OPTIMALITY CONDITIONS FOR BILEVEL PROGRAMMING PROBLEMS

J.J. YE and D.L. ZHU

DMS-618-IR

September 1992
[Revised July 1993]

OPTIMALITY CONDITIONS FOR BILEVEL PROGRAMMING PROBLEMS*

J.J. Ye[†] and D.L. Zhu[‡]

August 3, 1993

Abstract

The bilevel programming problem (BLPP) is a sequence of two optimization problems where the constraint region of the upper level problem is determinied 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 the usual constraint qualifications do not hold and the right constraint qualification is the calmness condition. It is also shown that for certain problems such as the linear bilevel programming problem and the minmax problem, the calmness condition is satisfied. First order necessary optimality conditions are given by using the nonsmooth analysis technique. Second-order sufficient optimality conditions are also given for the case where the lower level problem is unconstrained.

Key words: optimality conditions, bilevel programming problem, nonsmooth analysis, constraint qualification

*This work was supported in part by NSERC under grant WFA 0123160

[†]Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada V8W 3P4.

[‡]GERAD, Groupe d'études et de recherche en analyse des décisions, Montréal, Québec, Canada H3T 1V6.

1 Introduction

Let us consider a two-level hierachical system where the higher level (hereafter the "leader") and the lower level (hereafter the "follower") must find vectors $x \in X$ and $y \in Y$, respectively, to minimize their individual objective functions $F(x, y)$ and $f(x, y)$. The leader is assumed to select his decision vector $x \in X$ first and the follower to select his decision vector $y \in Y$ after that, where X and Y are nonempty subsets of R^n and R^m respectively. Under these assumptions on the order of play, the game will proceed as follows: Given any decision vector $x \in X$ chosen by the leader, the follower will select his decision vector $y \in Y$ (dependent on the decision vector x chosen by the leader) to minimize his objective $f(x, y)$ subject to contraints $g(x, y) \leq 0$. Let the solution set of the lower level problem be $S(x)$. Assume that the game is co-operative, i.e., if the follower's problem has several optimal solutions for a given x, 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 other words, given any decision vector $x \in X$ chosen by the leader, the follower faces the ordinary (*single level*) mathematical programming problem parametered in x:

$$
\begin{array}{lll}
(P_x) & \min & f(x,y) \\
 & \text{s.t.} & g(x,y) \leq 0 \\
 & y \in Y, &
\end{array}
$$

while the leader faces the *bilevel programming problem:*

$$
\begin{array}{lll}
\text{BLPP} & \min & F(x,y) \\
 & \text{over} & x \in X \ \text{ and all optimal solution } y \text{ of } (P_x),
\end{array}
$$

where $F, f : R^{n_1+n_2} \to R$, $g : R^{n_1+n_2} \to R^m$ are C^1 (continuously differentiable) functions.

The bilevel programming problem can be viewed as a two-person nonzero-sum game with perfect information where the order of play is specified at the outset and the players' strategy sets are no longer assumed to be disjoint. Minmax problems are special cases of bilevel programming problems where the sum of the objective functions of the two players is equal to zero.

Although numerical algorithms have been studied for some special BLPPs, to our knowledge, there are no applicable optimality conditions for the general bilevel programming BLPP to date. Bard [1] proposed a set of necessary conditions which was shown to be invalid by Clark and Westerberg [4]. The purpose of this paper is to fill this gap by providing applicable optimality conditions for bilevel programming programming problems. Notice that if we define the *value function of the lower level programming problem* as an extended value function $V : X \longrightarrow \bar{R}$ defined by

$$V(x) := \inf_{y \in Y} \{f(x,y) : g(x,y) \leq 0\}, \tag{1}$$

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

$$\begin{array}{lll} \text{SLPP} & \min & F(x,y) \\ & \text{s.t.} & f(x,y) - V(x) = 0 \\ & & g(x,y) \leq 0 \\ & & x \in X, \; y \in Y. \end{array}$$

