Optimization

ISSN: 0233-1934 (Print) 1029-4945 (Online) Journal homepage: https://www.tandfonline.com/loi/gopt20

Optimality conditions for bilevel programming problems

J. J. Ye & D. L. Zhu

Optimization, 1995, Vol. 33, pp. 9–27
Reprints available directly from the publisher
Photocopying permitted by license only

OPTIMALITY CONDITIONS FOR BILEVEL PROGRAMMING PROBLEMS

J. J. YE* and D. L. ZHU†

**Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada V8W 3P4*

†GERAD, École des Hautes Études Commerciales, 5255 Avenur decelles, Montréal, Québec, Canada H3T 1V6

(Received 28 February 1994; in final form 5 July 1994)

The bilevel programming problem (BLPP) 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. To obtain optimality conditions, we reformulate BLPP as a single level mathematical programming problem (SLPP) which involves the value function of the lower level problem. For this mathematical programming problem, it is shown that in general the usual constraint qualifications do not hold and the right constraint qualification is the calmness condition. It is also shown that the linear bilevel programming problem and the minmax problem satisfy the calmness condition automatically. A sufficient condition for the calmness for the bilevel programming problem with quadratic lower level problem and nondegenerate linear complementarity lower level problem are given. First order necessary optimality condition are given using nonsmooth analysis. Second order sufficient optimality conditions are also given for the case where the lower level problem is unconstrained.

KEY WORDS: Optimality conditions, bilevel programming problems, nonsmooth analysis, constraint qualification.

Mathematics Subject Classification 1991:
Primary: 65K05; Secondary: 49J52.

1. INTRODUCTION

Let us consider a two-level hierarchical system where the higher level (hereafter the "*leader*") and the lower level (hereafter the "*follower*") must find vectors $x \in R^n$ and $y \in R^m$, respectively to minimize their individual objective functions $F(x, y)$ and $f(x, y)$ subject to certain constraints. The leader is assumed to select his decision vector first and the follower after that. Under these assumptions on the order of the play, the game will proceed as follows: For any possible decision vector $x \in X \subset R^n$ chosen by the leader, the follower will react optimally by choosing his decision vector $y \in Y \subset R^m$ to minimize the objective function $f(x, y)$ subject to constraints $h(x, y) = 0$, $g(x, y) \leqslant 0$. Assume that if the *solution set* $S(x)$ of the follower's problem is not a singleton, the follower allows the leader to choose which of them is actually used. Hence the leader now chooses his optimal decision vection $x \in X$ and $y \in S(x)$ to minimize his objective $F(x, y)$. In mathematical terms, given any decision vector $x \in X$ chosen by the leader, the follower faces the ordinary (*single level*) mathematical programming problem parametered in x:

$$\begin{aligned}(P_x)\quad \min\quad & f(x,y)\\ \text{s.t.}\quad & h(x,y)=0\\ & g(x,y)\leqslant 0\\ & y\in Y,\end{aligned}$$

while the leader faces the *bilevel programming problem*:

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

where $F, f: R^{n+m}\to R$, $h: R^{n+m}\to R^c$, $g: R^{n+m}\to R^d$ are continuously differentiable, X and Y are closed subset of R^n and R^m respectively. We allow c or $d=0$ to signify the case in which there are no explicite equality or inequality constraints. In these cases it is clear below that certain references to such constraints are simply to be deleted.

The bilevel programming problem described above was first introduced in an economic model by Von Stackelberg [16]. There is a considering amount of works dealing with the existence, properties, approximations of solutions (e.g. [16], [10], [18]) and numerical methods (e.g. [1], [17], [12]). There are also some papers dealing with optimality conditions (e.g. [1], [19], [13], [6]). The classical approach to derive necessary optimality conditions for bilevel programming problem (see e.g. [1]) was to replace the solution set $S(x)$ of the lower level problem with the set of vectors $y\in R^m$ which are feasible for (P_x) and at which the Lagrange multiplier rules are satisfied, and to minimize over the original variables and the Lagrange multipliers. This approach requires the convexity of the lower level problem and the problem of constraint qualifications which ensure the necessary conditions are the Kuhn-Tucker type are usually neglected. Zhang [19] extends the classical approach to allow the nonsmooth problem data. Dempe [6] and Outrata [13] derive necessary conditions for the case where the solution set of the lower level problem $S(x)=\{y(x)\}$ is a singleton by minimizing the objective function $F(x, y(x))$ over all $x\in X$. This approach, however, requires that the solution set $S(x)$ is a singleton and the vector-valued function $y(x)$ has certain differentiability properties. The purpose of this paper is to derive optimality conditions for general bilevel programming problems without convexity assumption on the lower level problem and without the assumption that the solution set of the lower level problem $S(x)$ is a singleton. In particular, we will address the problem of constraint qualification.

We now describe our approach. Define the *value function of the lower level programming problem* as an extended value function $V: X\to \bar{R}$ by

$$V(x) := \inf_{y\in R^m}\{f(x,y): h(x,y)=0, g(x,y)\leqslant 0, y\in Y\}$$

where $\bar{R} := R\cup\{-\infty\}\cup\{+\infty\}$ is the extended real line and $\inf\{\varnothing\}=+\infty$ by convention. Then BLPP can be reformulated as the following single level mathematical programming problem:

$$\begin{aligned} \text{SLPP} \quad \min \quad & F(x,y) \\ \text{s.t.} \quad & f(x,y)-V(x)=0 \\ & h(x,y)=0 \\ & g(x,y)\leqslant 0 \\ & x\in X, y\in Y. \end{aligned}$$

It is known that $V(x)$ is not smooth in general even in the case where all problem data $F(x,y)$, $f(x,y)$, $h(x,y)$, $g(x,y)$ are continuously differentiable. In §2, we identify conditions under which the value function $V(x)$ is Lipschitz continuous, estimate the Clarke generalized gradient of $V(x)$ and apply the generalized Lagrange multiplier rule to derive a necessary optimality condition of Fritz-John type for BLPP. The difficulty with the derivation of a necessary condition of Kuhn–Tucker type is with the constraint qualification. It is tempting to develop a necessary condition of Kuhn–Tucker type by using the single level formulation SLPP by assuming a usual constraint qualification commonly used in mathematical programming problem such as Mangasarian-Fromovitz condition (which are sometimes referred to as Cottle condition [4]) holds for SLPP. In §3, we show that the usual constraint qualifications such as linear independence condition, Slater condition and Mangasarian-Fromovitz condition in general do not hold for SLPP due to the special structure of the constraint $f(x,y)-V(x)=0$ and that the (partial) calmness condition is the right constraint qualification condition. In §4, we show that the bilevel programming problem with linear lower level problem and the minmax problem satisfy the calmness condition. In §5, we identify the uniformly weak sharp minimum as a sufficient condition for calmness and give a sufficient condition for calmness for the bilevel programming problem where the lower level problem is quadratic. An example is also given to show that the uniformly weak sharp minimum is sufficient but not necessary for calmness. Sufficient optimality conditions for the unconstrained bilevel programming problem are given in §6.

2. NECESSARY CONDITIONS OF FRITZ-JOHN TYPE

In this section, we study the (generalized) differentiability of the value function for the lower level problem and derive the necessary condition of Fritz-John type for BLPP.

