively.

The second aspect is that the efficiency of solving linear (BP) through the (WDP) approach directly was lower than we expected before the experiments. From Table 1, among the five cases where the infeasibility of both is lower than 10^{-4} (that is, problems 1, 7, 11, 18, 20), there were only two cases where the (WDP) approach had advantages over the MPEC approach in the objective values. This may be caused by the increase of dimension for (WDP). On the other hand, it reminds us that (WDP) is still not an ordinary constrained optimization problem in many cases, although it is no longer an MPEC.

Delightfully, we observed that, among the five cases mentioned above, the (WDP) ap-

Table 1: Numerical results by solving WDP and MPEC as nonlinear programs

Problem	Method	Upper-Level Objective Value	Infeasibility	CPU Time
1	MPCC	-3.1463e+12	2.1900e-11	1.4554e+02
	WDP	-3.3923e+08	5.6843e-14	6.4458e+01
2	MPCC	-2.5857e+09	4.8005e-07	7.7537e+01
	WDP	-3.0284e+13	1.4604e+09	1.0761e+02
3	MPCC	-3.7596e+12	9.0261e-05	2.4301e+02
	WDP	-2.2363e+22	3.2708e+10	2.3440e+02
4	MPCC	-1.0616e+12	1.3316e+05	1.0887e+02
	WDP	-2.9227e+21	1.6209e+07	1.6791e+02
5	MPCC	-1.1837e+10	1.7053e-13	4.9124e+02
	WDP	-6.3853e+17	8.1583e+05	9.1536e+01
6	MPCC	-2.5239e+05	6.2759e+01	1.1706e+02
	WDP	-1.5364e+08	4.0963e-08	6.3872e+01
7	MPCC	-1.0972e+10	2.6241e-06	1.1103e+02
	WDP	-2.3061e+09	3.0972e-08	3.7846e+01
8	MPCC	-1.1108e+10	1.8574e-04	1.6481e+02
	WDP	-6.3271e+06	7.1054e-15	4.6037e+02
9	MPCC	-3.5679e+11	4.4173e+02	1.3203e+02
	WDP	-3.8552e+12	2.3625e-04	9.3630e+01
10	MPCC	-1.6421e+11	5.1916e+04	9.1017e+01
	WDP	-3.1135e+13	3.1974e-04	1.2583e+02
11	MPCC	-8.8731e+09	4.1478e-06	3.5987e+02
	WDP	-4.9310e+12	4.4921e-06	3.6791e+02
12	MPCC	-5.9250e+00	6.1676e+02	5.8209e+01
	WDP	-1.3248e+10	3.1468e+02	7.7642e+02
13	MPCC	-3.7345e+10	1.8982e-06	6.2414e+02
	WDP	-7.3841e+10	2.2366e+02	4.4411e+02
14	MPCC	-4.9281e+00	3.3777e+02	6.2954e+01
	WDP	-6.5233e+12	1.4604e-01	3.2927e+02
15	MPCC	-6.4304e+08	3.5476e-08	6.7769e+02
	WDP	-3.3281e+10	1.2171e+01	2.8304e+02
16	MPCC	-1.2443e+01	3.5345e+02	1.1613e+02
	WDP	-9.6905e+12	1.0489e+02	3.1826e+02
17	MPCC	-7.2061e+09	8.1419e+00	8.5065e+02
	WDP	-2.5564e+11	5.9209e+02	4.0125e+02
18	MPCC	-1.0965e+08	7.0720e-08	3.1691e+02
	WDP	-4.3347e+07	3.3506e-08	1.4159e+02
19	MPCC	-4.2493e+00	5.9103e+02	5.2648e+01
	WDP	-7.2757e+15	1.9079e+02	7.5048e+02
20	MPCC	-2.1158e+08	2.0894e-07	4.7101e+02
	WDP	-5.2877e+10	8.5212e-06	9.7612e+01

^a Problems 1-10 were generated with $n = 50, p = 40; m = 60, q = 50$.^b Problems 11-20 were generated with $n = 50, p = 40; m = 100, q = 80$.

proach had greater advantages than the MPEC approach in the CPU times. Our experiments on relaxation methods given in the next subsection show that the (WDP) approach had overall advantages over the MPEC approach in both objective values and CPU times in our tests.

