SIAM J. OPTIM.
Vol. 20, No. 4, pp. 1885–1905

NEW NECESSARY OPTIMALITY CONDITIONS FOR BILEVEL PROGRAMS BY COMBINING THE MPEC AND VALUE FUNCTION APPROACHES*

JANE J. YE† AND DAOLI ZHU‡

Abstract. The bilevel program is a sequence of two optimization problems where the constraint region of the upper level problem is determined implicitly by the solution set to the lower level problem. The classical approach to solving such a problem is to replace the lower level problem by its Karush–Kuhn–Tucker (KKT) condition and solve the resulting mathematical programming problem with equilibrium constraints (MPEC). In general the classical approach is not valid for nonconvex bilevel programming problems. The value function approach uses the value function of the lower level problem to define an equivalent single level problem. But the resulting problem requires a strong assumption, such as the partial calmness condition, for the KKT condition to hold. In this paper we combine the classical and the value function approaches to derive new necessary optimality conditions under rather weak conditions. The required conditions are even weaker in the case where the classical approach or the value function approach alone is applicable.

Key words. necessary optimality conditions, partial calmness, constraint qualifications, nonsmooth analysis, value function, bilevel programming problems

AMS subject classifications. 90C46, 90C26

DOI. 10.1137/080725088

1. Introduction.

In this paper we consider the following bilevel programming problem (BLPP):

$$\begin{aligned} \text{BLPP} \qquad & \min_{x,y} F(x,y) \\ & \text{s.t. } y \in S(x), \\ & \qquad G(x,y) \le 0, \end{aligned}$$

where $S(x)$ denotes the set of solutions of the lower level problem,

$$\begin{aligned} P_x: \qquad & \min_{y} f(x,y) \\ & \text{s.t. } g(x,y) \le 0, \end{aligned}$$

and $F, f : R^n \times R^m \to R$, $G : R^n \times R^m \to R^q$, $g : R^n \times R^m \to R^p$. We allow p or q to be zero to signify the case in which there are no explicit inequality constraints. In these cases it is clear below that certain references to such constraints are simply to be deleted.

To simplify the exposition and to concentrate on the main ideas, we assume that F, G, f, g are twice continuously differentiable mappings on finite dimensional spaces, and we do not include equality constraints. The results can be easily generalized to the

*Received by the editors May 27, 2008; accepted for publication (in revised form) November 19, 2009; published electronically March 3, 2010.
http://www.siam.org/journals/siopt/20-4/72508.html
†Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 3R4 (janeye@uvic.ca). The work of this author was partly supported by NSERC.
‡School of Management, Fudan University, Shanghai 200433, People's Republic of China (dlzhu@fudan.edu.cn). This research was supported by the National Natural Science Foundation of China, NSFC.

case in the presence of equality constraints in a straightforward manner. The results can be generalized to the case where F, G, f, g are Lipschitz continuous mappings on infinite dimensional spaces using appropriate variational analysis (see, e.g., [14, 15]).

Although the bilevel programming problem was introduced to the optimization community only in the 1970s by Bracken and McGill [4], the first formulation of a simpler case was introduced and used on the market economy by von Stackelberg [19] in 1934 and hence is also known in economic game theory as a Stackelberg game. Bilevel programming problems can be used to model a two-level hierarchical system where the two decision makers make their decisions on different levels of hierarchy. The game can be described as follows: For any possible decision vector $x \in R^n$ chosen by the leader, the follower will react optimally by choosing his decision vector $y \in R^m$ to minimize his objective function $f(x, y)$ subject to constraints $g(x, y) \leq 0$. If the lower level solution set $S(x)$ is a singleton, then the leader chooses x (consequently the follower chooses $y = S(x)$) such that the leader's objective function $F(x, S(x))$ is minimized subject to the constraints $G(x, S(x)) \leq 0$. However, if the solution set $S(x)$ of the lower level problem is not a singleton, two approaches have been taken in the literature. The first is to assume that the follower cooperates with the leader. In this optimistic version of the bilevel problem, the follower allows the leader to choose which of them is actually used, and hence the leader is to choose his optimal decision vector $x \in R^n$ (consequently the follower will choose $y \in S(x)$) such that the leader's objective function $F(x, y)$ is minimized for all (x, y) with $y \in S(x)$ and $G(x, y) \leq 0$. The problem BLPP models this optimistic version. The second approach argues for a conservative strategy that redefines the $\min_{x,y}$ in the BLPP as $\min_x \max_y$. This pessimistic approach assumes the worst, that there is no cooperation from the follower. Bilevel programming has been an important research area over the last twenty years, and many researchers have made contributions to the area. The reader is referred to monographs [2, 8, 18] for applications of bilevel programming and recent developments on the subject, and to [9, 20] for a bibliography review.

In this paper we adopt the following standard notation. For any two vectors a, b in R^n, we denote by $\langle a, b\rangle$ their inner product. Given a function $F : R^n \to R^m$, we denote its Jacobian by $\nabla F(z) \in R^{m\times n}$. If $m = 1$, the gradient $\nabla F(z) \in R^n$ is considered as a column vector. Given a function $f : R^n \times R^m \to R$, we define $\nabla^2_{xy} f(x, y) := \nabla_x(\nabla_y f(x, y))$, $\nabla^2_{yy} f(x, y) := \nabla_y(\nabla_y f(x, y))$. For a subset $A \subseteq R^n$, we denote by int$A$, co$A$ the interior and the convex hull of A, respectively. For a matrix $A \in R^{n\times m}$, $A^\top$ is its transpose.

To illustrate our approach we first review some background material for nonsmooth analysis. Let $\varphi : R^n \to R$ be Lipschitz continuous near $\bar{x}$; the directional derivative and the Clarke generalized directional derivative of φ at $\bar{x}$ in direction $d \in R^n$ are defined by

$$\varphi'(\bar{x}; d) := \lim_{t\downarrow 0} \frac{\varphi(\bar{x} + td) - \varphi(\bar{x})}{t},$$
$$\varphi^\circ(\bar{x}; d) := \limsup_{x\to\bar{x}, t\downarrow 0} \frac{\varphi(x + td) - \varphi(x)}{t},$$

respectively. The Clarke generalized gradient at $\bar{x}$ is a convex and compact subset of R^n defined by

$$\partial\varphi(\bar{x}) := \left\{\xi \in R^n : \xi^\top d \leq \varphi^\circ(\bar{x}; d) \quad \forall d \in R^n\right\}.$$

When φ is convex, the Clarke generalized gradient coincides with the subgradient in the sense of convex analysis, i.e.,

$$\partial\varphi(\bar{x}) := \left\{\xi \in R^n : \xi^\top(x-\bar{x}) \le \varphi(x) - \varphi(\bar{x}) \quad \forall x \in R^n\right\}.$$

When φ is continuously differentiable, the Clarke generalized gradient is reduced to the usual gradient, i.e., $\partial\varphi(\bar{x}) = \{\nabla\varphi(\bar{x})\}$. The following calculation rules for Clarke generalized gradients will be useful in the paper.

PROPOSITION 1.1 (see [5, 6]). *Let $f, g : R^n \to R$ be Lipschitz continuous near $\bar{x} \in R^n$ and α, β be any real numbers. Then*

$$\partial(\alpha f + \beta g)(\bar{x}) \subseteq \alpha\partial f(\bar{x}) + \beta\partial g(\bar{x}).$$

1.1. The first order approach and MPEC. The classical Karush–Kuhn–Tucker (KKT) approach (also called the first order approach) for deriving necessary optimality conditions for BLPP (see, e.g., [3]) was to consider the following one-level optimization problem where the lower level problem has been replaced by its KKT conditions:

$$\begin{aligned}
\text{KP} \quad & \min_{x,y,u} F(x,y) \\
& \text{s.t. } \nabla_y f(x,y) + u\nabla_y g(x,y) = 0, \\
& \qquad g(x,y) \le 0, \quad u \ge 0, \quad \langle g(x,y), u\rangle = 0, \\
& \qquad G(x,y) \le 0,
\end{aligned}$$

where $u\nabla_y g(x,y) := \sum_{i=1}^p u_i \nabla_y g_i(x,y)$. This approach, however, is only applicable to *convex* bilevel programming problems, i.e., those bilevel programming problems where the lower level problem is convex; i.e., $f(x,\cdot)$, $g_i(x,\cdot)$ are convex functions, and a certain constraint qualification is satisfied for the lower level. Moreover, the resulting one-level problem KP belongs to the class of mathematical programs with equilibrium constraints, or MPECs [12, 16], defined as follows:

$$\begin{aligned}
\text{MPEC} \quad & \min f(z) \\
& \text{s.t. } G(z) \ge 0,\ H(z) \ge 0,\ G(z)^\top H(z) = 0, \\
& \qquad g(z) \le 0,\ h(z) = 0,
\end{aligned}$$

where $f : R^n \to R, G, H : R^n \to R^m,\ g : R^n \to R^p,\ h : R^n \to R^q$. For simplicity and easy reference in this section we assume that f is Lipschitz near z^* and that all other functions are continuously differentiable. It is known that the usual constraint qualifications such as Mangasarian Fromovitz constraint qualification (MFCQ) will never hold (see [30, Proposition 1.1]). Recently some optimality conditions for MPECs and the corresponding constraint qualifications have been developed. For easy reference we review only those that will be used in this paper; see [24, 26] for more detailed discussions.