The significance of this formulation was elaborated by Chen and Florian in [3]. However, $V(x)$ is not smooth in general even in the case where all problem data $F(x,y), f(x,y), g(x,y)$ are smooth functions. In fact, under certain assumptions, $V(x)$ is known to be locally Lipschitz continuous. Therefore, nonsmooth anslysis will be used to derive a Lagrange multipler rule. The difficulty with the derivation of a necessary condition is with the constraint qualification. It will be shown that the usual constraint qualifications for mathematical programming problems are not applicable to SLPP (equivalently, for bilevel programming problems) and that the calmness condition is the right constraint qualification condition. It will be shown that for linear bilevel programming problems and minmax problems the calmness condition is satisfied. Sufficient conditions for optimality are also given for the case where the lower level problem (P_x) is not subject to the inequality constraints $g(x,y) \leq 0$.

2 Nonsmooth analysis background

In this section we shall give a concise review of the material on nonsmooth analysis which will be required. Our references are Clarke [5], [6] and Rockafallar [11].

Let C be a nonempty closed set in R^n. A vector $\zeta \in R^n$ is a *proximal normal* to C at point $\bar{x} \in C$ if for $t > 0$ sufficiently small, the unique point of C nearest to

$\bar{x} + t\zeta$ (in the Euclidean norm) is $\bar{x}$. It is a *limiting proximal normal* if there exist points $x^k \in C, x^k \to \bar{x}$, and proximal normals ζ^k to C at x^k, such that $\zeta^k \to \zeta$. Let the *limiting proximal normal cone* be the set

$$\hat{N}_C(\bar{x}) := \{\zeta : \zeta \text{ is a limiting proximal normal to } C \text{ at } \bar{x}\}.$$

Consider now a lower semicontinuous function $\phi : R^n \to R \cup \{+\infty\}$ and a point $\bar{x} \in R^n$ where ϕ is finite. The *epigraph* of ϕ is the set

$$\text{epi}\phi(x) := \{(x, \alpha) \in R^n \times R : \alpha \geq \phi(x)\}.$$

A *limiting subgradient* of ϕ at $\bar{x}$ is the set

$$\hat{\partial}\phi(\bar{x}) := \{\zeta : (\zeta, -1) \in \hat{N}_{\text{epi}\phi}(\bar{x}, \phi(\bar{x}))\}.$$

A *singular limiting subgradient* of ϕ at $\bar{x}$ is the set

$$\hat{\partial}^\infty\phi(\bar{x}) := \{\zeta : (\zeta, 0) \in \hat{N}_{\text{epi}\phi}(\bar{x}, \phi(\bar{x}))\}.$$

We shall say that $\phi(\cdot)$ is *Lipschitz continuous near* x provided that there exist a constant $K > 0$ and an open neighborhood O_x of x such that

$$|\phi(x^1) - \phi(x^2)| \leq K||x^1 - x^2|| \qquad \forall x^1, x^2 \in O_x.$$

The limiting subgradient is a smaller object than the *Clarke generalized gradient.* In fact, if ϕ is Lipschitz continuous near x, we have $\partial\phi(x) = \text{co}\hat{\partial}\phi(x)$, where ∂ and coA denote the Clarke generalized gradient and convex hull of set A respectively. For the definition and the precise relation between the limiting subgradient and the Clarke generalized gradient, the reader is referred to Clarke [6] and Rockafellar [11] for more details.

The following proposition summarizes the requisites regarding limiting subgradients and Clarke normal cones.

Proposition 2.1 (a) If C is a nonempty closed convex set, the Clarke normal cone to C, denoted by N_C coincides with the normal cone in the sense of convex analysis, i.e., any vector $\zeta \in N_C(\bar{x})$ if and only if

$$\langle \zeta, x - \bar{x} \rangle \leq 0, \quad \forall x \in C.$$

(b) [Proposition 1.2, Clarke [6]] The function $\phi(\cdot)$ is Lipschitz near x if and only if $\hat{\partial}^\infty\phi(x) = \{0\}$.