6.3 Solving (WDP) and MPEC by relaxation methods

In this subsection, we report our numerical results in solving (WDP) and MPEC by relaxation methods. The relaxation method for (WDP) has been studied in Section 5, while the relaxation scheme for MPEC employed here follows from the work [39], by using

$$\begin{aligned} \text{MPEC}(t) \quad \min \quad & c_1^T x + c_2^T y \\ \text{s.t.} \quad & A_1 x \leq b_1, \quad d_2 + B_2^T u_1 + u_2 - u_3 = 0, \\ & u_1 \geq 0, \quad u_2 \geq 0, \quad u_3 \geq 0, \quad A_2 x + B_2 y \leq b_2, \quad l_b \leq y \leq u_b, \\ & u_1^T (A_2 x + B_2 y - b_2) + u_2^T (y - u_b) + u_3^T (l_b - y) \leq t \end{aligned}$$

to approximate the MPEC (6.2), where $t > 0$ is an approximation parameter.

We implemented Algorithm 1 given below to solve the tested problems.

Algorithm 1

Step 0 Choose \tilde{x}^0 such that $A_1 \tilde{x}^0 \leq b_1$, an initial parameter $t_0 > 0$, a parameter $\sigma \in (0, 1)$, and some accuracy tolerances $\{\epsilon, \epsilon_{\text{SQP}}, \delta_{\text{min}}\}$. Set $k := 0$.

Step 1 Solve the lower-level program in (6.1) with $x = \tilde{x}^k$ by *linprog* to get \tilde{y}^k and \tilde{u}^k . Set $\tilde{w}^k := (\tilde{x}^k, \tilde{y}^k, \tilde{z}^k, \tilde{u}^k)$ in the case of (WDP) and $\tilde{w}^k := (\tilde{x}^k, \tilde{y}^k, \tilde{u}^k)$ in the case of MPEC. If \tilde{w}^k satisfies an termination criterion with ϵ as the accuracy tolerance, stop. Otherwise, go to Step 2.

Step 2 Solve WDP(t_k) or MPEC(t_k) by *fmincon* with \tilde{w}^k as the starting point and ϵ_{SQP} as the accuracy tolerance to obtain an iterate $w^k := (x^k, y^k, z^k, u^k)$ or $w^k := (x^k, y^k, u^k)$. If w^k satisfies the above termination criterion, stop. Otherwise, go to Step 3.

Step 3 Set $\tilde{x}^{k+1} := \tilde{x}^k$ and $t_{k+1} := \max\{\sigma t_k, \delta_{\text{min}}\}$. Let $k := k + 1$ and go to Step 1.

Note that, for the linear bilevel programs, the feasibility of \tilde{w}^k to (WDP) or MPEC are evident. In our experiments, we chose \tilde{x}^0 to be the origins and set $t_0 = 1$, $\sigma = 10^{-1}$, $\epsilon = 10^{-8}$, and $\epsilon_{\text{SQP}} = \delta_{\text{min}} = 10^{-16}$. The termination criterions for both (WDP) and MPEC were defined as the residuals of some kinds of optimality conditions. See Table 2 for the numerical results on the tested problems.

From Table 2, we can see that there were 18 and 15 cases with sufficient feasibility for the (WDP) and MPEC approaches respectively. Among the 14 cases with sufficient feasibility for both, the (WDP) approach had overall advantages over the MPEC approach in both objective values and CPU times. In particular, compared with Table 1, the CPU times did not increase significantly. This indicates that, at least in our tests, the relaxation method based on the (WDP) reformulation was quite efficient.