Given a feasible vector z^* of MPEC, we define the following index sets:

$$\begin{aligned}
& I_g := \{i : g_i(z^*) = 0\}, \\
& \alpha := \alpha(z^*) = \{i : G_i(z^*) = 0, H_i(z^*) > 0\}, \\
& \beta := \beta(z^*) = \{i : G_i(z^*) = 0, H_i(z^*) = 0\}, \\
& \gamma := \gamma(z^*) = \{i : G_i(z^*) > 0, H_i(z^*) = 0\}.
\end{aligned}$$

DEFINITION 1.1 (MPEC stationary conditions). *A feasible point z^* of MPEC is called a strong stationary point (S-stationary point) if there exists $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in R^{p+q+2m}$ such that the following conditions hold:*

$$(1.1)\quad 0 \in \partial f(z^*) + \sum_{i \in I_g} \lambda_i^g \nabla g_i(z^*) + \sum_{i=1}^{q} \lambda_i^h \nabla h_i(z^*) - \sum_{i=1}^{m} [\lambda_i^G \nabla G_i(z^*) + \lambda_i^H \nabla H_i(z^*)],$$

$$(1.2)\quad \lambda_i \geq 0, \quad i \in I_g, \qquad \lambda_i^G = 0, \quad i \in \gamma, \qquad \lambda_i^H = 0, \quad i \in \alpha,$$
$$\lambda_i^G \geq 0,\ \lambda_i^H \geq 0 \quad \forall i \in \beta.$$

A feasible point z^* of MPEC is called a piecewise stationary point (P-stationary point) if for each partition of the index set β into P, Q there exists $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in R^{p+q+2m}$ such that (1.1)–(1.2) *and the the following condition hold:*

$$\lambda_i^G \geq 0 \quad \forall i \in P, \qquad \lambda_i^H \geq 0 \quad \forall i \in Q.$$

Remark 1.1. Actually it is well known that the S-stationary condition is equivalent to the classical KKT condition for MPEC. A P-stationary point is equivalent to a Bouligrand stationary (B-stationary) point in the sense of [17] and is equivalent to a B-stationary point in the classical sense of [12] if a certain constraint qualification for each branch of the MPEC holds.

DEFINITION 1.2 (MPEC constraint qualifications). *Let z^* be a feasible point of MPEC. We say that the MPEC linear independence constraint qualification (MPEC LICQ) holds at z^* if the gradient vectors*

$$\nabla g_i(z^*),\ \ i \in I_g, \quad \nabla h_i(z^*),\ \ i = 1, \ldots, q, \quad \nabla G_i(z^*),\ \ i \in \alpha \cup \beta, \quad \nabla H_i(z^*),\ \ i \in \gamma \cup \beta,$$

are linearly independent. We say that the MPEC linear constraint qualification (MPEC linear CQ) holds if all functions G, H, g, h are affine. We say that MPEC piecewise MFCQ holds at z^* if MFCQ holds at z^* for each branch of MPEC corresponding to partition P, Q of index set β defined as

$$\begin{array}{lll}
\text{MPEC}_{P\cup Q} & \min & f(z) \\
& \text{s.t.} & G_i(z) = 0, \quad i \in \alpha, \qquad H_i(z) = 0, \quad i \in \gamma, \\
& & G_i(z) \geq 0,\ H_i(z) = 0, \quad i \in P, \\
& & G_i(z) = 0,\ H_i(z) \geq 0, \quad i \in Q, \\
& & g(z) \leq 0,\ h(z) = 0.
\end{array}$$

THEOREM 1.1 (MPEC necessary optimality conditions). *If MPEC LICQ holds, then z^* is S-stationary. If either MPEC linear CQ or MPEC piecewise MFCQ holds, then z^* is P-stationary.*

Proof. Using the nonsmooth Lagrange multiplier rule of Clarke [5, Theorem 6.1.1], the results can be proved as for the smooth case in [12, 23]. ∎

1.2. Value function approach. In Ye and Zhu [28, 29], the following value function approach is taken to reformulating the BLPP. Define *the value function of the lower level problem* as an extended value function $V : R^n \to \bar{R}$ by

$$V(x) := \inf_y \{f(x, y) : \quad g(x, y) \leq 0\},$$

where $\bar{R} := R \cup \{-\infty\} \cup \{+\infty\}$ is the extended real line and $\inf\{\emptyset\} = +\infty$ by convention. Then it is obvious that the BLPP can be reformulated as the following one-level optimization problem involving the value function:

$$\text{(1.3)} \qquad \begin{aligned} \text{VP} \qquad & \min\ F(x,y) \\ & \text{s.t. } f(x,y) - V(x) \le 0, \\ & \quad g(x,y) \le 0, \\ & \quad G(x,y) \le 0. \end{aligned}$$

The above one-level problem is completely equivalent to the BLPP without any convexity assumption on the lower level problem. However, there are two issues that need to be addressed when using this approach. First, it is well known that $V(x)$ may not be differentiable in general even in the case where all defining functions f, g are continuously differentiable, and hence the problem VP is in general a nonsmooth problem. To use the generalized Lagrange multiplier rule of Clarke [5], $V(x)$ is required to be Lipschitz continuous. This issue can be addressed as follows.

Let $x \in R^n$. For any $y \in S(x)$ we denote the set of KKT multipliers for the lower level problem P_x at y as follows:

$$M^1(x,y) := \left\{ u \in R^p : \begin{array}{l} 0 = \nabla_y f(x,y) + \sum_{i=1}^p u_i \nabla_y g_i(x,y), \\ u \ge 0,\ \sum_{i=1}^p u_i g_i(x,y) = 0 \end{array} \right\}.$$

Recall that a set-valued map Y is called uniformly bounded around $\bar{x}$ if there exists a neighborhood U of $\bar{x}$ such that the set $\cup_{x\in U} Y(x)$ is bounded. The following result is a special case of Clarke [5, Theorem 6.5.2].

PROPOSITION 1.2 (Clarke). *Assume that the set-valued map $Y(x) := \{y \in R^m : g(x,y) \le 0\}$ is uniformly bounded around $\bar{x}$. Suppose that MFCQ holds at y' for all $y' \in S(\bar{x})$. Then the value function $V(x)$ is Lipschitz continuous near $\bar{x}$ and*

$$\partial V(\bar{x}) \subseteq coW(\bar{x}),$$

where

$$W(\bar{x}) := \{\nabla_x f(\bar{x},y') + u'\nabla_x g(\bar{x},y') : y' \in S(\bar{x}), u' \in M^1(\bar{x},y')\}. \tag{1.4}$$

Secondly, since the constraint (1.3) is actually an equality constraint, the nonsmooth MFCQ for the one-level problem SP will never be satisfied [28, Proposition 3.2]. We now explain this from another perspective.

PROPOSITION 1.3. *Let $(\bar{x},\bar{y})$ be a local solution of* VP. *Then the Fritz John–type necessary optimality condition for* VP *always holds trivially. That is, there always exists a nonzero abnormal multiplier for problem* VP.

Proof. The Fritz John–type necessary optimality condition for VP is the existence of $r \ge 0$, $\mu \ge 0$, η^g, η^h, η^G, η^H not all equal to zero such that

$$\begin{aligned} & 0 \in r\nabla F(\bar{x},\bar{y}) + \mu(\nabla f(\bar{x},\bar{y}) - \partial V(\bar{x}) \times \{0\}) + \nabla g(\bar{x},\bar{y})^T \eta^g + \nabla G(\bar{x},\bar{y})^T \eta^G, \\ & \eta_i^G \ge 0,\ i \in I_G, \quad \eta_i^G = 0,\ i \notin I_G, \\ & \eta_i^g \ge 0,\ i \in I_g, \quad \eta_i^g = 0,\ i \notin I_g. \end{aligned}$$

However, for all y satisfying $g(x,y) \le 0$, one has $f(x,y) - V(x) \ge 0$ and $f(x,y) - V(x) = 0$ if $y \in S(x)$; hence $(\bar{x},\bar{y})$ is a minimizer of the problem

$$\begin{aligned} \min_{x,y}\ & f(x,y) - V(x) \\ \text{s.t. }\ & g(x,y) \le 0. \end{aligned}$$

The generalized Lagrange multiplier rule of Clarke [5, Theorem 6.1.1] applied to the above problem implies the existence of $\mu \geq 0,\ \eta^g$ not all equal to zero such that

$$0 \in \mu(\nabla f(\bar{x}, \bar{y}) - \partial V(\bar{x})) + \nabla g(\bar{x}, \bar{y})^\top \eta^g,$$
$$\eta_i^g \geq 0,\ i \in I_g, \quad \eta_i^g = 0,\ i \notin I_g,$$

which implies that the Fritz John condition for problem VP holds trivially with $r = 0,\ \eta^G = 0$. □

To cope with the second difficulty, the following partial calmness condition was suggested by [28, 29] as a condition under which a local optimal solution of BLPP satisfies the generalized Lagrange multiplier rule of Clarke.

DEFINITION 1.3 (partial calmness for VP). *Let $(\bar{x}, \bar{y})$ be a local solution of* VP. *We say that* VP *is partially calm at $(\bar{x}, \bar{y})$ if there exists $\mu > 0$ such that $(\bar{x}, \bar{y})$ is a local solution of the following partially penalized problem:*

$$\begin{aligned} (\mathrm{VP})_\mu \quad & \min\ F(x,y) + \mu(f(x,y) - V(x)) \\ & \text{s.t. } g(x,y) \leq 0, \\ & \qquad G(x,y) \leq 0. \end{aligned}$$

Sufficient condition for partial calmness can be found in [22, 28, 29, 30]. Under the partial calmness condition, the constraint (1.3) which causes the usual constraint qualification to fail is moved to the objective function, and hence the generalized Lagrange multiplier rule of Clarke is applied to the problem $(\mathrm{VP})_\mu$ to obtain a KKT-type necessary optimality condition in [28, Theorem 3.1]. The value function approach was further developed in Ye [25, 27] using other constraint qualifications such as the Abadie constraint qualification. For the case where the value function is convex, it was shown in [25, 27] that the resulting KKT condition takes a simpler form in which only one solution of the lower level optimal problem is involved. Under the partial calmness condition this simpler KKT condition was proved to hold under the assumption of inner semicontinuity of the solution mapping of the lower level program [11].