(c) [*c.f.* Proposition 2.2.4, Clarke [5]] The function $\phi(\cdot)$ is continuously differentiable on an open neighborhood of x if and only if $\hat{\partial}\phi(x)$ is single valued on that neighborhood. In this case, the limiting subgradient coincides with the usual gradient.

(d) [Proposition 1.5, Clarke [6]] Let ϕ and $\psi : R^n \to R \cup \{+\infty\}$ be lower semi-continuous functions finite at x, with $\hat{\partial}^\infty\phi(x) \cap -\hat{\partial}^\infty\psi(x) = \{0\}$. Then we have

$$\hat{\partial}(\phi + \psi)(x) \subset \hat{\partial}\phi(x) + \hat{\partial}\psi(x). \tag{2}$$

(e) For any $\alpha \geq 0$, one has

$$\hat{\partial}(\alpha\phi)(x) = \alpha\hat{\partial}\phi(x).$$

Let us now consider the general mathematical programming problem in the form

$$\begin{array}{lll} (P) & \text{minimize} & f(z) \\ & \text{subject to} & h(z) = 0 \\ & & g(z) \leq 0 \\ & & z \in C, \end{array}$$

where $f : R^n \to R,\ h : R^n \to R^l,\ g : R^n \to R^m,\ C \subset R^n$. We assume that C is a closed subset of R^n, each funtion f, h, g is Lipschitz continuous near any given point of C and $h(z) = (h_1(z), \cdots, h_l(z)), g(z) = (g_1(z), \cdots, g_m(z))$. We allow l or $m = 0$ to signify the case in which there are no explicite equality or inequality constraints. In these cases it is clear that they can be ignaed.

Imbed (P) in a parametrized family $P(u, v)$ of mathematical programming problems:

$$\begin{array}{lll} P(u,v) & \min & f(z) \\ & \text{s. t.} & h(z) + u = 0 \\ & & g(z) + v \leq 0 \\ & & z \in C. \end{array}$$

Hypothesis (H) Suppose that the solution to problem (P) exists and the sets $\{z : f(z) \leq r, |h(z)| \leq 1, g(z) \leq 1, z \in C\}$ are compact for all $r \in R$.

Define the *value function* of Problem (P) $V : R^l \times R^m \to R \cup \{+\infty\}$ by

$$V(u,v) := \inf\{f(z) : h(z) + u = 0, g(z) + v \leq 0, z \in C\}.$$

Note that Hypothesis (H) implies that the minimum in the above expression is attained finitely, unless the feasible set for $P(u,v)$ is empty, in which case $V(u,v) = +\infty$.

Let Σ denote the set of all solutions to Problem (P).

Let z be feasible for (P). The set of *normal multipliers* corresponding to z is

$$M^1(z) := \{(r,s) : 0 \in \partial f(z) + \partial\langle s, h(z)\rangle + \partial\langle r, g(z)\rangle + N_C(z), r \geq 0, \langle r, g(z)\rangle = 0\}$$

and the set of *abnormal multipliers* corresponding to z is

$$M^0(z) := \{(r,s) : 0 \in \partial\langle s, h(z)\rangle + \partial\langle r, g(z)\rangle + N_C(z), r \geq 0, \langle r, g(z)\rangle = 0\}.$$

The following theorem which is buried in [Theorem 6.5.2, Clarke [5]] gives an estimate of $\hat{\partial} V(0,0)$ and $\hat{\partial}^\infty V(0,0)$ in terms of Lagrange multipliers.

Theorem 2.1 *Under Hypothesis* (H), V *is lower semicontinuous and one has*

$$\hat{\partial} V(0,0) \subset \cup_{z\in\Sigma} M^1(z) \tag{3}$$

$$\hat{\partial}^\infty V(0,0) \subset \cup_{z\in\Sigma} M^0(z). \tag{4}$$

Proof. In the proof of [Theorem 6.5.2, Clarke [5]], if we replace the generalized gradient ∂ by the limiting subgradient $\hat{\partial}$, the argument goes through without modification and we have the above conlusion. ∎

Similarly, define $W(u) := \inf\{f(z) : h(z) + u = 0, g(z) \leq 0, z \in C\}$. The methods and results of [Theorem 6.52., Clarke [5]] can be used to give the following result.