Table 2: Numerical results for WDP and MPEC by relaxation methods

Problem	Method	Upper-Level Objective Value	Infeasibility	CPU Time
1	MPCC	-3.6778e+22	5.6843e-14	4.8954e+02
	WDP	-1.5743e+24	5.6843e-14	1.6175e+03
2	MPCC	-2.1392e+11	0.0000e+00	9.5534e+02
	WDP	-1.0887e+18	0.0000e+00	8.8564e+02
3	MPCC	-2.9363e+11	6.8764e-08	1.8591e+02
	WDP	-5.8726e+10	2.8422e-14	1.0515e+02
4	MPCC	-2.5200e+08	0.0000e+00	1.8158e+02
	WDP	-1.7895e+10	2.7830e-11	1.1953e+03
5	MPCC	-5.2419e+10	1.1377e-11	1.2217e+03
	WDP	-2.6210e+11	3.3307e-16	3.9862e+02
6	MPCC	-3.3349e+10	1.6620e-11	1.0940e+02
	WDP	-1.7702e+15	5.6843e-14	1.1025e+02
7	MPCC	-1.1298e+10	5.6843e-14	9.9915e+02
	WDP	-3.5136e+16	5.6843e-14	8.3334e+02
8	MPCC	-1.4392e+16	0.0000e+00	3.4346e+03
	WDP	-3.3058e+10	1.1559e-11	1.4444e+03
9	MPCC	-7.0267e+09	6.8374e-07	1.0926e+03
	WDP	-2.3466e+15	5.6843e-14	1.1597e+02
10	MPCC	-3.4852e+17	2.8422e-14	3.0776e+03
	WDP	-2.3240e+17	2.8422e-14	8.2931e+01
11	MPCC	-5.5652e+10	1.8271e-12	1.2597e+03
	WDP	-1.2063e+11	0.0000e+00	7.7513e+02
12	MPCC	-6.6887e+10	9.2383e-07	1.4390e+03
	WDP	-2.7658e+11	2.2737e-13	2.7867e+03
13	MPCC	-1.0222e+10	9.9172e-07	2.0430e+03
	WDP	-5.1128e+10	1.3984e-12	7.0663e+02
14	MPCC	-1.6065e+10	1.7830e-12	2.3482e+03
	WDP	-7.3393e+11	5.6843e-14	2.5877e+03
15	MPCC	-1.5798e+11	0.0000e+00	7.3459e+03
	WDP	-3.9495e+12	2.1346e-32	5.8045e+02
16	MPCC	-1.0787e+11	1.2280e-10	1.9163e+03
	WDP	-3.3889e+08	5.6843e-14	3.1293e+02
17	MPCC	-2.5465e+19	0.0000e+00	2.4133e+03
	WDP	-6.3969e+10	4.7536e-07	6.6342e+03
18	MPCC	-2.7215e+10	7.4023e-13	1.0163e+03
	WDP	-3.2015e+11	5.6843e-14	2.1804e+03
19	MPCC	-1.4700e+10	9.2752e-13	1.3563e+03
	WDP	-5.6359e+07	1.1369e-13	9.1482e+02
20	MPCC	-7.6390e+10	4.2411e-07	4.5808e+02
	WDP	-1.5278e+10	3.1483e-09	2.2327e+03

7 Conclusions

We have shown that, under mild conditions such as pseudoconvexity or invexity, the bilevel program (BP) is equivalent to the new single-level optimization problem (WDP) in the globally or locally optimal sense. Unlike the existing single-level reformulations (SP), (VP), and (LDP), (WDP) does not involve any function without analytic expression. Compared with another existing single-level reformulation MPEC, we have shown by an example that (WDP) may satisfy the MFCQ at its feasible points, unlike the MPEC that fails to satisfy the MFCQ at any feasible point, and hence (WDP) is relatively easy to deal with. Based on our realization that (WDP) is still not an ordinary constrained optimization problem in some cases, we have proposed a relaxation method for solving (WDP) and established a convergence analysis. Comprehensive numerical experiments indicate that, although solving (WDP) directly did not perform very well in our tests, the relaxation method based on the (WDP) reformulation was quite efficient. As a future work, we will devote to analyzing the (WDP) reformulation from theoretical and numerical aspects so as to develop more efficient algorithms for the bilevel program (BP).