Unfortunately optimal solutions of many *nonconvex* bilevel programming problems where the lower level problem is not convex do not satisfy the KKT conditions derived by using either the MPEC approach or the value function approach. In this paper we derive a new KKT condition by using a new approach which is a combination of the classical KKT and the value function approach. The resulting necessary optimality condition is much more likely to hold since it contains the ones derived by using the MPEC and the value function approach as special cases. In the case where the lower level problem is convex or the lower level problem has a unique solution and a unique multiplier at the optimal solution of the bilevel problem, the resulting KKT condition reduces to either the S-stationary condition or P-stationary condition by using the MPEC approach, and in the case where the multipliers corresponding to the KKT conditions of the lower level problem are zero, it reduces to the KKT condition derived by using the value function approach. And even in these cases our sufficient conditions for the KKT conditions to hold at a local optimal solution are new and weaker than some of the existing conditions.

We organize our paper as follows. In section 2, we use a simple example to illustrate our approach and in the meantime derive a KKT condition for the special case of the bilevel programming problem where the lower level has no constraints. In section 3 we provide our results for the general bilevel programming problem. We conclude our paper with some illustrative examples in section 4. In particular we

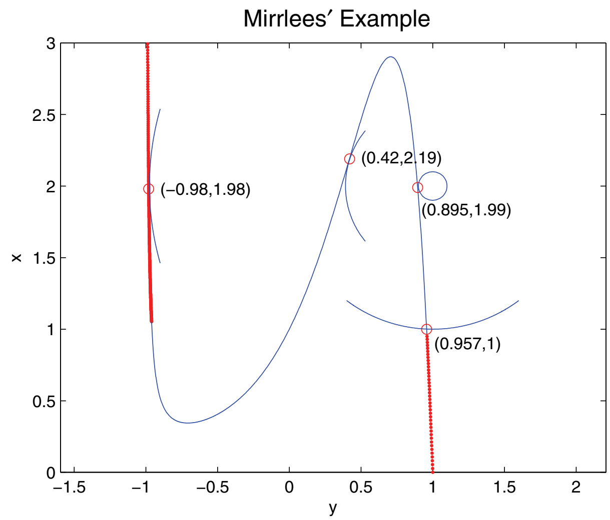

FIG. 1. *The first order approach fails.*

include an example for which our new necessary condition applies but neither the classical KKT condition nor the value function approach applies.

2. An example. To illustrate the difficulties of bilevel programming problems and our approach, we consider the following example from Mirrlees [13].

Example 2.1 (Mirrlees' problem).

$$\begin{array}{ll} \text{BLPP} & \min\ F(x,y) = (x-2)^2 + (y-1)^2 \\ & \text{s.t. } y \text{ minimizes } \ f(x,y) = -xe^{-(y+1)^2} - e^{-(y-1)^2}. \end{array}$$

The first order necessary condition for minimization of the lower level objective function with respect to y is

$$x(y+1)e^{-(y+1)^2} + (y-1)e^{-(y-1)^2} = 0.$$

Hence each x and y, stationary points of the lower level problem, are related by the equation

$$x = \frac{1-y}{1+y} e^{4y}, \tag{2.1}$$

which is a smooth and connected curve as shown in Figure 1.

Since the objective of the lower level problem is not convex in variable y, for each fixed x, not all corresponding y's lying on the curve are global optimal solutions of the lower level problem. The true global optimal solutions for the lower level problem are identified as a disconnected curve with a jump at $\bar{x} = 1$ (the darker curve in Figure 1), which represents the feasible region of the bilevel program.

Although at $\bar{x} = 1$ both $\bar{y} = 0.957$ and $y_1 = -0.957$ are the global optimal solutions of the lower level problem, the global optimal solution of the Mirrlees' problem is shown to be $\bar{x} = 1$, $\bar{y} = 0.957$. However, by the classical first order approach, one would consider the optimization problem with the first order condition (2.1) as a constraint and find three solutions $(x, y) = (0.895, 1.99)$, $(x, y) = (0.42, 2.19)$, $(x, y) = (-0.98, 1.98)$ to the Lagrange multiplier rule. Since these three solutions do not include the optimal solution $(\bar{x}, \bar{y}) = (1, 0.957)$, by using the first order approach, one would miss the optimal solution of the bilevel program completely.

The solution of Mirrlees' problem will not change if we add the constraint $y \in [-2, 2]$ into the lower level problem. Hence $\bar{x} = 1$, $\bar{y} = 0.957$ is still the optimal solution to the bilevel programming problem

$$\begin{aligned}&\min\ F(x, y) = (x-2)^2 + (y-1)^2\\ &\text{s.t. } y \in S(x),\end{aligned}$$

where $S(x)$ is the solution of the lower level problem

$$\begin{aligned}&\min_{y}\ f(x, y) := -xe^{-(y+1)^2} - e^{-(y-1)^2}\\ &\text{s.t. } y \in [-2, 2].\end{aligned}$$

The value function for the lower level problem is

$$V(x) = \min_{y \in [-2,2]} -xe^{-(y+1)^2} - e^{-(y-1)^2}.$$

By the value function approach, Mirrlees' problem (Example 2.1) is equivalent to the following single level optimization problem:

$$\begin{aligned}\text{VP}\qquad &\min\ F(x, y) = (x-2)^2 + (y-1)^2\\ &\text{s.t. } f(x, y) - V(x) \le 0,\\ &\qquad y \in [-2, 2].\end{aligned}$$

The following result can be obtained from Danskin's theorem, stated on page 99 of [6] (also see [7]).

PROPOSITION 2.1 (Danskin's theorem). *Let $S \subseteq R^m$ be a compact set and let $f(x, y)$ be a function on $R^n \times R^m$ that is continuously differentiable at $\bar{x}$. Then the value function*

$$V(x) = \min\{f(x, y) : y \in S\}$$

is Lipschitz continuous and directionally differentiable at $\bar{x}$. Let $S(\bar{x})$ denote the solution of the lower level problem. Then the Clarke generalized gradient at $\bar{x}$ and the directional derivative of V at $\bar{x}$ in the direction dx are

$$\partial V(\bar{x}) = co\{\nabla_x f(\bar{x}, y') : y' \in S(\bar{x})\}$$

and

$$V'(\bar{x}; dx) = \min\{\nabla_x f(\bar{x}, y')^T dx : y' \in S(\bar{x})\}.$$

By Danskin's theorem, the one-level problem VP is an optimization problem with Lipschitz continuous and continuously differentiable problem data. By the nonsmooth generalized Lagrange multiplier rule of Clarke [5, Theorem 6.1.1], there exist nonnegative scalars λ, μ not all zero such that

$$\begin{aligned}&0 \in \lambda \nabla_x F(\bar{x}, \bar{y}) + \mu[\nabla_x f(\bar{x}, \bar{y}) - \partial V(\bar{x})],\\ &0 = \lambda \nabla_y F(\bar{x}, \bar{y}) + \mu \nabla_y f(\bar{x}, \bar{y}).\end{aligned}$$

The above condition is a nonsmooth version of the Fritz John necessary optimality condition. In the case where the Lagrange multiplier corresponding to the objective function λ is equal to zero, the above condition does not include the objective function and hence does not provide any useful information about the optimality. We now show that this is the case for Mirrlees' problem. Since $\bar{y} = 0.957$ is an interior point of $[-2, 2]$ and is an optimal solution of the lower level problem, $\nabla_y f(\bar{x}, \bar{y}) = 0$. But $\nabla_y F(\bar{x}, \bar{y}) \neq 0$. So the above Lagrange multiplier rule holds only if $\lambda = 0$. This shows that for the equivalent problem obtained by using the value function approach, the KKT condition does not hold at the optimal solution. Hence the value function approach fails to provide a KKT necessary optimality condition.

To explain our new approach, we now consider the following optimization problem in which we add the first order condition to the problem VP:

$$\begin{aligned} \text{CP} \qquad & \min\ F(x,y) = (x-2)^2 + (y-1)^2 \\ & \text{s.t. } f(x,y) - V(x) \leq 0, \\ & \quad \nabla_y f(x,y) = 0, \\ & \quad y \in [-2, 2]. \end{aligned} \tag{2.2}$$

Since the first order condition holds at $(\bar{x}, \bar{y})$, the constraint (2.2) is redundant, and hence $(\bar{x}, \bar{y})$ is still the optimal solution of the above problem. By the nonsmooth generalized Lagrange multiplier rule of Clarke [5], there exist $\lambda \geq 0,\ \mu \geq 0,\ \beta \in R^m$ not all zero such that

$$\begin{aligned} &0 \in \lambda \nabla_x F(\bar{x}, \bar{y}) + \mu[\nabla_x f(\bar{x}, \bar{y}) - \partial V(\bar{x})] + \nabla^2_{xy} f(\bar{x}, \bar{y})^T \beta, \\ &0 = \lambda \nabla_y F(\bar{x}, \bar{y}) + \nabla^2_{yy} f(\bar{x}, \bar{y})^T \beta. \end{aligned}$$

As verified in Mirrlees [13], at $\bar{x} = 1$ both $\bar{y} = 0.957$ and $y_1 = -0.957$ are solutions of the lower level problem $P_{\bar{x}}$. Hence by Proposition 2.1,

$$\begin{aligned} \partial V(\bar{x}) &= co\{\nabla_x f(\bar{x}, \bar{y}), \nabla_x f(\bar{x}, y_1)\} \\ &= \{\alpha \nabla_x f(\bar{x}, \bar{y}) + (1-\alpha) \nabla_x f(\bar{x}, y_1) : \alpha \in [0,1]\}. \end{aligned}$$

Hence the nonsmooth generalized Lagrange multiplier rule of Clarke implies the existence of $\lambda \geq,\ \mu \geq 0,\ \beta \in R^m$ not all zero and $\alpha \in [0,1]$ such that

$$\begin{aligned} &0 = \lambda \nabla_x F(\bar{x}, \bar{y}) + \mu(1-\alpha)[\nabla_x f(\bar{x}, \bar{y}) - \nabla_x f(\bar{x}, y_1)] + \nabla^2_{yx} f(\bar{x}, \bar{y})^T \beta, \\ &0 = \lambda \nabla_y F(\bar{x}, \bar{y}) + \nabla^2_{yy} f(\bar{x}, \bar{y})^T \beta. \end{aligned}$$

It is now easy to verify that the above condition holds for $\lambda = 1,\ \mu = 2.05/1-\alpha,\ \beta = -0.04918,\ \alpha \in [0, 1)$. Since multipliers for both the first order condition and the value function constraint are not zero, it is clear that both the first order approach and the value function approach fail for Mirrlees' problem. Actually Mirrlees pointed out in [13] that there are many similar bilevel programming problems in the class of moral-hazard model principal-agent problems in economics, and hence it is extremely important to find an approach that will solve these types of problems.