Let y be feasible for (P_x). The *normal* and *abnormal* multiplier sets for problem (P_x) corresponding to y are the sets defined respectively by

$$\begin{aligned} M_x^1(y) := \{(v,\pi)\in R^{c+d}: \quad & 0\in\nabla_y f(x,y)+\nabla_y h(x,y)^{\mathrm{T}}v+\nabla_y g(x,y)^{\mathrm{T}}\pi+N(y,Y), \\ & \pi\geqslant 0, \langle \pi, g(x,y)\rangle = 0\} \\ M_x^0(y) := \{(v,\pi)\in R^{c+d}: \quad & 0\in\nabla_y h(x,y)^{\mathrm{T}}v+\nabla_y g(x,y)^{\mathrm{T}}\pi+N(y,Y), \\ & \pi\geqslant 0, \langle \pi, g(x,y)\rangle = 0\} \end{aligned}$$

where $^{\mathsf{T}}$ denotes and the transpose and $N(y, Y)$ denotes the Clarke normal cone to Y at $y \in Y$.

Denote by $M_x^0 S(x) := \bigcup_{y \in S(x)} M_x^0(y)$. The following proposition gives a sufficient condition for Lipschitz continuity of the value function $V(x)$ and an estimate for the Clarke generalized gradient of V.

Proposition 2.1: *Suppose the solution set of the lower level problem $S(x)$ is nonempty and every $y \in S(x)$ lies in the interior of Y. If $M_x^0 S(x) = \{0\}$, then the value function $V(x)$ is Lipschitz continuous near x and one has the following estimates for the Clarke generalized gradient of $V(x)$:*

$$\partial V(x) \subset co\{\nabla_x f(x, y) + \nabla_x h(x, y)^t v + \nabla_x g(x, y)^t \pi : y \in S(x), (v, \pi) \in M_x^1(y)\}. \tag{1}$$

Proof: Using an observation due to Rockafellar, we can rewrite the value function $V(x)$ as follows:

$$V(x) = \min_{(\alpha, y) \in R^n \times R^m} \{f(\alpha, y) : h(\alpha, y) = 0, g(\alpha, y) \leqslant 0, \alpha - x = 0, y \in Y\}.$$

Application of methods and results in Corollary 1 of Theorem 6.5.2 of Clarke [5] to the above problem leads to the desired conclusion. ■

The following Fritz-John type Lagrange multiplier rule can be obtained readily from applying the nonsmooth Lagrange multiplier rule (Theorem 6.1.1 of Clarke [5]) to SLPP and using the estimates for $\partial V(x)$ (Proposition 2.1).

Theorem 2.1: [Fritz-John type Lagrange multiplier rule]. *Let (x^*, y^*) solve BLPP. Suppose that x^* lies in the interior of X and every $y \in S(x^*)$ lies in the interior of Y. Suppose that $M_{x^*}^0 S(x^*) = \{0\}$. Then there exists $\lambda \in \{0, 1\}$, $\mu \geqslant 0$, $s \in R^c$, $r \in R^d$ not all zero, positive integers I, J, $\lambda_{ij} \geqslant 0$, $\sum_{i=1}^I \sum_{j=1}^J \lambda_{ij} = 1$, $y_i \in S(x^*)$, $v_{ij} \in R^c$, $\pi_{ij} \in R^d$ such that*

$$\begin{aligned}
0 &= \lambda \nabla_x F(x^*, y^*) + \nabla_x h(x^*, y^*)^t s + \nabla_x g(x^*, y^*)^t r + \mu[\nabla_x f(x^*, y^*) \\
&\quad - \sum_{ij} \lambda_{ij}(\nabla_x f(x^*, y_i) + \nabla_x h(x^*, y_i)^t v_{ij} + \nabla_x g(x^*, y_i)^t \pi_{ij})] \\
0 &= \lambda \nabla_y F(x^*, y^*) + \nabla_y h(x^*, y^*)^t s + \nabla_y g(x^*, y^*)^t r + \mu \nabla_y f(x^*, y^*) \\
r &\geqslant 0, \langle r, g(x^*, y^*) \rangle = 0 \\
0 &= \nabla_y f(x^*, y_i) + \nabla_y h(x^*, y_i)^t v_{ij} + \nabla_y g(x^*, y_i)^t \pi_{ij} \\
\pi_{ij} &\geqslant 0, \langle \pi_{ij}, g(x^*, y_i) \rangle = 0.
\end{aligned}$$

Remark 2.1: $\mu \geqslant 0$ results from the fact that the feasible points of SLPP are the same as those satisfying the constraints with $f(x, y) - V(x) \leqslant 0$ instead of $f(x, y) - V(x) = 0$. The case where x^* is in the boundary of X and there are some $y \in S(x^*)$ lies in the boundary Y can be easily taken care by adding $N(x^*, X)$ and $N(y, Y)$ to the appropriate places and changing the corresponding equalities to inclusions. For a clear presentation, we will consider the case where x^* is in the interior of X and that every $y \in S(x^*)$ lies in the interior of Y.

Remark 2.2: Since the assumption $M_{x^*}^0 S(x^*) = \{0\}$ is only for the Lipschitz continuity of the value function $V(x)$ near x^*, it can be replaced by any condition which ensures the Lipschitz continuity of the value function.

3. NECESSARY CONDITIONS OF KUHN-TUCKER TYPE

The necessary conditions of Theorem 2.1 may not provide useful information regarding to optimality when the multiplier corresponding to F (which we have labeled λ) vanishes, since then the function F being minimized is not involved. Various constraint qualifications have been proposed in the literature to ensure that the multiplier corresponding to the objective function in a mathematical programming problem does not vanish.

Let us first discuss the issue of constraint qualification for a general mathematical programming problem with both equality and inequality constraints:

$$\begin{aligned} (P)\quad \min\quad & f(z) \\ \text{s.t.}\quad & h(z)=0 \\ & g(z)\leqslant 0 \\ & z\in C, \end{aligned}$$

where $f: R^k \to R$, $h: R^k \to R^l$, $g: R^k \to R^d$, $C \subset R^k$. We assume that C is a closed subset of R^k, each function f, h, g is Lipschitz continuous near any given point of C and $h(z)=(h_1(z), h_2(z), \dots, h_l(z))$, $g(z)=(g_1(z), \dots, g_d(z))$. We allow l or $d=0$ to signify the case in which there are no explicite equality or inequality constraints. In these cases it is clear below that certain references to such constraints are simply to be deleted.

Let z be feasible for (P). The *normal* and *abnormal* multiplier sets for problem (P) corresponding to z are the sets defined respectively by

$$\begin{aligned} M^1(z) := \{(r,s)\in R^{l+d}:\ & 0\in\partial f(z)+\partial h(z)^{\mathrm{T}}s+\partial g(z)^{\mathrm{T}}r+N(z,C), \\ & r\geqslant 0, \langle r, g(z)\rangle = 0\} \\ M^0(z) := \{(r,s)\in R^{l+d}:\ & 0\in\partial h(z)^{\mathrm{T}}r+\partial g(z)^{\mathrm{T}}s+N(z,C), \\ & r\geqslant 0, \langle r, g(z)\rangle = 0\}, \end{aligned}$$

where $\partial\phi(z)$ denotes the Clarke generalized gradient or the generalized Jacobian of ϕ at z. Note that the Kuhn-Tucker condition holds at a solution z to problem (P) is equivalent to $M^1(z)\neq\varnothing$. Since the generalized Fritz-John Lagrange multiplier rule, Theorem 6.1.1. of Clarke [5], can be rephrased as: if z solves (P) then one has

$$M^1(z)\cup[M^0(z)\backslash\{0\}]\neq\varnothing, \tag{2}$$

all constraint qualification can be classified into two categories: the type that makes structural assumptions about the data of the problem so that the set $M^0(z)$ of abnormal multipliers necessarily reduces to 0 (which would imply $M^1(z)\neq\varnothing$ by virtue of (2)), and the type that simply assures that $M^1(x)$ is nonempty (even though $M^0(z)$ may not reduce to 0).