Theorem 2.2 *Under Hypothese* (H), W *is lower semicontinuous and one has*

$$\hat{\partial} W(0) \subset \cup_{z\in\Sigma}\{s : 0 \in \partial f(z) + \partial\langle s, h(z)\rangle + \partial\langle r, g(z)\rangle + N_C(z), \quad r \geq 0, \langle r, g(z)\rangle = 0\}$$

$$\hat{\partial}^\infty W(0) \subset \cup_{z\in\Sigma}\{s : 0 \in \partial\langle s, h(z)\rangle + \partial\langle r, g(z)\rangle + N_C(z), \quad r \geq 0, \langle r, g(z)\rangle = 0\}.$$

For convenience, we now state a generalized Lagrange multiplier rule [Theorem 6.1.1 of Clarke [5]] for Problem (P).

Theorem 2.3 (Fritz-John type Lagrange multiplier rule) *Let $\bar{z}$ solve (P). Then there exist $\lambda \in \{0,1\}$, $r \geq 0$ and s not all zero, such that*

$$0 \in \lambda \partial f(\bar{z}) + \sum_i s_i \partial h_i(\bar{z}) + \sum_j r_j \partial g_j(\bar{z}) + N_C(\bar{z})$$

$$\langle r, g(\bar{z}) \rangle = 0.$$

Theorem 2.3 however does not ensure the existence of a normal multiplier. It is important in practices to find the condition under which the multiplier rule holds in a normal form since a multiplier rule in an abnormal form may not provide useful information regarding optimality due to the fact that the function being minimized is not involved. Various constraint qualifications have been proposed in the literature under which the multiplier rule holds in a normal form (see Bazaraa et al. [2] and Peterson [10] for the smooth case and Hiriart-Urruty [8] for the nonsmooth case). The following constraint qualification called *calmness condition*, first introdued by Clarke and Rockafellar lays claim to be the weakest constraint qualification.

Definition 2.1 Let $\bar{z}$ solve (P). The problem (P) is calm at $\bar{z}$ provided that there exist positive δ and M such that, for all (u,v) in δB, for all $z \in \bar{z} + \delta B$ which are feasible for $P(u,v)$, one has

$$f(z) - f(\bar{z}) + M\|(u,v)\| \geq 0.$$

The following theorem [Proposition 6.4.4, Clarke [5]] gives a Lagrange multiplier rule in a normal form.

Theorem 2.4 (Kuhn-Tucker type Lagrange multiplier rule) *Let $\bar{z}$ solves (P). If either $M^0(\bar{z}) = \{0\}$ or (P) is calm at $\bar{z}$, then the conclusion of Theorem 2.3 holds with $\lambda = 1$.*

Remark 2.1 The constraint qualification of the type $M^0(\bar{z}) = \{0\}$ is in fact a generalization of the well-known linear independence, Mangasarian-Fromowitz and Slater conditions since it is obvious that linear independence of $\nabla h_i(\bar{z}), \nabla g_j(\bar{z}), i = 1, \cdots, l, j \in \{1, \cdots, m\}, g_j(\bar{z}) = 0$ implies $M^0(\bar{z}) = \{0\}$ and it was shown in Clarke [5] that if either the Mangasarian-Fromowitz or the Slater condition holds, then $M^0(\bar{z}) = \{0\}$. Moreover, it was shown (*c.f.* [Corollary 5 of Clarke [5]]) that if $M^0(\bar{z}) = \{0\}$, then (P) is calm at $\bar{z}$. Therefore, the calmness condition is weaker than the constraint qualification $M^0(\bar{z}) = \{0\}$. In fact it is strictly weaker since there are situations where $M^0(\bar{z}) \neq \{0\}$ but problem (P) is calm at $\bar{z}$. It will be demonstrated later that bilevel programming problems belongs to this category.