Let us analyze why the value function approach fails even though the value function reformulation VP is completely equivalent to the original bilevel programming problem BLPP. Let us consider the perturbed objective function of the bilevel programming problem:

$$F_\mu(x,y) = F(x,y) + \mu(f(x,y) - V(x))$$

for some $\mu \geq 0$. By Danskin's theorem, since $\nabla_x f(\bar{x}, y_1) < \nabla_x f(\bar{x}, \bar{y}) < 0$ one has

$$V'(\bar{x}; dx) = \min\{\nabla_x f(\bar{x}, \bar{y})^T dx, \nabla_x f(\bar{x}, y_1)^T dx\} = \begin{cases} \nabla_x f(\bar{x}, y_1)^T dx & \text{if } dx > 0, \\ \nabla_x f(\bar{x}, \bar{y})^T dx & \text{if } dx \leq 0. \end{cases}$$

Hence the directional derivative of the perturbed objective function is

$$\begin{aligned} F'_\mu((\bar{x}, \bar{y}); (dx, dy)) &= \nabla F(\bar{x}, \bar{y})^T (dx, dy) + \mu[\nabla f(\bar{x}, \bar{y})^T (dx, dy) - V'(\bar{x}; dx)] \\ &= \nabla F(\bar{x}, \bar{y})^T (dx, dy) + \mu[\nabla_x f(\bar{x}, \bar{y})^T dx - V'(\bar{x}; dx)] \\ &= \begin{cases} \nabla F(\bar{x}, \bar{y})^T (dx, dy) + \mu(f_x(\bar{x}, \bar{y}) - f_x(\bar{x}, y_1))^T dx & \text{if } dx > 0 \\ \nabla F(\bar{x}, \bar{y})^T (dx, dy) & \text{if } dx \leq 0 \end{cases} \\ &= \begin{cases} -2dx - 0.86dy + \mu(e^{-0.043^2} - e^{-1.957^2})dx & \text{if } dx > 0 \\ -2dx - 0.86dy & \text{if } dx \leq 0. \end{cases} \end{aligned}$$

It is clear that for any $\mu \geq 0$ and any direction (dx, dy) such that $dx = 0, dy > 0$,

$$F'_\mu((\bar{x}, \bar{y}); (dx, dy)) < 0,$$

which implies that $(\bar{x}, \bar{y})$ is not a local minimum of the perturbed problem $\min F_\mu(x, y)$ for any $\mu \geq 0$, and hence the problem VP is not partially calm.

Let $M = \{(x, y) : \nabla_y f(x, y) = 0\}$ represent the set described by the first order condition (2.1). Then at the point $(\bar{x}, \bar{y}) \in M$, the tangent plane of M at point $(\bar{x}, \bar{y})$ is

$$T((\bar{x}, \bar{y}) : M) = \{(dx, dy) : \nabla^2_{xy} f(\bar{x}, \bar{y})dx + \nabla^2_{yy} f(\bar{x}, \bar{y})dy = 0\}.$$

From Figure 1, it is easy to see that for any nonzero direction $(dx, dy) \in T((\bar{x}, \bar{y}) : M)$, $\frac{dx}{dy} < 0$ and $|dx| > |dy|$. When $dx \leq 0,\ dy \geq 0$, and $|dx| \geq |dy|$, we have

$$F'_\mu((\bar{x}, \bar{y}); (dx, dy)) = 2(\bar{x} - 2)dx + 2(\bar{y} - 1)dy \geq 0,$$

and when $dx > 0, dy \leq 0$, since $\nabla_x f(\bar{x}, \bar{y}) - \nabla_x f(\bar{x}, y_1) > 0$ we can find $\mu > 0$ large enough such that

$$F'_\mu((\bar{x}, \bar{y}); (dx, dy)) = 2(\bar{x} - 2)dx + \mu[\nabla_x f(\bar{x}, \bar{y}) - \nabla_x f(\bar{x}, y_1)]^T dx + 2(\bar{y} - 1)dy \geq 0.$$

So in every direction of the tangent plane of the curve described by the first order condition, the perturbed objective function of the bilevel programming problem is nonnegative. That is, $d = (dx, dy) = (0, 0)$ is an optimal solution to the following linearized perturbed problem for some $\mu > 0$:

$$\begin{aligned} \min_d \quad & \Phi(d) := F'_\mu((\bar{x}, \bar{y}); d) \\ \text{s.t.} \quad & \nabla(\nabla_y f)(\bar{x}, \bar{y})d = 0. \end{aligned}$$

Since the directional derivative is a convex function as a function of the direction, the above problem is a convex program with linear constraints. Since $d = 0$ is an optimal solution, there exists $\xi \in \partial\Phi(0)$, the subgradient of Φ in the sense of convex analysis, and a multiplier β such that

$$0 = \xi + \nabla(\nabla_y f)(\bar{x}, \bar{y})^T \beta.$$

By the definition of subgradient in the sense of convex analysis, $\xi \in \partial\Phi(0)$ if and only if

$$F'_\mu((\bar{x}, \bar{y}); d) = \Phi(d) \geq \langle \xi, d \rangle \quad \forall d.$$

By the definition of the Clarke generalized gradient, the above inequality implies that $\xi \in \partial F_\mu(\bar{x}, \bar{y})$. By the calculus rule in Proposition 1.1 it follows that

$$\xi \in \partial F_\mu(\bar{x}, \bar{y}) \subseteq \nabla F(\bar{x}, \bar{y}) + \mu(\nabla f(\bar{x}, \bar{y}) - \partial V(\bar{x}) \times \{0\})$$

and consequently

$$0 \in \nabla F(\bar{x}, \bar{y}) + \mu(\nabla f(\bar{x}, \bar{y}) - \partial V(\bar{x}) \times \{0\}) + \nabla(\nabla_y f)(\bar{x}, \bar{y})^T \beta.$$

By Carathéodory's theorem, the convex subset $co\{\nabla_x f(\bar{x}, y') : y' \in S(\bar{x})\}$ of R^n can be represented by not more than $n+1$ elements at a time. Hence by Proposition 2.1,

$$\partial V(\bar{x}) = \left\{ \sum_{i=1}^{n+1} \lambda^i \nabla_x f(\bar{x}, y^i) : \lambda^i \geq 0, \sum_{i=1}^{n+1} \lambda^i = 1, y^i \in S(\bar{x}) \right\}.$$

Actually we have proved the following theorem.

THEOREM 2.1. *Let $(\bar{x}, \bar{y})$ be a local optimal solution to the bilevel programming problem*

$$\begin{aligned} &\min\ F(x, y) \\ &\quad y \in \arg\min_{y \in S} f(x, y), \end{aligned}$$

where S is a compact subset of R^m and F, f are continuously differentiable functions at $(\bar{x}, \bar{y})$. Suppose that $\bar{y} \in \text{int}S$ and there exists $\mu \geq 0$ such that

$$F'_\mu((\bar{x}, \bar{y}), d) \geq 0$$

for all $d = (dx, dy) \in R^{n+m}$ such that

$$\nabla^2_{yx} f(\bar{x}, \bar{y}) dx + \nabla^2_{yy} f(\bar{x}, \bar{y}) dy = 0.$$

Then there exist $\beta \in R^m$, $\lambda^i \geq 0$, $i = 1, \ldots, n+1$, $y^i \in S(\bar{x})$ such that $\sum_{i=1}^{n+1} \lambda^i = 1$ and

$$\begin{aligned} 0 &= \nabla_x F(\bar{x}, \bar{y}) + \mu \sum_{i=1}^{n+1} \lambda^i [\nabla_x f(\bar{x}, \bar{y}) - \nabla_x f(\bar{x}, y^i)] + \nabla^2_{xy} f(\bar{x}, \bar{y})^T \beta, \\ 0 &= \nabla_y F(\bar{x}, \bar{y}) + \nabla^2_{yy} f(\bar{x}, \bar{y})^T \beta. \end{aligned}$$

Remark 2.1. (i) According to [13], $(\bar{x}, \bar{y})$ is f-critical if $\bar{y} \in \arg\min_y f(\bar{x}, y)$, but there is no neighborhood N of $(\bar{x}, \bar{y})$ such that $(x', y') \in N, \nabla_y f(x', y') = 0$ imply that $y' \in \arg\min_y f(x', y)$. In [13, Theorem 2] Mirrlees showed that if $(\bar{x}, \bar{y})$, an optimal solution of the bilevel programming problem, is not f-critical, then the classical first order approach applies. For the case where an optimal solution of the bilevel programming problem is f-critical, such as in Mirrlees' problem, no general KKT condition was given in [13]. However by using a heuristic argument, Mirrlees suggested a KKT condition for the case of the bilevel programming problem for which there are two distinct maxima $\bar{y}$ and y^1 of $f(\bar{x}, \cdot)$, there being no others. It is easy to see that for this special case (omitting the upper level constraint $H = 0$ in Mirrlees' paper), the form of the KKT condition in (52) on page 17 of [13] coincides with our KKT condition in Theorem 2.1. Note that a rigorous proof of the KKT condition for this special case in [13] would require a constraint qualification, and our result in Theorem 2.1 fills this gap by providing a constraint qualification under which the KKT condition for the general case holds.