The most commonly used constraint qualifications for a general mathematical programming problem are the linear independence, Mangasarian-Fromovitz and Slater conditions. For problem (P), the nonsmooth extension of the classical linear independence and Mangasarian-Fromovitz conditions, first introduced by Hiriart-Urruty in [9] can be stated as follows:

(1) The (nonsmooth) *linear independence condition* holds at a feasible point z of (P) which lies in the interior of C provided that the vectors

$$\xi_j, \zeta_i, j = 1, 2, \ldots, l; i \in I := \{i: g(z) = 0, i = 1, 2, \ldots, d\}$$

are independent where ξ_j is any vector in $\partial h_j(z)$ and ζ_i is any vector in $\partial g_i(z)$.

(2) The (nonsmooth) *Mangasarian-Fromovitz condition* holds at a feasible point z of (P) which lies in the interior of C provided that the vectors ξ_j, $j = 1, 2, \ldots, l$ where ξ_j is any vector in $\partial h_j(z)$ are independent and for any $\xi_j \in \partial h_j(z)$, $j = 1, 2, \ldots, l$ and $\zeta_i \in \partial g(z)$, $i \in I$ there exists a vector $v \in R^n$ such that

$$\langle \xi_j, v \rangle = 0 \quad \forall j = 1, 2, \ldots, l$$
$$\langle \zeta_i, v \rangle < 0 \quad \forall i \in I.$$

The *generalized Slater condition* for problem (P) can be stated as follows: the (generalized) Slater condition holds at a feasible point z of (P) which lies in the interior of C provided

(1) h is continuous differentiable quasiconvex and quasiconcave, the vectors $\nabla h_j(z)$, $j = 1, 2, \ldots, l$ are independent;
(2) there exists a $z^0 \in C$ such that $g_i(z^0) < 0$ $\forall i \in I$ and $h(z^0) = 0$;
(3) For $i \in I$ g_i is continuously differentiable and pseudoconvex at z and for $i \notin I$ g_i is continuous.

It is well-known that the (generalized) Slater condition stated as above and the classical linear independence condition implies the classical Mangasarian-Fromovitz condition (see e.g. p. 171 of [4] for the proof). The fact that the nonsmooth linear independence stated as above implies the nonsmooth Mangasarian-Fromovitz condition can be proved in an exactly same way. The following result shows that the Mangasarian-Fromovitz condition is equivalent to $M^0(z) = \{0\}$.

Proposition 3.1: *The nonsmooth Mangasarian-Fromovitz condition holds at a feasible point z of (P) which lies in the interior of C if and only if $M^0(z) = \{0\}$.*

Proof: The proof that the nonsmooth Mangasarian-Fromovitz condition holds at a feasible point z of (P) implies $M^0(z) = \{0\}$ is exactly similar to the classical smooth case (see p. 235 of Clarke [5]).

Conversely, suppose that $M^0(z) = \{0\}$. Then the vectors ξ_j, $j = 1, 2, \ldots, l$ where $\xi_j \in \partial h(z)$ must be independent. Indeed, if not, there exists $r = (r_1, r_2, \ldots, r_l) \neq 0$ such that $0 = \sum_j r_j \xi_j$ which implies that $(r, 0) \in R^{l+d}$ is a nonzero vector in $M^0(z)$. For any $\xi_j \in \partial h_j(z)$, $j = 1, 2, \ldots, l$ and $\zeta_i \in \partial g_i(z)$, $i \in I$, there must exist a vector $v \in R^n$ such that

$$\langle \xi_j, v \rangle = 0 \quad \forall j = 1, 2, \ldots, l$$
$$\langle \zeta_i, v \rangle < 0 \quad \forall i \in I.$$

Indeed, if not, then by Motzkin's transposition theorem (see e.g. [8]), there must exist a solution (r, s) to the following system:

$$0 = \sum_j r_j \xi_j + \sum_{i \in I} s_i \zeta_i, \quad s_i \quad \text{not all zero}$$

i.e., $M^0(z)$ contains a nonzero vector. ■

Therefore the linear independence, Slater condition and Mangasarian-Fromovitz condition belong to the first category. Unfortunately, all the constraint qualifications fall in the first category in general do not hold for SLPP as shown in the following result.

Proposition 3.2: *Let (x, y) be a solution of* BLPP. *Suppose that x lies in the interior of X and every $y \in S(x)$ lies in the interior of Y. Assume that $M_x^0 S(x) = \{0\}$ and equality holds in* (1). *Then there exists a nontrivial abnormal multiplier for* SLPP, *i.e., the set of abnormal multipliers for* SLPP *corresponding to (x, y) contains a nonzero element.*

Proof: Since $M_x^0(y) = \{0\}$, there exist $(v, \pi) \in r^{c+d}$ such that

$$0 = \nabla_y f(x, y) + \nabla_y h(x, y) T v + \nabla_y g(x, y) T \pi$$

$$\pi \geqslant 0, \langle \pi, g(x, y) \rangle = 0.$$

Since $\nabla_x f(x, y) + \nabla_x h(x, y) T v + \nabla_x g(x, y) T \pi \in \partial V(x)$ by assumption that equality (1) holds in Proposition 2.1, we have

$$0 \in \nabla f(x, y) - \partial V(x) \times \{0\} + \nabla h(x, y) T v + \nabla g(x, y) T \pi$$

$$\pi \geqslant 0, \ \langle \pi, g(x, y) \rangle = 0$$

That is, $(1, v, \pi)$ is a nontrivial abnormal multiplier for SLPP corresponding to (x, y). The proof is complete. ■

The above discussion indicates that in general the usual constraint qualification is too strong for SLPP to hold. Therefore to find the *right constraint qualification* (i.e., it is satisfied by a large class of bilevel programming problems), one needs to look for constraint qualifications which *fall in the second category*. The following constraint qualification called *calmness condition*, first introduced by Clarke and Rockafellar, fall in the second cagegory. It is weaker then the constraint qualification $M^0(z) = \{0\}$ (see Corollary 5 of Theorem 6.5.2 in Clarke [5]). The calmness condition for SLPP is defined in terms of the following *fully perturbed* problem

$$\begin{aligned} P_{uv} \quad & \min \quad F(x, y) \\ & \text{s.t.} \quad f(x, y) - V(x) + u = 0 \qquad (3) \\ & \qquad h(x, y) + v_1 = 0 \\ & \qquad g(x, y) + v_2 \leqslant 0 \\ & \qquad x \in X, \ \ y \in Y, \end{aligned}$$

where $v = (v_1, v_2)$ and it is known (*c.f.* Clarke [5]) that the concept of calmness is closely related to a numerical technique called "exact penalization". Since we are particularly interested in perturbing the constraint (3) because it is the essential constraint that reflects the bilevel nature of the problem, we may consider the following partially perturbed problem:

$$
\begin{array}{lll}
P_u & \min & F(x,y) \\
 & \text{s.t.} & f(x,y) - V(x) + u = 0 \\
 & & h(x,y) = 0 \\
 & & g(x,y) \leqslant 0 \\
 & & x \in X, \quad y \in Y.
\end{array}
$$

Definition 3.1: *Let* (x,y) *solves* SLPP. SLPP *is partially calm at* (x,y) *provided that there exist* $\delta > 0$ *and* $\mu > 0$ *such that, for all* $u \in \delta B$, *for all* $(x',y') \in (x,y) + \delta B$ *which are feasible for* P_u, *we have*

$$F(x',y') - F(x,y) + \mu|u| \geqslant 0.$$

The partial calmness condition defined as above is obvious weaker than the calmness condition. In fact in Theorem 3.1, we will show that the partial calmness plus the usual constraint qualification for the lower level problem is a constraint qualification for SLPP.