References

- [1] Bard, J.F.: Practical Bilevel Optimization: Algorithms and Applications. Kluwer, Dordrecht (1998)
- [2] Clarke, F.H.: Optimization and Nonsmooth Analysis. SIAM, Philadelphia (1990)
- [3] Colson, B., Marcotte, P., Savard, G.: An overview of bilevel optimization. *Ann. Oper. Res.* **153**(1), 235–256 (2007)
- [4] Craven, B.D., Mond, B.: On converse duality in nonlinear programming. *Oper. Res.* **19**(4), 1075–1078 (1971)
- [5] Davis, C.: Theory of positive linear dependence. *Am. J. Math.* **76**(4), 733–746 (1954)
- [6] Dempe, S.: Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints. *Optimization* **52**, 333–359 (2003)
- [7] Dempe, S.: Foundations of Bilevel Programming. Kluwer, Dordrecht (2002)
- [8] Dempe, S., Dutta, J.: Is bilevel programming a special case of a mathematical program with complementarity constraints? *Math. Program.* **131**(1-2), 37–48 (2012)
- [9] Dempe, S., Dutta, J., Mordukhovich, B.: New necessary optimality conditions in optimistic bilevel programming. *Optimization* **56**(5-6), 577–604 (2007)

- [10] Dempe, S., Franke, S.: Solution of bilevel optimization problems using the KKT approach. *Optimization* **68**(8), 1471–1489 (2019)
- [11] Dempe, S., Zemkoho, A.B.: The bilevel programming problem: reformulations, constraint qualifications and optimality conditions. *Math. Program.* **138**(1-2), 447–473 (2013)
- [12] Dempe, S., Zemkoho, A.B.: The generalized mangasarian-fromowitz constraint qualification and optimality conditions for bilevel programs. *J. Optim. Theory Appl.* **148**(1), 46–68 (2011)
- [13] Franceschi, L., Frasconi, P., Salzo, S., Grazzi, R., Pontil, M.: Bilevel programming for hyperparameter optimization and meta-learning. *Proceedings of International Conference on Machine Learning*, 1563–1572 (2018)
- [14] Fukushima, M., Lin, G.H.: Smoothing methods for mathematical programs with equilibrium constraints. *Proceedings of the ICKS'04, IEEE Computer Society*, 206–213 (2004)
- [15] Fukushima, M., Luo, Z.Q., Pang, J.S.: A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints. *Comput. Optim. Appl.* **10**(1), 5–34 (1998)
- [16] Izmailov, A.F., Pogosyan, A.L., Solodov, M.V.: Semismooth Newton method for the lifted reformulation of mathematical programs with complementarity constraints. *Comput. Optim. Appl.* **51**(1), 199–221 (2012)
- [17] Izmailov, A.F., Solodov, M.V., Uskov, E.I.: Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. *SIAM J. Optim.* **22**(4), 1579–1606 (2012)
- [18] Guo, L., Lin, G.H., Ye, J.J.: Solving mathematical programs with equilibrium constraints. *J. Optim. Theory Appl.* **166**(1), 234–256 (2015)
- [19] Hansen, P., Jaumard, B., Savard, G.: New branch-and-bound rules for linear bilevel programming. *SIAM J. Sci. Stat. Comput.* **13**(5), 1194–1217 (1992)
- [20] Hanson, M., Mond, B.: Further generalizations of convexity in mathematical programming. *J. Inform. Optim. Sci.* **3**(1), 25–32 (1982)
- [21] Hoheisel, T., Kanzow, C., Schwartz, A.: Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Math. Program.* **137**(1-2), 257–288 (2013)
- [22] Jeroslow, R.G.: The polynomial hierarchy and a simple model for competitive analysis. *Math. Program.* **32**(2), 146–164 (1985)