(ii) In Araujo and Moreira [1, Theorem 3.1 and Corollary 2], a similar form of the KKT condition for a bilevel program is given where the leader's decision x belongs to an infinite dimensional Banach space and the lower level decision y is a real number in a compact interval. Note that although our results are stated for the finite dimensional space for simplicity, it is possible for us to generalize them to the case where x belongs to an infinite dimensional space. There are two restrictions with the approach of Araujo and Moreira. The first one, as the authors pointed out [1, p. 27], is that it is not possible to extend [1, Theorem 3.1 and Corollary 2] to the multidimensional case using their method. The second restriction is the possibility of generalizing their method to the general bilevel program where the lower level has a constraint $g(x, y) \leq 0$ depending on the leader's decision variable x. Since the method of Araujo and Moreira is based on writing the bilevel program as a sequential minimization problem with respect to y and x, it is unlikely that the method can be generalized.

3. Combined MPEC and the value function approach. In this section we consider the general BLPP defined as in section 1. Analysis of Example 2.1 suggests that we should consider the following combined problem:

$$
\begin{aligned}
\text{CP} \qquad & \min_{x,y,u} && F(x, y) \\
& \quad \text{s.t.} && f(x, y) - V(x) \leq 0, \\
&&& \nabla_y f(x, y) + u \nabla_y g(x, y) = 0, \\
&&& g(x, y) \leq 0, \quad u \geq 0, \quad \langle g(x, y), u \rangle = 0, \\
&&& G(x, y) \leq 0.
\end{aligned}
\tag{3.1}
$$

We now study the relationship of CP and BLPP in the following proposition.

PROPOSITION 3.1. *Let $(\bar{x}, \bar{y})$ be a local (global) optimal solution to* BLPP. *Suppose that at $\bar{y}$ the KKT condition holds for the lower level problem $P_{\bar{x}}$. Then there exists $\bar{u}$ such that $(\bar{x}, \bar{y}, \bar{u})$ is a local (global) optimal solution of* CP. *Conversely, suppose that $(\bar{x}, \bar{y}, \bar{u})$ is an optimal solution to CP restricting on $U(\bar{x}, \bar{y}) \times R^p$, where $U(\bar{x}, \bar{y})$ is a neighborhood of $(\bar{x}, \bar{y})$ and the KKT condition holds at $y \in S(x)$ for lower level problem P_x for all (x, y) close to $(\bar{x}, \bar{y})$. Then $(\bar{x}, \bar{y})$ is a local solution of* BLPP.

Proof. Let $(\bar{x}, \bar{y})$ be a local optimal solution to BLPP. Then there exists $U(\bar{x}, \bar{y})$, a neighborhood of $(\bar{x}, \bar{y})$ such that

$$
F(\bar{x}, \bar{y}) \leq F(x, y) \qquad \forall (x, y) \in U(\bar{x}, \bar{y}) \cap \mathcal{F},
\tag{3.2}
$$

where $\mathcal{F}$ denotes the feasible region of BLPP. Since $\bar{y}$ is the global optimal solution of the lower level problem $P_{\bar{x}}$, by the KKT condition, there exists a multiplier $\bar{u}$ such that

$$\begin{aligned}
&0 = \nabla_y f(\bar{x}, \bar{y}) + \sum_{i=1}^{p} \bar{u}_i \nabla_y g_i(\bar{x}, \bar{y}),\\
&\bar{u} \geq 0, \quad \sum_{i=1}^{p} \bar{u}_i g_i(\bar{x}, \bar{y}) = 0.
\end{aligned}$$

Therefore $(\bar{x}, \bar{y}, \bar{u})$ is a feasible solution to problem CP. Now let (x, y, u) be a feasible solution to problem CP such that $(x, y) \in U(\bar{x}, \bar{y})$. Then it is obvious that (x, y) is a feasible solution of BLPP. By (3.2), $(\bar{x}, \bar{y}, \bar{u})$ is a local (global) optimal solution of CP.

Conversely, suppose that $(\bar{x}, \bar{y}, \bar{u})$ is an optimal solution to CP on $U(\bar{x}, \bar{y}) \times R^p$. Then

$$F(\bar{x}, \bar{y}) \leq F(x, y) \qquad \forall (x, y, u) \in (U(\bar{x}, \bar{y}) \times R^p) \cap \mathcal{F}_1, \tag{3.3}$$

where $\mathcal{F}_1$ is the feasible region of CP. Let $(x, y) \in U(\bar{x}, \bar{y})$ be a feasible solution of BLPP; then, by the assumption, the KKT condition holds, and hence there exists u such that (x, y, u) is a feasible solution of problem CP. The optimality of $(\bar{x}, \bar{y})$ for problem BLPP follows from (3.3). ∎

Suppose that the value function $V(x)$ is Lipschitz continuous near the optimal solution; then the problem CP is an MPEC with continuously differentiable and Lipschitz continuous problem data. However, due to the value function constraint (3.1), we can argue as in Proposition 1.3 that the usual MPEC constraint qualifications such as MPEC LICQ and MPEC piecewise MFCQ will never hold. Since the value function is usually not linear, the MPEC linear CQ is unlikely to hold as well. We propose the following partial calmness condition for CP.

DEFINITION 3.1 (partial calmness for CP). *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local solution of* CP. *We say that* CP *is partially calm at $(\bar{x}, \bar{y}, \bar{u})$ if there exists $\mu > 0$ such that $(\bar{x}, \bar{y}, \bar{u})$ is a local solution of the following partially penalized problem:*

$$\begin{aligned}
(\mathrm{CP})_\mu \quad & \min\ F(x, y) + \mu(f(x, y) - V(x))\\
& \text{s.t. } \nabla_y f(x, y) + u \nabla_y g(x, y) = 0,\\
& \qquad g(x, y) \leq 0,\\
& \qquad G(x, y) \leq 0.
\end{aligned} \tag{3.4}$$

Take as given a feasible vector $(\bar{x}, \bar{y}, \bar{u})$ in the feasible region of $(CP)_\mu$. We define the following index sets:

$$\begin{aligned}
I_G = I_G(\bar{x}, \bar{y}) &:= \{i : G_i(\bar{x}, \bar{y}) = 0\},\\
I_g = I_g(\bar{x}, \bar{y}, \bar{u}) &:= \{i : g_i(\bar{x}, \bar{y}) = 0, \bar{u}_i > 0\},\\
I_u = I_u(\bar{x}, \bar{y}, \bar{u}) &:= \{i : g_i(\bar{x}, \bar{y}) < 0, \bar{u}_i = 0\},\\
I_0 = I_0(\bar{x}, \bar{y}, \bar{u}) &:= \{i : g_i(\bar{x}, \bar{y}) = 0, \bar{u}_i = 0\}.
\end{aligned}$$

DEFINITION 3.2 (stationary conditions for CP based on the value function). *Let $(\bar{x}, \bar{y}, \bar{u})$ be a feasible solution to CP. We say that $(\bar{x}, \bar{y}, \bar{u})$ is an S-stationary point* *based on the value function if there exist* $\mu \geq 0,\ \beta \in R^m,\ \eta^g \in R^p,\ \eta^G \in R^q$ *such that*

$$0 \in \nabla F(\bar{x},\bar{y}) + \mu[\nabla f(\bar{x},\bar{y}) - \partial V(\bar{x}) \times \{0\}] + \nabla(\nabla_y f + \bar{u}\nabla_y g)(\bar{x},\bar{y})^T \beta + \nabla g(\bar{x},\bar{y})^T \eta^g + \nabla G(\bar{x},\bar{y})^T \eta^G, \tag{3.5}$$

$$\eta_i^G \geq 0, \quad i \in I_G, \quad \eta_i^G = 0, \quad i \notin I_G, \tag{3.6}$$

$$\eta_i^g = 0, \quad i \in I_u, \quad (\nabla_y g(\bar{x},\bar{y})\beta)_i = 0, \quad i \in I_g, \tag{3.7}$$
$$\eta_i^g \geq 0, \quad (\nabla_y g(\bar{x},\bar{y})\beta)_i \geq 0, \quad i \in I_0.$$

We say that $(\bar{x},\bar{y},\bar{u})$ *is a P-stationary point based on the value function if for each partition of the index set* I_0 *into* P, Q *there exist* $\mu \geq 0,\ \beta \in R^m,\ \eta^g \in R^p,\ \eta^G \in R^q$ *such that* (3.5)–(3.7) *and the following condition hold:*

$$\eta_i^g \geq 0, \quad i \in P, \qquad (\nabla_y g(\bar{x},\bar{y})\beta)_i \geq 0, \quad i \in Q.$$

Applying Theorem 1.1 to the problem $(CP)_\mu$, we obtain the following result immediately.

PROPOSITION 3.2. *Let* $(\bar{x},\bar{y},\bar{u})$ *be a local solution to CP. Suppose that the value function* $V(x)$ *is Lipschitz continuous near* $\bar{x}$ *and that* CP *is partially calm at* $(\bar{x},\bar{y},\bar{u})$.