The concept of partial calmness is actually equivalent to the "exact penalization" as shown in the following proposition whose proof is straightforward and will be omitted.

Proposition 3.3: *Suppose that* (x,y) *solves* SLPP. *Then* SLPP *is partially calm at* (x,y) *if and only if there exists a* $\mu > 0$ *such that* (x,y) *is a local optimal solution to the following penalized problem*

$$
\begin{array}{lll}
\widetilde{\text{SLPP}} & \min & F(x,y) + \mu(f(x,y) - V(x)) \\
 & \text{s.t.} & h(x,y) = 0 \\
 & & g(x,y) \leqslant 0 \\
 & & x \in X, \quad y \in Y.
\end{array}
$$

Theorem 3.1: *Let* (x^*,y^*) *solve* BLPP *and let* SLPP *be partially calm at* (x^*,y^*). *Suppose* x^* *lies in the interior of* X *and every* $y \in S(x^*)$ *lies in the interior of* Y. *Assume that* $M^0_{x^*}S(x^*) = \{0\}$. *Then the conclusion of Theorem 2.1 holds with* $\lambda = 1$ *and* $\mu > 0$.

Proof: By virtue of Proposition 3.3 there exists $\mu > 0$ such that (x^*,y^*) is an optimal solution to the penalized problem $\widetilde{\text{SLPP}}$. $M^0_{x^*}S(x^*) = \{0\}$ implies that $M_{x^*}(y^*) = \{0\}$. By Proposition 3.1, $M^0_{x^*}(y^*) = \{0\}$ if and only if the Mangasarian-Fromovitz condition holds at any y^*, i.e., $\nabla_y h_j(x^*,y^*)$ for $j = 1,2,\ldots,c$ are linearly independent and there exists a vector $v \in R^m$ such that

$$\langle \nabla_y h_j(x^*,y^*), v \rangle = 0, \quad j = 1,2,\ldots,c$$

$$\langle \nabla_y g_i(x^*,y^*), v \rangle < 0, \quad \text{if } i \in I := \{i : g_i(x^*,y^*) = 0\}.$$

Therefore $\nabla h_j(x^*, y^*)$ for $j = 1, 2, \ldots, c$ are linearly independent and there exists a vector $w = (0, v) \in R^{n+m}$ such that

$$\langle \nabla h_j(x^*, y^*), w \rangle = 0, \quad j = 1, 2, \ldots, c$$

$$\langle \nabla g_i(x^*, y^*), w \rangle < 0, \quad \text{if } i \in I := \{i : g_i(x^*, y^*) = 0\}.$$

That is, the Mangasarian-Fromovitz condition holds for the penalized problem $\widetilde{\text{SLPP}}$.

Applying the Kuhn-Tucker Lagrange multiplier rule (Proposition 6.4.4 of Clarke [5]) to $\widetilde{\text{SLPP}}$ and the estimate for $\partial V(x^*)$, the desired result follows. ■

Remark 3.1: The assumption $M^0_{x^*} S(x^*) = \{0\}$ can be replaced by any condition which ensures both the Lipschitz continuity of the valued function and the calmness of the penalized problem $\widetilde{\text{SLPP}}$.

4. OPTIMALITY CONDITIONS FOR THE MINIMAX PROBLEM AND THE BILEVEL PROGRAMMING PROBLEM WITH LINEAR LOWER LEVEL PROBLEM

4.1 The Minimax Problem

Consider the following *minmax problem*:

$$\min_{x \in X} \max_{y \in R^m} \{\phi(x, y) : h(x, y) = 0, g(x, y) \leqslant 0\},$$

where $\phi : R^{n+m} \to R$, $h : R^{n+m} \to R^c$, $g : R^{n+m} \to R^d$ are continuously differentiable. The minmax problem is a special case of BLPP with $-f(x, y) = F(x, y) := \phi(x, y)$.

Let the value function $V(x) := \inf\{-\phi(x, y) : h(x, y) = 0, g(x, y) \leqslant 0, y \in R^m\}$. Then the minmax problem is equivalent to the following problem

$$\begin{aligned} \min \quad & -V(x) \\ \text{s.t.} \quad & h(x, y) = 0 \\ & g(x, y) \leqslant 0 \\ & x \in X, \ y \in R^m, \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \min \quad & F(x, y) + (f(x, y) - V(x)) \\ \text{s.t.} \quad & h(x, y) = 0 \\ & g(x, y) \leqslant 0 \\ & x \in X, \ y \in R^m. \end{aligned}$$

Therefore by the equivalence of the exact penalty and the partial calmness (Proposition 3.3), the minmax problem is partially calm at any solution (x, y).

The necessary condition of Kuhn-Tucker type Theorem 3.1. for the minmax problem can be stated as follows:

Theorem 4.1: *Let (x^*, y^*) solve the minmax problem. Suppose x^* is in the interior of X. Suppose that $M^0_{x^*}S(x^*) = \{0\}$. Then there exists $s \in R^c$, $r \in R^d$, positive integers I, J, $\lambda_{ij} \geqslant 0$, $\sum_{i=1}^I \sum_{j=1}^J \lambda_{ij} = 1$, $y_i \in S(x^*)$, $v_{ij} \in R^c$, $\pi_{ij} \in R^d$ such that*

$$\begin{aligned}
0 &= \nabla_x h(x^*, y^*)^{\mathrm{T}} s + \nabla_x g(x^*, y^*)^{\mathrm{T}} r \\
&\quad + \sum_{ij} \lambda_{ij}(\nabla_x \phi(x^*, y_i) - \nabla_x h(x^*, y_i)^{\mathrm{T}} v_{ij} - \nabla_x g(x^*, y_i)^{\mathrm{T}} \pi_{ij}) \\
0 &= \nabla_y h(x^*, y^*)^{\mathrm{T}} s + \nabla_y g(x^*, y^*)^{\mathrm{T}} r \\
r &\geqslant 0, \langle r, g(x^*, y^*) \rangle = 0 \\
0 &= \nabla_y \phi(x^*, y_i) - \nabla_y h(x^*, y_i)^{\mathrm{T}} v_{ij} - \nabla_y g(x^*, y_i)^{\mathrm{T}} \pi_{ij} \\
\pi_{ij} &\geqslant 0, \langle \pi_{ij}, g(x^*, y_i) \rangle = 0.
\end{aligned}$$

In the unconstrained case, the above necessary condition reduces to the one derived by Schmitendorf in [15].

4.2 The Bilevel Programming Problem with Linear Lower Level Problem

Consider the following *bilevel programming problem with linear lower level problem*:

$$\begin{aligned}
\text{LBLPP} \quad & \min \quad F(x, y) \\
& \text{s.t.} \quad x \in R^n, y \in \arg\min_{y \in R^m} \{a^t x + b^t y : Cx + Dy - q \leqslant 0\}
\end{aligned}$$

where F is continuously differentiable, $a \in R^n$, $b \in R^m$, $C \in R^{d \times n}$, $D \in R^{d \times m}$, $q \in R^d$.