3 Differentiability of the value function of the lower level problem

In this section, we study the (generalized) differentiability of the value function for the lower level mathematical progamming problem.

Thoughout this paper, suppose the following assumptions hold for BLPP (equivalently for SLPP):

(A1) F, f, g are continuous differentiable.

(A2) The solution to the BLPP are in the interior of $X \times Y$.

Since $V(x)$ is in general nondifferentiable regardless of assumptions made on g, SLPP is in general a problem with nonsmooth problem data even though the original BLPP has smooth problem data. Since the nonsmoothness of SLPP only appears in $V(x)$ under assumption (A1), we now discuss the (general) differentiablity of $V(x)$. To apply the multiplier rules (Theorem 2.2 and Theorem 2.3), we must find conditions under which $V(x)$ is locally Lipschitz continuous. Thanks to Theorem 2.2, we have the following result.

Let y be feasible for (P_x), i.e. $y \in Y$ and $g(x,y) \leq 0$. Let $\lambda \in \{0,1\}$. The index λ multiplier set corresponding to y is

$$M_x^{\lambda}(y) = \{\pi : \; 0 = \lambda \nabla_y f(x,y) + \nabla_y g(x,y)^T \pi, \; \pi \geq 0, \langle \pi, g(x,y) \rangle = 0\}.$$

Proposition 3.1 Suppose $\Sigma_x \neq \emptyset$ and all solutions to (P_x) lie in the interior of Y. If $M_x^0(\Sigma_x) = \{0\}$, then $V(x)$ is Lipschitz continuous near x and

$$\hat{\partial} V(x) \subset \{\nabla_x f(x,y) + \nabla_x g(x,y)^T \pi : \text{ for some } y \in \Sigma_x \text{ and } \pi \in M_x^1(y)\},$$

where $M_x^0(\Sigma_x) := \cup_{y \in \Sigma_x} M_x^0(y)$.

Proof. Using an observation due to Rockafellar, we can rewrite (P_x) as follows:

$$\begin{array}{ll} \min & f(\alpha, y) \\ \text{s.t.} & g(\alpha, y) \leq 0 \\ & \alpha - x = 0 \\ & y \in Y. \end{array}$$

Then we have

$$V(x) = \min_{(\alpha,y)\in X\times Y}\{f(\alpha,y) : g(\alpha,y) \leq 0, \alpha - x = 0\}.$$

Application of Theorem 2.2 and Proposition 2.1 to the above problem leads to the following estimates:

$$\begin{aligned}
\hat{\partial} V(x) \quad &\subset \quad \cup_{y\in\Sigma_x}\{s : 0 = \nabla_x f(x,y) + \nabla_x g(x,y)^\top \pi - s, \\
&\qquad\qquad 0 = \nabla_y f(x,y) + \nabla_y g(x,y)^\top \pi \\
&\qquad\qquad\qquad \pi \geq 0, \langle \pi, g(x,y)\rangle = 0\} \qquad (5) \\
\hat{\partial}^\infty V(x) \quad &\subset \quad \cup_{y\in\Sigma_x}\{s : 0 = \nabla_x g(x,y)^\top \pi - s, \\
&\qquad\qquad 0 = \nabla_y g(x,y)^\top \pi \\
&\qquad\qquad\qquad \pi \geq 0, \langle \pi, g(x,y)\rangle = 0\}. \qquad (6)
\end{aligned}$$

It is easy to see that the assumption $M_x^0(\Sigma_x) = \{0\}$ implies that the right hand side of (6) is equal to $\{0\}$. Hence $\hat{\partial}^\infty V(x) = \{0\}$. Therefore by virtue of (b) of Proposition 2.1, $V(x)$ is Lipschitz continuous near x. By rearranging the right hand side of (5), we have the desired inclusion. ■