(i) *If MPEC LICQ holds for* $(CP)_\mu$ *at* $(\bar{x},\bar{y},\bar{u})$, *then* $(\bar{x},\bar{y},\bar{u})$ *is an S-stationary point based on the value function.*

(ii) *If either MPEC linear CQ or MPEC piecewise MFCQ holds at* $(\bar{x},\bar{y},\bar{u})$, *then* $(\bar{x},\bar{y},\bar{u})$ *is a P-stationary point based on the value function.*

However, the partial calmness condition may be difficult to verify. In the rest of this section we introduce some conditions which are weaker and easier to verify than the partial calmness condition.

It can be shown easily that the linearization cone of the feasible region of CP_μ can be described as follows.

DEFINITION 3.3 (linearization cone). *Let* $\widetilde{\mathcal{F}}$ *denote the feasible region of the problem* CP_μ. *The linearization cone of* $\widetilde{\mathcal{F}}$ *at* $(\bar{x},\bar{y},\bar{u})$ *is the cone defined by*

$$\mathcal{L}((\bar{x},\bar{y},\bar{u});\widetilde{\mathcal{F}}) := \left\{ \begin{array}{l} (d,v) \\ \in R^{n+m} \times R^p \end{array} : \begin{array}{ll} \nabla(\nabla_y f + \bar{u}\nabla_y g)(\bar{x},\bar{y})d + \nabla_y g(\bar{x},\bar{y})^T v = 0, & \\ \nabla G_i(\bar{x},\bar{y})^T d \leq 0, & i \in I_G, \\ \nabla g_i(\bar{x},\bar{y})^T d = 0, & i \in I_g, \\ v_i = 0, & i \in I_u, \\ \nabla g_i(\bar{x},\bar{y})^T d \leq 0, v_i \geq 0, & i \in I_0 \end{array} \right\}.$$

Since the feasible region of an MPEC may be nonconvex, it is unreasonable to expect that the usual linearization cone of the feasible region $\widetilde{\mathcal{F}}$ is equal to the tangent cone of the feasible region. However, in the MPEC literature, it is known that under weak assumptions the MPEC linearization cone defined as follows is equal to the tangent cone of the feasible region. When the tangent cone is equal to the MPEC linearization cone it is said that the MPEC Abadie constraint qualification holds. The reader is referred to Ye [26] for sufficient conditions for the MPEC Abadie constraint qualification to hold.

DEFINITION 3.4 (MPEC linearization cone). *The MPEC linearization cone of* $\widetilde{\mathcal{F}}$

at $(\bar{x}, \bar{y}, \bar{u})$ *is the cone defined by*

$$\mathcal{L}^{MPEC}((\bar{x},\bar{y},\bar{u});\widetilde{\mathcal{F}}) := \left\{ (d,v) \in R^{n+m} \times R^p \ : \ \begin{array}{ll} \nabla(\nabla_y f + \bar{u}\nabla_y g)(\bar{x},\bar{y})h + \nabla_y g(\bar{x},\bar{y})^T v = 0, & \\ \nabla G_i(\bar{x},\bar{y})^T d \leq 0, & i \in I_G, \\ \nabla g_i(\bar{x},\bar{y})^T d = 0, & i \in I_g, \\ v_i = 0, & i \in I_u, \\ \left\{ \begin{array}{l} \nabla g_i(\bar{x},\bar{y})^T d \cdot v_i = 0, \\ \nabla g_i(\bar{x},\bar{y})^T d \leq 0,\ v_i \geq 0, \end{array} \right. & i \in I_0 \end{array} \right\}.$$

THEOREM 3.1. *Let* $(\bar{x}, \bar{y}, \bar{u})$ *be a local solution to CP. Suppose that the value function* $V(x)$ *is Lipschitz continuous near* $\bar{x}$*.*

If

$$F_\mu^\circ((\bar{x},\bar{y});d) \geq 0 \quad \forall (d,v) \in \mathcal{L}^{MPEC}((\bar{x},\bar{y},\bar{u});\widetilde{\mathcal{F}}) \tag{3.8}$$

for some $\mu \geq 0$*, then* $(\bar{x}, \bar{y}, \bar{u})$ *is a P-stationary point of CP.*

If

$$F_\mu^\circ((\bar{x},\bar{y});d) \geq 0 \quad \forall (d,v) \in \mathcal{L}((\bar{x},\bar{y},\bar{u});\widetilde{\mathcal{F}}) \tag{3.9}$$

holds for some $\mu \geq 0$*, then* $(\bar{x}, \bar{y}, \bar{u})$ *is an S-stationary point of CP.*

Proof. By (3.8), $(d, v) = (0, v)$ is an optimal solution to the following linearized problem:

$$\begin{array}{rl} \min\limits_{(d,v)} & \Phi(d) := F_\mu^\circ((\bar{x},\bar{y});d) \\ \text{s.t.} & \nabla(\nabla_y f + \bar{u}\nabla_y g)(\bar{x},\bar{y})d + \nabla_y g(\bar{x},\bar{y})^T v = 0, \\ & \nabla G_i(\bar{x},\bar{y})d \leq 0, \quad i \in I_G, \\ & \nabla g_i(\bar{x},\bar{y})d = 0, \quad i \in I_g, \\ & \qquad v_i = 0, \quad i \in I_u, \\ & \left\{ \begin{array}{l} \nabla g_i(\bar{x},\bar{y})d \cdot v_i = 0, \\ \nabla g_i(\bar{x},\bar{y})d \leq 0, v_i \geq 0, \end{array} \right. \quad i \in I_0. \end{array} \tag{3.10}$$

The objective function of the above problem is a nonsmooth convex function, and the constraint functions are all linear in variable (d, v). Hence the MPEC linear CQ holds. Applying Theorem 1.1, we conclude that for each partition P, Q of I_0 there exist multipliers $\beta \in R^m$, $\eta^g \in R^p$, $\eta^G \in R^q$, $\eta^u \in R^p$ such that

$$\begin{array}{l} 0 \in \partial\Phi(0) + \nabla(\nabla_y f + \bar{u}\nabla_y g)(\bar{x},\bar{y})^T \beta + \nabla g(\bar{x},\bar{y})^T \eta^g + \nabla G(\bar{x},\bar{y})^T \eta^G, \\ 0 = \nabla_y g(\bar{x},\bar{y})\beta - \eta^u, \\ \eta_i^G \geq 0, \quad i \in I_G, \quad \eta_i^G = 0, \quad i \notin I_G, \\ \eta_i^g = 0, \quad i \in I_u, \quad \eta_i^u = 0, \quad i \in I_g, \\ \eta_i^g \geq 0, \quad i \in P, \quad \eta_i^u \geq 0, \quad i \in Q. \end{array}$$

By the calculus rules for Clarke generalized gradients, one has

$$\partial\Phi(0) \subseteq \nabla F(\bar{x},\bar{y}) + \mu\{\nabla f(\bar{x},\bar{y}) - \partial V(\bar{x}) \times \{0\}\}.$$

The conclusion that $(\bar{x}, \bar{y}, \bar{u})$ is a P-stationary point of CP follows from replacing η^u by $\nabla_y g(\bar{x}, \bar{y})\beta$.

Now suppose that (3.9) holds. Then $(d, v) = (0, v)$ is an optimal solution to the following linearized problem:

$$
\begin{aligned}
\min_{(d,v)} \quad & \Phi(d) := F_\mu^\circ((\bar{x}, \bar{y}); d) \\
\text{s.t.} \quad & \nabla(\nabla_y f + \bar{u}\nabla_y g)(\bar{x}, \bar{y})d + \nabla_y g(\bar{x}, \bar{y})^T v = 0, \\
& \nabla G_i(\bar{x}, \bar{y})d \le 0, \quad i \in I_G, \\
& \nabla g_i(\bar{x}, \bar{y})d = 0, \quad i \in I_g, \\
& \qquad\quad v_i = 0, \quad i \in I_u, \\
& \nabla g_i(\bar{x}, \bar{y})d \le 0,\ v_i \ge 0, \quad i \in I_0.
\end{aligned}
$$

The conclusion that $(\bar{x}, \bar{y}, \bar{u})$ is an S-stationary point of CP follows from applying the nonsmooth KKT condition to the above optimization problem. □

The necessary optimality conditions obtained in Theorem 3.1 involve the Clarke generalized directional derivative and the Clarke generalized gradient of the value function. In some practical circumstance, calculating the Clarke generalized gradients may be difficult or impossible. We now introduce two new conditions under which our new necessary optimality conditions hold. Our new conditions do not involve either the Clarke generalized directional derivative or the Clarke generalized gradient of the value function.

DEFINITION 3.5. *Take as given a feasible vector $(\bar{x}, \bar{y}, \bar{u})$ of (CP_μ). We say that CP is MPEC-weakly calm at $(\bar{x}, \bar{y}, \bar{u})$ with modulus μ if*

$$
(3.11)\quad [\nabla F(\bar{x}, \bar{y}) + \mu\nabla f(\bar{x}, \bar{y})]^T d - \mu \min_{\xi\in W(\bar{x})} \xi dx \ge 0 \quad \forall (d, v) \in \mathcal{L}^{MPEC}((\bar{x}, \bar{y}, \bar{u}); \widetilde{\mathcal{F}}).
$$

We say that CP is weakly calm at $(\bar{x}, \bar{y}, \bar{u})$ with modulus μ if

$$
(3.12)\quad [\nabla F(\bar{x}, \bar{y}) + \mu\nabla f(\bar{x}, \bar{y})]^T d - \mu \min_{\xi\in W(\bar{x})} \xi dx \ge 0 \qquad \forall (d, v) \in \mathcal{L}((\bar{x}, \bar{y}, \bar{u}); \widetilde{\mathcal{F}}).
$$

Since

$$
F_\mu^\circ((\bar{x}, \bar{y}); d) = [\nabla F(\bar{x}, \bar{y}) + \mu\nabla f(\bar{x}, \bar{y})]^T d + \mu(-V)^\circ(\bar{x}; dx)
$$

and

$$
(-V)^\circ(\bar{x}; dx) = \max_{\xi\in\partial(-V)(\bar{x})}\{\xi dx\} \le \max_{\xi\in -W(\bar{x})}\{\xi dx\} = \max_{\xi\in W(\bar{x})}\{-\xi dx\} = -\min_{\xi\in W(\bar{x})}\{\xi dx\},
$$

the MPEC-weakly calmness condition and the weakly calmness condition are weaker than conditions (3.8) and (3.9), respectively. In fact for the Mirrlees problem, section 2 has illustrated that the weakly calmness condition does not hold.