The equivalent formulation of problem LBLPP is

$$\begin{aligned}
\text{SLBLPP} \quad & \min \quad F(x, y) \\
& \text{s.t.} \quad a^t x + b^t y - V(x) = 0 \\
& \qquad\quad Cx + Dy - q \leqslant 0 \\
& \qquad\quad x \in R^n, y \in R^m,
\end{aligned}$$

where $V(x) := \min\{a^t x + b^t y : Cx + Dy - q \leqslant 0, y \in R^m\}$.

We now show that the bilevel programming problem with linear lower level problem is partially calm.

Proposition 4.1: *Let (x^*, y^*) solve* LBLPP, *then* SLBLPP *is partially calm at (x^*, y^*).*

Proof: Fix an arbitrary $\delta > 0$. Let $u \in \delta B$. Suppose that $(x', y') \in (x^*, y^*) + \delta B$ is feasible for the perturbed problem P_u, i.e.,

$$\begin{aligned}
a^t x' + b^t y' - V(x') + u &= 0 \qquad (4) \\
Cx' + Dy' - q &\leqslant 0.
\end{aligned}$$

Let $y(x')$ be a solution of the linear lower level problem, i.e.,

$$a^t x' + b^t y(x') - V(x') = 0 \tag{4}$$

$$Cx' + Dy(x') - q \leqslant 0.$$

Then $(x', y(x'))$ is feasible for LBLPP and by the optimality of (x^*, y^*), one has

$$\begin{aligned} F(x', y') - F(x^*, y^*) &\geqslant F(x', y') - F(x', y(x')) \\ &\geqslant -K_F \| y' - y(x') \|, \end{aligned} \tag{5}$$

where K_F is the Lipschitz constant of the function $y \to F(x, y)$. By definition of the value function, since (x', y') is feasible for P_u, u must be nonpositive. Therefore one has

$$\begin{aligned} \| y' - y(x') \| &= \min_{\varepsilon, y} \{\varepsilon : -\varepsilon e \leqslant y' - y \leqslant \varepsilon e, a^t x' + b^t y - V(x') = 0, Cx' + Dy - q \leqslant 0\} \\ &= \max_{\xi_1, \xi_2, \xi_3, \xi_4} \{(\xi_1 - \xi_2)^t y' + \xi_3(a^t x' - V(x')) + \xi_4^t (Cx' - q): \\ &\quad \xi_1 - \xi_2 - \xi_3 b - D^t \xi_4 = 0, e^t \xi_1 + e^t \xi_2 = 1, \xi_i \geqslant 0, i = 1, 1, 3, 4\} \\ &= \max \{\xi_3(a^t x' + b^t y' - V(x')) + \xi_4^t (Cx' + Dy' - q): \\ &\quad e^t \xi_1 + e^t \xi_2 = 1, \xi_i \geqslant 0, i = 1, 2, 3, 4\} \end{aligned} \tag{6}$$

by linear programming duality

$$= \xi_3(x', y')(a^t x' + b^t y' - V(x')) + \xi_4^t(x', y')(Cx' + Dy' - q)$$

where $(\xi_1(x', y'), \xi_2(x', y'), \xi_3(x', y'), \xi_4(x', y'))$ is a solution to the maximization problem (6)

$$\leqslant \xi_3(x', y')(a^t x' + b^t y' - V(x')) = \xi_3(x', y')(-u) = \xi_3(x', y')|u|, \tag{7}$$

where $e \in R^m$ is the vector with all components equal to 1. Since $\xi_3(x', y')$ may be chosen as a component of a solution vertex of the linear programming problem (6) and since the feasible region of (6) is independent of (x', y') and has a finite number of vertices, one has

$$\xi_3(x', y') \leqslant M := \max \{\xi_3 : (\xi_1, \xi_2, \xi_3, \xi_4) \text{ is a vertex of the constraint region of } (6)\}. \tag{8}$$

Combining (5), (7) and (8), we conclude that problem SLBLPP is partially calm with $\mu = K_F M$. ■

Since the value function for the linear lower level problem is convex therefore Lipschitz continuous and the corresponding constraints of the penalized problem SLPP is a linear system therefore regular, in the view of Theorem 3.1 and Remark 3.1 we have the Kuhn-Tucker type Lagrange multiplier rule for LBLPP in the following form

Theorem 4.2: *Let (x^*, y^*) be a solution to* LBLPP. *Then there exist $r \in R^d$, $\mu > 0$, positive integers I, J, $\lambda_{ij} \geqslant 0$, $\sum_{i=1}^{I} \sum_{j=1}^{J} \lambda_{ij} = 1$, $y_i \in S(x^*)$, $\pi_{ij} \in R^d$ such that*

$$0 = \nabla_x F(x^*, y^*) + C^t r - \mu C^t \left(\sum_{ij} \lambda_{ij} \pi_{ij} \right)$$

$$0 = \nabla_y F(x^*, y^*) + D^t r + \mu b$$

$$r \geqslant 0, \langle r, Cx^* + Dy^* - q \rangle = 0$$

$$0 = b + D^t \pi_{ij}$$

$$\pi_{ij} \geqslant 0, \langle \pi_{ij}, Cx^* + Dy_i - q \rangle = 0.$$

5. UNIFORMLY WEAK SHARP MINIMUM AS A SUFFICIENT CONDITION FOR PARTIAL CALMNESS

A function defined on a set $\Omega \subset R^n$ is said to have a *sharp minimum* at $\bar{x} \in \Omega$ if there exists an $\alpha > 0$ such that

$$\phi(x) - \phi(\bar{x}) \geqslant \alpha \|x - \bar{x}\| \quad \forall x \in \Omega.$$

It is obvious that the notion of a sharp minimum implies uniqueness of the solution of the problem. In order to include the possibility of a non-unique solution set, Ferris [7] extended the notion of a sharp minimum to a weak sharp minimum as follows: Let S be the non-empty optimal solution set of the problem

$$\min_{x \in \Omega} \phi(x).$$

We say ϕ has a *weak sharp minimum* (on Ω) if there exists an $\alpha > 0$ such that

$$\phi(x) - \phi(y) \geqslant \alpha \operatorname{dist}(x, S) \quad \forall y \in S, x \in \Omega.$$

We now extend the notion of a weak sharp minimum to the family of parametric mathematical programming problem $\{(P_x)\colon x \in X\}$. We say that $\{(P_x)\colon x \in X\}$ has a uniformly weak sharp minimum if there exists an $\alpha > 0$ such that

$$f(x, y) - V(x) \geqslant \operatorname{dist}(y, S(x)) \quad \forall (x, y) \in X \times Y$$
$$\text{s.t. } h(x, y) = 0,\ g(x, y) \leqslant 0. \tag{9}$$

The following result indicates that the uniformly weak sharp minimum is a sufficient condition for partial calmness.