- [23] Kannappan, P.: Duality theorems for convex programming without constraint qualification. *J. Austral. Math. Soc.* **36**(2), 253–266 (1984)
- [24] Kunapuli, G., Bennett, K.P., Hu, J., Pang, J.S.: Classification model selection via bilevel programming. *Optim. Methods Software* **23**(4), 475–489 (2008)
- [25] Lampariello, L., Sagratella, S.: Numerically tractable optimistic bilevel problems. *Comput. Optim. Appl.* **76**, 277–303 (2020)
- [26] Lin, G.H., Chen, X., Fukushima, M.: Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization. *Math. Program.* **116**(1-2), 343–368 (2009)
- [27] Lin, G.H., Fukushima, M.: Hybrid approach with active set identification for mathematical programs with complementarity constraints. *J. Optim. Theory Appl.* **128**(1), 1–28 (2006)
- [28] Lin, G.H., Xu, M., Ye, J.J.: On solving simple bilevel programs with a nonconvex lower level program. *Math. Program.* **144**(1-2), 277–305 (2014)
- [29] Liu, R., Mu, P., Yuan, X., Zeng, S., Zhang, J.: A generic first-order algorithmic framework for bi-level programming beyond lower-level singleton. *Proceedings of International Conference on Machine Learning*, 6305–6315 (2020)
- [30] Loridan, P., Morgan, J.: Weak via strong stackelberg problem: new results. *J. Glob. Optim.* **8**(3), 263–287 (1996)
- [31] Luo, Z.Q., Pang, J.S., Ralph, D.: *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge (1996)
- [32] Osuna-Gómez, R., Rufián-Lizana, A., Ruiz-Canales, P.: Invex functions and generalized convexity in multiobjective programming. *J. Optim. Theo. Appl.* **98**(3), 651–661 (1998)
- [33] Ouattara, A., Aswani, A.: Duality approach to bilevel programs with a convex lower level. *Proceedings of the Annual American Control Conference*, 1388–1395. IEEE (2018)
- [34] Outrata, J.V.: On the numerical solution of a class of stackelberg problems. *Z. Oper. Res.* **34**(4), 255–277 (1990)
- [35] Pang, J.S., Fukushima, M.: Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints. *Comput. Optim. Appl.* **13**(1-3), 111–136 (1999)
- [36] Polak, E.: *Optimization: Algorithms and Consistent Approximations*. Springer-Verlag, New York (1997)

- [37] Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. *SIAM J. Optim.* **10**(4), 963–981 (2000)
- [38] Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Math. Oper. Res.* **25**(1), 1–22 (2000)
- [39] Scholtes, S.: Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM J. Optim.* **11**(4), 918–936 (2001)
- [40] Stackelberg, H.: *The Theory of the Market Economy*. Oxford University Press (1952)
- [41] Steffensen, S., Ulbrich, M.: A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM J. Optim.* **20**(5), 2504–2539 (2010)
- [42] Weir, T., Mond, B.: Pre-invex functions in multiple objective optimization. *J. Math. Anal. Appl.* **136**(1), 29–38 (1988)
- [43] Wolfe, P.: A duality theorem for non-linear programming. *Quart. Appl. Math.* **19**(3), 239–244 (1961)
- [44] Xu, M., Ye, J.J.: A smoothing augmented lagrangian method for solving simple bilevel programs. *Comput. Optim. Appl.* **59**(1-2), 353–377 (2014)
- [45] Ye, J.J.: Necessary optimality conditions for multiobjective bilevel programs. *Math. Oper. Res.* **36**(1), 165–184 (2011)
- [46] Ye, J.J., Zhu, D.: Optimality conditions for bilevel programming problems. *Optimization* **33**(1), 9–27 (1995)
- [47] Ye, J.J., Zhu, D.: New necessary optimality conditions for bilevel programs by combining the mpec and value function approaches. *SIAM J. Optim.* **20**(4), 1885–1905 (2010)