THEOREM 3.2. *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local solution to CP. Suppose that the set $W(\bar{x})$ as defined in* (1.4) *is nonempty and compact.*

If CP is MPEC-weakly calm at $(\bar{x}, \bar{y}, \bar{u})$ with modulus $\mu \ge 0$, then for each partition of the index set I_0 into P, Q there exist $\lambda^i \ge 0$, $\sum_{i=1}^{n+1} \lambda^i = 1$, $y^i \in S(\bar{x})$, $u^i \in$

$M^1(\bar{x}, y^i)$, $i = 1, 2, \ldots, n+1$, and $\beta \in R^m$, $\eta^g \in R^p$, $\eta^G \in R^q$ such that

$$0 = \nabla_x F(\bar{x}, \bar{y}) + \mu \sum_{i=1}^{n+1} \lambda^i (\nabla_x f(\bar{x}, \bar{y}) - \nabla_x f(\bar{x}, y^i) - u^i \nabla_x g(\bar{x}, y^i))$$

$$+\nabla_x(\nabla_y f + \bar{u}\nabla_y g)(\bar{x}, \bar{y})^T \beta + \nabla_x g(\bar{x}, \bar{y})^T \eta^g + \nabla_x G(\bar{x}, \bar{y})^T \eta^G, \tag{3.13}$$

$$0 = \nabla_y F(\bar{x}, \bar{y}) + \mu \nabla_y f(\bar{x}, \bar{y})$$

$$+\nabla_y(\nabla_y f + \bar{u}\nabla_y g)(\bar{x}, \bar{y})^T \beta + \nabla_y g(\bar{x}, \bar{y})^T \eta^g + \nabla_y G(\bar{x}, \bar{y})^T \eta^G, \tag{3.14}$$

$$\eta_i^G \geq 0, \quad i \in I_G, \qquad \eta_i^G = 0, \quad i \notin I_G, \tag{3.15}$$

$$\eta_i^g = 0, \quad i \in I_u, \quad (\nabla_y g(\bar{x}, \bar{y})\beta)_i = 0, \quad i \in I_g, \tag{3.16}$$

$$\eta_i^g \geq 0, \quad i \in P, \quad (\nabla_y g(\bar{x}, \bar{y})\beta)_i \geq 0, \quad i \in Q.$$

If CP is weakly calm with modulus $\mu \geq 0$ at $(\bar{x}, \bar{y}, \bar{u})$, then there exist $\lambda^i \geq 0$, $\sum_{i=1}^{n+1} \lambda^i = 1$, $y^i \in S(\bar{x})$, $u^i \in M^1(\bar{x}, y^i)$, $i = 1, 2, \ldots, n+1$, and $\beta \in R^m$, $\eta^g \in R^p$, $\eta^G \in R^q$ such that (3.13)–(3.16) *holds and*

$$\eta_i^g \geq 0, \quad (\nabla_y g(\bar{x}, \bar{y})\beta)_i \geq 0, \quad i \in I_0.$$

Proof. Suppose that CP is MPEC-weakly calm. Then (3.11) holds at $(\bar{x}, \bar{y}, \bar{u})$ for some $\mu \geq 0$. Therefore $(d, v) = (0, v)$ is an optimal solution to the following linearized problem:

$$\begin{aligned}
\min_{(h,v)} \quad & [\nabla F(\bar{x}, \bar{y}) + \mu \nabla f(\bar{x}, \bar{y})]^T d - \mu \min_{\xi \in W(\bar{x})} \xi^T dx \\
\text{s.t.} \quad & \nabla(\nabla_y f + \bar{u}\nabla_y g)(\bar{x}, \bar{y}) d + \nabla_y g(\bar{x}, \bar{y})^T v = 0, \\
& \nabla G_i(\bar{x}, \bar{y}) d \leq 0, \quad i \in I_G, \\
& \nabla g_i(\bar{x}, \bar{y}) d = 0, \quad i \in I_g, \\
& \qquad v_i = 0, \quad i \in I_u, \\
& \left\{ \begin{array}{l} \nabla g_i(\bar{x}, \bar{y}) d \cdot v_i = 0, \\ \nabla g_i(\bar{x}, \bar{y}) d \leq 0,\ v_i \geq 0, \end{array} \right. \qquad i \in I_0.
\end{aligned} \tag{3.17}$$

Let $\phi(z) := \min_{\xi \in W(\bar{x})} \xi^T z$. Since the set $W(\bar{x})$ is assumed to be nonempty and compact, by Proposition 2.1 one has $\partial\phi(0) = coW(\bar{x})$. Therefore the Clarke generalized gradient of the objective function of the above problem at $(d, v) = (0, v)$ is included in the set described by

$$\nabla F(\bar{x}, \bar{y}) + \mu[\nabla f(\bar{x}, \bar{y}) - coW(\bar{x}) \times \{0\}].$$

By Carathéodory's theorem, the convex set $coW(\bar{x}) \subseteq R^n$ can be represented by not more than $n+1$ elements at a time. Therefore

$$coW(\bar{x}) = \left\{ \begin{array}{ll} \sum_{i=1}^{n+1} \lambda^i (\nabla_x f(\bar{x}, y^i) + u^i \nabla_x g(\bar{x}, y^i)) : & y^i \in S(\bar{x}),\ u^i \in M^1(\bar{x}, y^i), \\ & \lambda^i \geq 0,\ \sum_{i=1}^{n+1} \lambda^i = 1. \end{array} \right\}$$

As in the proof of Theorem 3.1, the desired result follows from applying Theorem 1.1, and we omit the proof. ∎

Remark 3.1. (i) A sufficient but not necessary condition for the set $W(\bar{x})$ to be nonempty and compact is that the MFCQ hold at every optimal solution of the lower level problem $P_{\bar{x}}$ and that the set-valued map $Y(x) := \{y \in R^m : g(x, y) \leq 0\}$ be uniformly bounded around $\bar{x}$.

(ii) The new S- or P-type optimality conditions obtained in Theorem 3.2 are in general weaker than the S- or P-stationary conditions, respectively. However, they are the most suitable surrogates for the S- or P-stationary conditions, since by the sensitivity of the value function

$$\partial V(\bar{x}) \subseteq coW(\bar{x})$$

and the equality holds under certain conditions.

4. Illustrative examples. In this section we use some nonconvex bilevel programming problems to illustrate the necessary optimality conditions we obtain in Theorem 3.2. We first examine the example given by Vogel in his Ph.D. thesis and later included by Dempe in his recent monograph.

Example 4.1 (see Vogel [21], Dempe [10]).

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

It is easy to see that

$$S(x) = \begin{cases} \{x\} & \text{if } -3 \leq x < -2, \\ \{-2, 1\} & \text{if } x = -2, \\ \{1\} & \text{if } -2 < x \leq 1, \\ \{x\} & \text{if } 1 < x \leq 2, \end{cases}$$

and the optimal solution of the bilevel problem is $\bar{x} = \bar{y} = -2$.

If the classical KKT approach is used, then the resulting one-level problem is

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

However, the optimal solution of the above problem is $\bar{x} \leq -1,\ \bar{y} = -1,\ \bar{u} = 0$, and the optimal solution of the bilevel program and the corresponding multiplier $\bar{x} = \bar{y} = -2,\ \bar{u} = 9$ do not satisfy the KKT condition of the above problem. Hence the classical KKT approach fails for this example. Now

$$S(\bar{x}) = \{-2, 1\}, \quad M^1(\bar{x}, -2) = \{9\}, \quad M^1(\bar{x}, 1) = \{0\}, \quad W(\bar{x}) = \{0, 9\}$$

and for any $d = (dx, dy) \in R^2$,

$$\begin{aligned} &\nabla F(\bar{x},\bar{y})^T d + \mu \nabla f(\bar{x},\bar{y})^T d - \mu \min_{\xi \in W(\bar{x})} \xi dx \\ &= \begin{cases} (-2+9\mu)dy - 9\mu dx & \text{if } dx < 0 \\ (-2+9\mu)dy & \text{if } dx \geq 0 \end{cases} \\ &\geq 0 \quad \text{for } \mu = 2/9. \end{aligned}$$

Hence the problem is weakly calm with modulus $\mu = 2/9$. The KKT condition from Theorem 3.2 implies the existence of $\lambda^1 \geq 0,\ \lambda^2 \geq 0,\ \lambda^1 + \lambda^2 = 1,\ y^1 = \bar{y} = -2,\ y^2 =$

$1,\ u^1 = \bar{u} = 9,\ u^2 = 0,\ \beta \in R,\ \eta^g \in R$ such that

$$\begin{aligned}
&0 = \mu\lambda^1(-\bar{u}) + \mu\lambda^2(-u^2) + \eta^g,\\
&0 = 2(\bar{y}+1) + \mu(3\bar{y}^2 - 3) + 6\bar{y}\beta + (-1)\eta^g,\\
&\beta = 0.
\end{aligned}$$

The above KKT condition holds at $(\bar{x}, \bar{y}, \bar{u}) = (-2, -2, 9)$ with $\lambda^1 = 0,\ \lambda^2 = 1,\ \beta = 0,\ \eta^g = 0,\ \mu = 2/9$. Since μ is not equal to zero but β is equal to zero, the value function approach works but the classical KKT approach fails. This example shows that our theorem includes the KKT condition derived by using the value function approach as a special case for the nonconvex bilevel programming problem.