Proposition 5.1: *Suppose $F(x, y)$ is Lipschitz continuous in y uniformly in x with Lipschitz constant L_F and $\{(P_x)\colon x \in X\}$ has a uniformly weak sharp minimum. Let (x^*, y^*) solve BLPP. Then SLPP is partially calm at (x^*, y^*).*

Proof: Fix any arbitrary $\delta > 0$. Let $u \in \delta B$. Let $(x, y) \in (x^*, y^*) + \delta B$ be feasible for the perturbed problem P_u and $y(x)$ by the projection of x onto the set $S(x)$. Then by

Lipschitz contibuity of F and uniformly weak sharp minimum property (9), one has

$$\begin{aligned} F(x,y)-F(x^*,y^*) &\geqslant F(x,y)-F(x,y(x)) \\ &\geqslant -K_F\|y-y(x)\| \\ &\geqslant -K_F\alpha(f(x,y)-V(x)) \\ &= -K_F\alpha(-u) = -K_F\alpha|u|, \end{aligned}$$

i.e., SLPP is partially calm in (x^*, y^*). ■

Now consider the bilevel programming problem where the lower level problem is the following parameteric quadratic programming problem:

$$\begin{aligned} QP_x \quad \min \quad & f(x,y) := \tfrac{1}{2}\langle y, Qy\rangle + b^t y + \langle y, Ex\rangle \\ \text{s.t.} \quad & y\in\Omega_x, \end{aligned}$$

where $Q\in R^{m\times m}$ is a symmetric and positive semidefined matrix, $b\in R^m$, $E\in R^{m\times n}$ and for any given $x\in X$, Ω_x is a polyhedral in R^{n+m}. The following result gives a sufficient condition for the uniformly weak sharp minimum of $\{(QP_x): x\in X\}$ and hence a sufficient condition for the partial calmness of the BLPP where the lower level problem is (QP_x). It is the generalization of the necessity part of Theorem 3.2 of Burke and Ferris [2]) to the parametric case. Although the proof technique are taken from Burke and Ferris [2], we include it here for the completeness.

Proposition 5.2: *Suppose either*

$$\nabla_y f(x,y) = 0 \quad \forall x\in X, y\in S(x) \tag{10}$$

or there exists a $\alpha > 0$ *such that*

$$\|\nabla_y f(x,y)\| \geqslant \alpha > 0, \quad \forall x\in X, y\in S(x) \quad \text{s.t.}\ \|\nabla_y f(x,y)\| \neq 0. \tag{11}$$

Then $\{\{(QP_x): x\in X\}$ *has a uniformly weak sharp minimum provided*

$$(\ker(\nabla_y^2 f(x,\bar y))^{\perp} \subset \operatorname{span}(\nabla_y f(x,\bar y)) + N(y,\Omega_x), \quad \forall y\in S(x) \tag{12}$$

or equivalently

$$(\nabla_y f(x,\bar y))^{\perp} \cap T(y,\Omega_x) \subset \ker(\nabla_y^2 f(x,\bar y)) \quad \forall x\in X, y\in S(x), \tag{13}$$

where $\bar y$ *is any element in* $S(x)$, $A^{\perp}: \{y\in R^m: \langle y,x\rangle = 0\ \forall x\in A\}$ *denotes the subspace pependicular to* A, $\operatorname{span}(d)$ *represents the subspace generated by* d, $T(y,A)$ *and* $\ker(A)$ *denote the tangent cone to* A *at* y *and the nullspace of the matrix* A *respectively.*

Before we prove the above result we first state the following discription of the solution set of a convex program given by Mangasarian [11].

Lemma 5.1: *Let* S *be the set of of solutions to the problem* $\min\{f(y): y\in\Omega\}$ *where* $f: R^n \to R$ *is a twice continuous convex function and* Ω *is a convex subset of* R^n. *Let* $\bar y\in S$. *Then*

$$S = \{y\in\Omega: \nabla f(y) = \nabla f(\bar y), \langle\nabla f(\bar y), y-\bar y\rangle = 0\}.$$

It is clear that for (QP_x), this gives the solution set $S(x)$ as

$$S(x)=\Omega_x\cap\{y:\langle\nabla_y f(x,\bar{y}),y-\bar{y}\rangle=0\}\cap\{y:\nabla_y^2 f(x,\bar{y})(y-\bar{y})=0\}$$

and since Ω_x is a polyhedral one has

$$T(y,S(x))=T(y,\Omega_x)\cap(\nabla_y f(x,\bar{y}))^{\perp}\cap\ker(\nabla_y^2 f(x,\bar{y})) \tag{14}$$

by virtue of Corollaries 16.4.2 and 23.8.1 of Rockafellar [14].

Proof of Proposition 5.2: By virtue of Theorem 2.6 of Burke and Ferris [2], it suffices to show that for all $x\in X$, $y\in S(x)$, there exists an $\alpha>0$ such that

$$f_2'(x,y;d)\geqslant\alpha\|d\| \quad \forall d\in T(y,\Omega_x)\cap N(y,S(x)),$$

where $f_2'(x,y;d)$ denotes the directional derivative of f with respect y in direction d. Note that (14) and (12) implies that

$$\begin{aligned}K(x)&:=T(y,S(x))^o=N(y,\Omega_x)+\operatorname{span}(\nabla_y f(x,\bar{y}))+(\ker(\nabla_y^2 f(x,\bar{y}))^{\perp}\\&=N(y,\Omega_x)+\operatorname{span}(\nabla_y f(x,\bar{y})),\end{aligned}$$

where $A^o:=\{x^*\in R^m:\langle x^*,x\rangle\leqslant 0\ \forall x\in A\}$ denotes the polar of A. Therefore span $(\nabla_y f(x,\bar{y}))=$ recession cone of $K(x)$. It follows from p. 65 of Rockafellar [14] that

$$K(x)=\operatorname{span}(\nabla_y f(x,\bar{y}))+(K(x)\cap(\nabla_y f(x,\bar{y}))^{\perp}). \tag{15}$$

Since $d\in T(y,\Omega_x)\cap N(y,S(x))$, one has

$$\begin{aligned}\alpha\|d\|&=\alpha\operatorname{dist}(d,T(y,S(x))\\&=\alpha\sup\{\langle z(x),d\rangle:z(x)\in B\cap K(x)\},\end{aligned}$$

where the last equality follows from Theorem 3.1 of Burke and Han [3]. By virtue of (15), $z(x)\in B\cap K(x)$ implies

$$z(x)=\lambda(x)\nabla_y f(x,\bar{y})+\zeta(x)$$

with $|\lambda(x)|\leqslant\eta(x)$ where

$$\eta(x)=\begin{cases}1/\|\nabla_y f(x,\bar{y})\| & \text{if } \|\nabla_y f(x,\bar{y})\|\neq 0\\ 0 & \text{otherwise}\end{cases}$$

and $\zeta(x)\in K(x)\cap(\nabla_y f(x,\bar{y}))^{\perp}$. Therefore

$$\begin{aligned}\alpha\|d\|&=\alpha\sup\{\langle\lambda(x)\nabla_y f(x,\bar{y})+\zeta(x),d\rangle:|\lambda(x)|\leqslant\eta(x),\\&\qquad\zeta(x)\in N(y,\Omega_x)\cap(\nabla_y f(x,\bar{y}))^{\perp}\}\\&=\alpha\sup\{\langle\lambda(x)\nabla_y f(x,\bar{y}),d\rangle:|\lambda(x)|\leqslant\eta(x)\}\\&\leqslant\alpha\eta(x)\langle\nabla_y f(x,\bar{y}),d\rangle\\&\leqslant\langle\nabla_y f(x,\bar{y}),d\rangle=\langle\nabla_y f(x,y),d\rangle=f_2'(x,y;d). \quad ■\end{aligned}$$

Remark 5.1: A generalization of Proposition 5.2 to the case where Ω_x is a nonpolyhedral convex set can be easily obtained. Observe that the argument given above only employs the polyhedrality of Ω_x to establish that (14) holds. However by Corollaries 16.4.2 and 23.8.1 of Rockafellar [14], (14) also holds under the assumption

$$ri\,(\Omega_x - \bar{y}) \cap (\nabla_y f(x, \bar{y}))^{\perp} \cap \ker(\nabla_y^2 f(x, \bar{y})) \neq \varnothing, \tag{16}$$

where riA debotes the relative interior of A. Therefore when the Ω_x is nonpolyhedral convex set the conclusion of Proposition 5.2 is true under the additional assumption (16).