Remark 3.1 If the lower level problem (P_x) has a unique solution $(\bar{y})$ and a unique optimal multiplier π at $\bar{y}$ i.e., $\Sigma_x = \{\bar{y}\}$ and $M_x^1(\bar{y}) = \{\pi\}$, then $V(x)$ is C^1 and

$$\nabla V(x) = \nabla_x f(x,\bar{y}) + \nabla_x g(x,\bar{y})^\top \pi.$$

If the solutions of (P_x) lie in the interior of its constraint set, then

$$\hat{\partial} V(x) \subset \{\nabla_x f(x,y), \text{some } y \in \Sigma_x\}.$$

4 Necessary optimality conditions for BLPP

In this section, we give generalized Lagrange rules for BLPP and study the issue of constraint qualification.

The following Fritz-John type optimality conditions can be obtained readily from the nonsmooth Lagrange multiplier rule (Theorem 2.3), the estimates for the limiting subgradients of the value function $V(x)$ (Proposition 3.1) and the fact that $\partial V(x) = \text{co}\hat{\partial} V(x)$.

Theorem 4.1 *Let (x, y) solves* BLPP. *Suppose that $M_x^0(\Sigma_x) = \{0\}$. Then there exists $\lambda \in \{0, 1\}$, $r \geq 0$, $s \geq 0$ not all zero, integers I, J, $\lambda_{ij} \geq 0$, $\sum_{i=1}^{I} \sum_{j=1}^{J} \lambda_{ij} = 1$, $y_i \in \Sigma_x, \pi_{ij} \geq 0$ such that*

$$0 = \lambda \nabla_x F(x, y) + \nabla_x g(x, y)^\top r + s[\nabla_x f(x, y) - \sum_{ij} \lambda_{ij}(\nabla_x f(x, y_i) + \nabla_x g(x, y_i)^\top \pi_{ij})]$$

$$0 = \lambda \nabla_y F(x, y) + \nabla_y g(x, y)^\top r + s \nabla_y f(x, y)$$

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

$$0 = \nabla_y f(x, y_i) + \nabla_y g(x, y_i)^\top \pi_{ij}$$

$$\langle \pi_{ij}, g(x, y_i) \rangle = 0.$$

It should be noted that $s \geq 0$ results from the fact the feasible solutions of SLPP is the same as those satisfying the constraints with $f(x, y) - V(x) \leq 0$ instead of $f(x, y) - V(x) = 0$.

Now we focus our attention on the problem of constraint qualification. We first show that the usual constraint qualifications, i.e. the linear independence, Mangasarian-Fromowitz, and Slater conditions do not hold in general for SLPP.

1. The linear independence condition and its nonsmooth extension.
Since all data of SLPP are differentiable in y, the linear independence condition or its nonsmooth extension require that the vectors

$$\nabla_y f(x, y), \ \nabla_y g_i(x, y), i \in \bar{I},$$

where $\bar{I} = \{i | g_i(x, y) = 0, i = 1, ..., m\}$ be independent. On the other hand, $M_x^0(\Sigma_x) = \{0\}$ implies that $M_x^1(\Sigma_x) \neq \emptyset$ (see the proof of Proposition 3.1), i.e. there exist $r_i \geq 0, i \in \bar{I}$ such that

$$\nabla_y f(x, y) + \sum_{i \in \bar{I}} r_i \nabla_y g_i(x, y) = 0.$$

This is a contradiction. Therefore, the linear independence condition and any of its nonsmooth extension do not hold for SLPP.

2. The Magasarian-Fromowitz condition and its nonsmooth extension.
Since all data of SLPP are differentiable in y, the Magasarian Fromowitz condition and its nonsmooth extension would imply that

(a) $\nabla_y f(x,y) \neq 0$

(b) There exists a vector $v \in R^n$ such that

$$\langle \nabla_y f(x,y), v \rangle = 0$$
$$\langle \nabla_y g_i(x,y), v \rangle < 0, \text{if } i \in \bar{I}$$

On the other hand since y is an optimal solution of (P_x), there exist $r \geq 0$ such that

$$\nabla_y f(x,y) + \sum_{i \in \bar{I}} r_i \nabla_y g_i(x,y) = 0.$$

If all $r_i = 0$, this contradicts (a). If there is a $j \in \bar{I}$ such that $r_j \neq 0$, by taking the inner product with v of both sides in the equation above, a contradiction to (b) results. So the Magasarian Fromowitz condition and any of its nonsmooth extension will not hold for SLPP.