We now modify the objective function of the upper level problem so that it becomes the following problem.

Example 4.2.

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

It is easy to see that

$$S(x) = \begin{cases} \{x\} & \text{if } -3 < x < -2,\\ \{-2, 1\} & \text{if } x = -2,\\ \{1\} & \text{if } -2 < x \leq 1,\\ \{x\} & \text{if } 1 < x \leq 2, \end{cases}$$

and the optimal solution of the bilevel problem is $\bar{x} = \bar{y} = 1$. Now we have

$$S(\bar{x}) = \{1\}, \qquad M^1(\bar{x}, 1) = \{0\},$$

and since $I^g = I^u = \emptyset$ and $I^0 \neq \emptyset$, the linearization cone of the feasible region is

$$\mathcal{L}((\bar{x}, \bar{y}, \bar{u}); \widetilde{\mathcal{F}}) = \{(dx, dy, du) : 6dy = dv, dy \geq dx, dv \geq 0\}.$$

Since, for all $\mu \geq 0$,

$$\nabla F(\bar{x}, \bar{y})^T d + \mu \nabla f(\bar{x}, \bar{y})^T d = 2dy \geq 0 \qquad \forall (dx, dy, dv) \in \mathcal{L}((\bar{x}, \bar{y}, \bar{u}); \widetilde{\mathcal{F}}),$$

the problem is weakly calm at $(\bar{x}, \bar{y}, \bar{u})$ with modulus $\mu \geq 0$. The KKT condition

$$\begin{aligned}
&0 = 2(\bar{x} - 1) + \eta^g,\\
&0 = 2\bar{y} + \mu(3\bar{y}^2 - 3) + 6\bar{y}\beta + (-1)\eta^g,\\
&\eta^g \geq 0, \quad \beta \leq 0,
\end{aligned}$$

holds at $(\bar{x}, \bar{y}, \bar{u}) = (1, 1, 0)$ with $\beta = -1/3,\ \eta^g = 0,\ \mu \geq 0$.

Since the solution set $S(\bar{x})$ is a singleton and the corresponding multiplier is a singleton as well for this example, our KKT condition actually reduces to the KKT condition derived by using the classical KKT approach. But the problem in this example is a nonconvex bilevel programming problem. Hence this example illustrates that our result provides a sufficient condition under which the KKT condition derived by the classical KKT approach is valid even for a nonconvex bilevel programming problem.

Example 4.3.

$$\begin{aligned}
&\min\ (x-0.5)^2+(y-2)^2\\
&\text{s.t. } y\in S(x):=\arg\min_y\{y^3-3y : y\ge x-3\},\\
&\quad 0\le x\le 4.
\end{aligned}$$

The solution set of the lower level problem can be easily obtained as

$$S(x)=\begin{cases}\{x-3\} & \text{if } 0\le x<1,\\ \{-2,1\} & \text{if } x=1,\\ \{1\} & \text{if } 1<x\le 4,\end{cases}$$

and it is easy to verify that the optimal solution of the bilevel problem is $\bar{x}=\bar{y}=1$. Now

$$S(\bar{x})=\{-2,1\},\quad M^1(\bar{x},-2)=\{9\},\quad M^1(\bar{x},1)=\{0\},\quad W(\bar{x})=\{0,9\}.$$

Since $g(\bar{x},\bar{y})=1-3-1=-3<0$ and $\bar{u}=0$, $I^g=I^0=\emptyset$ and $I^u\neq\emptyset$. The linearization cone of the feasible region is

$$\mathcal{L}((\bar{x},\bar{y},\bar{u});\widetilde{\mathcal{F}})=\{(dx,dy,du):6dy=0\}.$$

So one has

$$\begin{aligned}
&\nabla F(\bar{x},\bar{y})^T d+\mu\nabla f(\bar{x},\bar{y})^T d-\mu\min_{\xi\in W(\bar{x})}\xi dx\\
&=\begin{cases} dx-2dy-9\mu dx & \text{if } dx<0\\ dx-2dy & \text{if } dx\ge 0\end{cases}\\
&\ge 0\quad \forall (dx,dy,dv)\in\mathcal{L}((\bar{x},\bar{y},\bar{u});\widetilde{\mathcal{F}}) \text{ and } \mu=1/9,
\end{aligned}$$

and hence the problem is weakly calm with modulus $\mu=1/9$. The KKT condition

$$\begin{aligned}
&0=2(\bar{x}-0.5)+\mu\lambda^1(-\bar{u})+\mu\lambda^2(-u^2)+\eta^g,\\
&0=2(\bar{y}-2)+\mu(3\bar{y}^2-3)+6\bar{y}\beta+(-1)\eta^g,\\
&\eta^g=0,
\end{aligned}$$

holds at $(\bar{x},\bar{y},\bar{u})=(1,1,9)$ with $\lambda^1=1,\ \lambda^2=0,\ u^2=0,\ \beta=1/3,\ \eta^g=0,\ \mu=1/9$. Note that since neither β nor μ is equal to zero, both the classical KKT approach and the value function approach fail, and only the combined KKT and the value function approach will work for this example.

Acknowledgments. We would like to thank two anonymous referees for their valuable comments which helped us to improve the presentation of the material in this paper.

REFERENCES

[1] A. Araujo and H. Moreira, *A general Lagrangian approach for non-concave moral hazard problems*, J. Math. Econom., 35 (2001), pp. 17–39.

[2] J. F. Bard, *Practical Bilevel Optimization: Algorithms and Applications*, Kluwer Academic Publications, Dordrecht, The Netherlands, 1998.
[3] J. F. Bard and J. E. Falk, *An explicit solution to the multilevel programming problem*, Comput. Oper. Res., 9 (1982), pp. 77–100.
[4] J. Bracken and J. McGill, *Mathematical programs with optimization problems in the constraints*, Oper. Res., 21 (1973), pp. 37–44.
[5] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983.
[6] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer, New York, 1998.
[7] J. M. Danskin, *The Theory of Max-Min and Its Applications to Weapons Allocation Problems*, Springer, New York, 1967.
[8] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.
[9] S. Dempe, *Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints*, Optimization, 52 (2003), pp. 333–359.
[10] S. Dempe, *Bilevel programming*, in Essays and Surveys in Global Optimization, GERAD 25 Anniv. Ser. 7, Springer, New York, 2005, pp. 165–193.
[11] S. Dempe, J. Dutta, and B. S. Mordukhovich, *New necessary optimality conditions in optimistic bilevel programming*, Optimization, 56 (2007), pp. 577–604.
[12] Z. Q. Luo, J. S. Pang, and D. Ralph, *Mathematical Programs with Equilibrium Constraints,* Cambridge University Press, Cambridge, UK, 1996.
[13] J. Mirrlees, *The theory of moral hazard and unobservable behaviour—Part* I, Rev. Econom. Stud., 66 (1999), pp. 3–22.
[14] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation (Vol.* 1: *Basic Theory; Vol.* 2: *Applications)*, Springer, New York, 2004.
[15] B. S. Mordukhovich, N. M. Nam, and N. D. Yen, *Subgradients of marginal functions in parametric mathematical programming*, Math. Program. Ser. B, 116 (2009), pp. 369–396.
[16] J. V. Outrata, M. Kočvara, and J. Zowe, *Nonsmooth Approach to Optimization Problem with Equilibrium Constraints: Theory, Application and Numerical Results*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
[17] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1–22.
[18] K. Shimizu, Y. Ishizuka, and J. F. Bard, *Nondifferentiable and Two-Level Mathematical Programming*, Kluwer Academic Publishers, Boston, 1997.
[19] H. von Stackelberg, *The Theory of the Market Economy*, Oxford University Press, Oxford, England, 1954 (English translation of Marktform und Gleichgewicht, Springer-Verlag, Berlin, 1934).
[20] L. N. Vincente and P. H. Calamai, *Bilevel and multilevel programming: A bibliography review*, J. Global Optim., 5 (1994), pp. 291–306.
[21] S. Vogel, *Zwei-Ebenen-Optimierungsaufgaben mit nichtkonvexer Zielfunktion in der unterren Ebene: Pfadverfolgung und Sprünge*, Ph.D. thesis, Technische Universität Bergakademie Freiberg, Freiberg, Germany, 2002.
[22] J. J. Ye, *New uniform parametric error bounds*, J. Optim. Appl., 98 (1998), pp. 197–219.
[23] J. J. Ye, *Optimality conditions for optimization problems with complementarity constraints*, SIAM J. Optim., 9 (1999), pp. 374–387.
[24] J. J. Ye, *Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints*, SIAM J. Optim., 10 (2000), pp. 943–962.
[25] J. J. Ye, *Nondifferentiable multiplier rules for optimization and bilevel optimization problems*, SIAM J. Optim., 15 (2004), pp. 252–274.
[26] J. J. Ye, *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints*, J. Math. Anal. Appl., 307 (2005), pp. 350–369.
[27] J. J. Ye, *Constraint qualifications and KKT conditions for bilevel programming problems*, Math. Oper. Res., 31 (2006), pp. 811–824.
[28] J. J. Ye and D. L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization, 33 (1995), pp. 9–27.
[29] J. J. Ye and D. L. Zhu, *A note on optimality conditions for bilevel programming problems*, Optimization, 39 (1997), pp. 361–366.
[30] J. J. Ye, D. L. Zhu, and Q. J. Zhu, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM J. Optim., 7 (1997), pp. 481–507.