The following result is an easy consequence of Propositions 5.1 and 5.2.

Corrollary 5.1: *The bilevel programming problem where the lower level problem is the quadratic programming problem QP_x is partially calm provided either* (10) *or* (11) *and either* (12) *or* (13) *hold.*

Remark 5.2: Note that it is easy to verify that when P, Q, and E, are all equal to zero, assumptions of Proposition 5.2 are satisfied. Therefore the BLPP with linear lower level problem is partially calm as it was proved directly in Proposition 4.1.

Consider now a BLPP where the lower level problem is the following parametric linear complementarity problem:

$$\begin{aligned} LCP_x \quad \min \quad & f(x,y) := \langle y, My + q + Ex \rangle \\ \text{s.t.} \quad & My + q + Ex \geqslant 0, y \geqslant 0 \end{aligned}$$

where $M \in R^{m \times m}$ is a symmetric and positive semidefinite matrix, $q \in R^m$ and $E \in R^{m \times m}$. Let A_i denote the ith row of a matrix A and d_i denote the ith component of a vector d. We make a nondegeneracy assumption that there is a solution $\hat{y}(x)$ of LCP_x which satisfies

$$I_x(\hat{y}(x)) \cap J_x((\hat{y}(x)) = \varnothing,$$

where $I_x(y) := \{i : M_i y + q_i + E_i x = 0\}$ and $J_x(y) := \{j : y_j = 0\}$. Obviously LCP_x is a special case of QP_x. As in the nonparametric case (See Theorem 3.7 of [2]), it is easy to verify that under either (10) or (11), (13) always holds. Therefore the following is an immediate consequence of Corollary 5.1.

Corollary 5.2: *The bilevel programming problem where the lower level problem is the nondegenerate linear complementarity problem LCP_x is partially calm provided either* (10) *or* (11) *hold.*

We now give an example to illustrate the use of Corollary 5.1.

Example 5.1:

$$\begin{aligned} &\min \quad F(x, y_1, y_2) \\ &\text{s.t} \quad x \in [-0.5, 0.5], (y_1, y_2) \in \arg\min_{y1,y2} \{\tfrac{1}{2} y_1^2 + y_1 x : 1 \leqslant y_1 \leqslant 2, 1 \leqslant y_2 \leqslant 2\}. \end{aligned}$$

The set of optimal solution for the lower level problem is

$$S(x) = \{1\} \times [1,2], \quad \forall x \in [-0.5, 0.5]$$

For all $x \in [-0.5, 0.5]$, $y \in S(x) = \{1\} \times [1,2]$, $\|\nabla_y f(x,y)\| = \|(y_1 + x, 0)\| = 1 + x \geqslant 1/2 > 0$. So assumption (11) is satisfied. Since $(\nabla_y f(x,y))^{\perp} = \{0\} \times R$, $\ker \nabla_y^2 f(x,y) = \{0\} \times R$ for all $(x,y) \in [-0.5, 0.5] \times \{1\} \times [1,2]$ and

$$T(y, \Omega_x) = [0, \infty] \times T(y_2, [1,2]) \quad \forall x \in [-0.5, 0.5], y \in S(x)$$

assumption (13) is satisfied. By Corollary 5.1, the BLPP considered is partially calm.

Now we give an example to show that uniformly weak sharp minimum is only a sufficient condition for partial calmness but not necessary.

Consider problem

$$\begin{aligned} &\min \quad x + y \\ &\text{s.t.} \quad x \in [-1, 1], y \in \arg\min_y \{-2xy + y^2, -1 \leqslant y \leqslant 1\}. \end{aligned}$$

The solution set of the lower level problem $S(x) = \{x\}$ for $x \in [-1, 1]$. It is obvious that for any x, the lower level problem does not have a sharp minimum hence no uniformly weak sharp minimum.

The value function $V(x) = -x^2$, for all $x \in [-1, 1]$. The equivalent formulation SLPP is

$$\begin{aligned} \min_{(x,y) \in R^2} \quad & x + y \\ \text{s.t.} \quad & (x - y)^2 = 0 \\ & -1, \leqslant x \leqslant 1 \\ & -1 \leqslant y \leqslant 1. \end{aligned}$$

It is obvious that $(x, y) = (-1, -1)$ is the unique optimal solution of the original problem and the following penalized problem

$$\begin{aligned} \min \quad & x + y + \mu(x - y)^2 \\ \text{s.t.} \quad & -1 \leqslant x \leqslant 1 \\ & -1 \leqslant y \leqslant 1 \end{aligned}$$

for any $\mu > 0$. Therefore by the equivalence of the exact penalty and the partial calmness (Proposition 3.3), the problem considered is partially calm.

6. SUFFICIENT OPTIMALITY CONDITIONS

In this section, we give sufficient optimality conditions for the following bilevel programming problem:

$$\begin{aligned} (P) \quad \min \quad & F(x, y) \\ \text{s.t.} \quad & x \in X, y \in \arg\min_y \{f(x, y), y \in Y\}, \end{aligned}$$

where $X \subset R^n$, $Y \subset R^m$, $F: R^{n+m} \to R$ and $f: R^{n+m} \to R$ are twice continuously differentiable.

Theorem 6.1: *Let (x^*, y^*) be feasible for (P) which lies in the interior of $X \times Y$. If (x^*, y^*) satisfies the necessary condition of Kuhn-Tucker type, i.e., there exist $\mu \geqslant 0$, a positive integer I, $\lambda_i \geqslant 0$, $\sum_{i=1}^{I} \lambda_i = 1$, $y_i \in S(x^*)$ such that*

$$0 = \nabla_x F(x^*, y^*) + \mu(\nabla_x f(x^*, y^*) - \sum_i \lambda_i \nabla_x f(x^*, y_i)) \tag{17}$$

$$0 = \nabla_y F(x^*, y^*) + \mu \nabla_y f(x^*, y^*) \tag{18}$$

and the matrix

$$\frac{\partial^2}{\partial(x, y)^2}[F(x, y) + \mu(f(x, y) - \sum_i \lambda_i f(x, y_i))]$$

is positive semidefinite for any $x \in R^n$, $y \in R^m$ then (x^*, y^*) is an optimal solution to (P).

Proof: Take any $x^0 \in X$ and $y^0 \in S(x^0)$. By Taylor's theorem, we have

$$F(x^0, y^0) + \mu\left(f(x^0, y^0) - \sum_i \lambda_i f(x^0, y_i)\right) - F(x^*, y^*) - \mu\left(f(x^*, y^*) - \sum_i \lambda_i f(x^*, y_i)\right)$$

$$= \left[\nabla F(x^*, y^*) - \mu\left(\nabla f(x^*, y^*) - \sum_i \lambda_i \nabla_x f(x^*, y_i) \times \{0\}\right)\right][(x^0, y^0) - (x^*, y^*)]$$

$$+ \tfrac{1}{2}[(x^0, y^0) - (x^*, y^*)]^{-1} \frac{\partial^2}{\partial(x, y)^2}\left[F(x, y) + \mu\left(f(x, y) - \sum_i \lambda_i f(x, y_i)\right)\right]_{(x,y) = (\tilde{x}, \tilde{y})}$$

$$[(x^0, y^0) - (x^*, y^*)]$$

where $(\tilde{x}, \tilde{y}) = \rho(x^0, y^0) + (1 - \rho)(x^*, y^*)$, $0 \leqslant \rho \leqslant 1$. Since the first term on the right hand side of the above equation is zero by virtue of (17) and (18) and the second term on the right hand side is greater and equal to zero by assumption, we have

$$F(x^0, y^0) + \mu\left(f(x^0, y^0) - \sum_i \lambda_i f(x^0, y_i)\right) \geqslant F(x^*, y^*) - \mu\left(f(x^*, y^*) - \sum_i \lambda_i f(x^*, y_i)\right) \tag{19}$$

Since $y_i \in S(x^*)$, we have

$$f(x^*, y_i) = \min_{y \in Y} f(x^*, y) = f(x^*, y^*).$$

Therefore,

$$f(x^*, y^*) - \sum_i \lambda_i f(x^*, y_i) = 0. \tag{20}$$

Since

$$f(x^0, y_i) \geqslant \min_{y \in Y} f(x^0, y) = f(x^0, y^0)$$

and $\mu \geqslant 0$, we have

$$\mu(f(x^0, y^0) - \sum_i \lambda_i f(x^0, y_i)) \leqslant 0 \tag{21}$$

By virtue of (19), (20) and (21), we have

$$F(x^0, y^0) \geqslant F(x^0, y^0) + \mu(f(x^0, y^0) - \sum_i \lambda_i f(x^0, y_i)) \geqslant F(x^*, y^*).$$

That is, (x^*, y^*) is an optimal solution for (P_3). ■

Theorem 6.2: *Let $(x^*, y^*) \in X \times Y$ be feasible for (P) which lies in the interior of $X \times Y$. Suppose that $S(x^*) = \{y^*\}$ is a singleton. If the necessary condition of Kuhn-Tucker type is satisfied at (x^*, y^*), i.e.*

$$0 = \nabla F(x^*, y^*) \tag{22}$$

and $(\partial^2/\partial(x, y)^2)F(x^*, y^*)$ is positive definite for (x^*, y^*), then (x^*, y^*) is a strict local minimum point of (P).

Proof: If (x^*, y^*) is not a strict local minimum point, there exists a sequence of feasible points $\{x_k, y_k\}$ converging to $\{x^*, y^*)$ such that for each k,

$$F(x_k, y_k) \leqslant F(x^*, y^*). \tag{23}$$

Write each (x_k, y_k) in the form $(x_k, y_k) = (x^*, y^*) + \delta_k s_k$ where $|s_k| = 1$ and $\delta_k > 0$ for each k. Clearly, $\delta_k > 0$ $\delta_k \to 0$ since $S(x^*)$ is a singleton and $y_k \in S(x_k)$ and the sequence $\{s_k\}$, being bounded, must have a convergent subsequence converging to some s^*. For convenience of notation, we assume that the sequence $\{s_k\}$ is itself convergent to s^*. Now by Taylor's theorem, we have

$$F(x_k, y_k) - F(x^*, y^*) = \nabla F(x^*, y^*)\delta_k s_k + \frac{1}{2}\delta_k^2 s_k^{-1} \frac{\partial^2}{\partial(x, y)^2} F(\tilde{x}_k, \tilde{y}_k) s_k$$

where $(\tilde{x}_k, \tilde{y}_k) = \rho(x_k, y_k) + (1 - \rho)(x^*, y^*)$, $0 \leqslant \rho \leqslant 1$. By the first order condition (22) and the assumption (23), we have

$$\frac{1}{2}\delta_k^2 s_k^{-1} \frac{\partial^2}{\partial(x, y)^2} F(\tilde{x}_k, \tilde{y}_k) s_k \leqslant 0$$

which yields a contradiction as $k \to \infty$. Therefore (x^*, y^*) is a strict local minimum point of (P) and the proof of the theorem is complete. ■

7. CONCLUSION AND FUTURE RESEARCH

In this paper, we have identified the difficulty in deriving optimality conditions for the BLPP. We have shown that in general the usual constraint qualifications do not hold for the equivalent single level problem SLPP and the right constraint qualification is the (partial) calmness condition. Moreover, we have shown that the calmness condition is satisfied automatically for any bilevel programming problem with linear lower level problem and the minmax problem. We identify the uniformly weak sharp minimum as a sufficient condition for the (partial) calmness. As a consequence, sufficient conditions for calmness condition for the bilevel problem with quadratic lower level problem and nondegenerate linear complementarity problem are given. Searching for sufficient conditions for the calmness condition for the bilevel programming problem with more general lower level problem will be a subject of the future research. Sufficient conditions for general constrained BLPP is also a subject of the future study.

Acknowledgements

The authors would like to thank Professor P. Marcoote for carefully reading this paper and giving helpful comments. The research of the first author was supported by NSERC under grant WFA 0123160.

References

[1] Bard J. F. and Falk, J. E. (1982) An explicit solution to the multilevel programming problem, *Computers and Operations Research* **9**, 77–100
[2] Burke J. V. and Ferris, M. C. (1993) Weak sharp minima in mathematical programming, *SIAM J. Control Optim.* **31**, 37–46
[3] Burke J. V. and Han, S. P. (1986) A Gauss-Newton approach to solving generalized inequalities, *Math. of Oper. Res.* **11**, 632–643
[4] Bazaraa M. S. and Shetty, C. M. (1979) "Nonlinear Programming", John Wiley and Sons, New York
[5] Clarke, F. H. (1983) "Optimization and Nonsmooth Analysis", Wiley-Interscience, New York
[6] Dempe, S. (1992) A necessary and sufficient optimality for bilevel programming problems, Optimization **25**, 341–354
[7] Ferris, M. C. Weak sharp minima and penalty functions in mathematical programming, Technical Report 779, Computer Sciences Department, University of Wisconsin, June 1988
[8] Mangasarian O. L. and Fromovitz, S. (1967) The Fritz John necessary optimality conditions in the presence of equality and inequality constraints, *Journal of Mathematical Analysis and Applications* **17**, 37–47
[9] Hiriart-Urruty J. B. (1979) Refinement of necessary optimality conditions in nondifferentiable programming I, *Appl. Math. Optim.* **5**, 63–82
[10] Loridan, P. and Morgan, J. (1989) A theoretical approximation scheme for Stackelberg problem, *Journal of Optimization Theory and Applications* **11**, 95–110
[11] Mangasarian, O. L. (1988) A simple characterization of solution sets of convex programs, *Operations Research Letters* 7, 21–26
[12] Outrata, J. V. (1990) On the numerical solution of a class of Stackelberg problems, *Zeitschrift für Operations Research* **4**, 255–278
[13] Outrata, J. V. (1993) Necessary optimality conditions for Stackelberg problems, *Journal of Optimization Theory and Applications*, **76**, 305–320
[14] Rockafellar, R. T. (1970) "Convex Analysis", Princeton University Press, Princeton, NJ
[15] Schmitendorf, W. E. (1977) Necessary conditions and sufficient conditions for static minmax problems, *Journal of Mathematical Analysis and Applications* **57**, 683–693
[16] Von Stackelberg, H. (1952) "The Theory of the Market Economy", Oxford University Press, Oxford. England
[17] Shimizu K. and Aiyoshi, E. (1981) A new computational method for Stackelberg and minimax problems by use of a penalty method, *IEEE Transactions on Automatic Control* **26**, 460–466
[18] Simaan M. and Cruz, J. B. (1973) On the Stackelberg strategy in nonzero-sum games, *Journal of Optimization Theory and Applications* **11**, 535–555
[19] Zhang, R. Problems of hierarchical optimization in finite dimensions, To appear in *SIAM J. Optimization*