3. The Slater condition

The Slater condition does not apply to SLPP due to the existence of the equality constraint $f(x,y) - V(x) = 0$.

In fact, the following Proposition demonstrates that any constraint qualification which implies $M^0(\Sigma) = \{0\}$ does not hold for SLPP.

Proposition 4.1 Suppose that $M_x^0(\Sigma_x) = \{0\}$, then $M^0(\Sigma)$, the abnormal multiplier set of SLPP, contains nonzero element.

Proof. $\forall (x,y) \in \Sigma$, y is a solution of P_x, i.e. $y \in \Sigma_x$. Since $M_x^0(\Sigma_x) = \{0\}$, there exist $\pi \geq 0$, $\pi \in M_x^1(y)$ such that

$$0 = \nabla_y f(x,y) + \nabla_y g(x,y)^\top \pi$$
$$\langle \pi, g(x,y) \rangle = 0.$$

Since $\nabla_x f(x,y) + \nabla_x g(x,y)^\top \pi \in \hat{\partial} V(x) \subset \partial V(x)$ by Proposition 3.1, we have

$$\begin{aligned} 0 &\in \nabla g(x,y)^\top \pi + (\nabla f(x,y) - \partial V(x)) \\ 0 &= \langle \pi, g(x,y) \rangle, \pi \geq 0. \end{aligned}$$

That is $(\pi, 1) \in M^0(x,y)$, therefore $(M^0(x,y) \backslash \{0\}) \neq \emptyset$. ∎

The discussion above indicates the difficulty with the constraint qualification for SLPP. We now state the Kuhn-Tucker type necessary optimality condition for SLPP. It is a direct consequence of Theorem 2.4 and Proposition 3.1.

Theorem 4.2 *Let* (x, y) *be a solution of* SLPP. *Suppose that* $M_x^0(\Sigma_x) = \{0\}$. *If* SLPP *is calm at* (x, y)*, then the conclusion of Theorem 4.1 holds with* $\lambda = 1$.

The calmness condition for SLPP is defined in terms of the following *fully perturbed* problem

$$
\begin{array}{llll}
P_{uv} & \min & F(x,y) & \\
& \text{s.t.} & f(x,y) - V(x) + u = 0 & (7) \\
& & g(x,y) + v \leq 0 & (8) \\
& & x \in X,\ y \in Y, &
\end{array}
$$

and it is known (*c.f.* Clarke [5]) that the concept of calmness is closedly related to a numerical technique called "exact penalization". Since it is not a easy task to check a problem is calm, we now search for certain weaker constraint qualification. Since the calmness condition is related to the exact penalization for both constraints (7) and (8) and it is clear from the above discussion that the infeasibility of $M^0(x, y) = \{0\}$ results from the very constraint (7), we may consider the following partially perturbed problem:

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

Definition 4.1 Let (x, y) solves SLPP. SLPP is partially calm at (x, y) provided that there exist $\delta > 0$ and $M > 0$ such that, for all u in δB, for all $(x', y') \in (x, y) + \delta B$ which are feasible for P_u , we have

$$F(x', y') - F(x, y) + M|u| \geq 0.$$

It is easy to see that the partial calmness condition is weaker than the calmness condition and therefore easier to check.

The concept of partial calmness is actually equivalent to the "exact penalization" as shown in the following proposition.

Proposition 4.2 Suppose (x, y) solves SLPP. Then SLPP is partially calm at (x, y) if only if (x, y) is a local optimal solution to the following penalized problem

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

where $M > 0$.

Proof. The conclusion follows easily from the fact that $f(x, y) - V(x) \geq 0$. ∎

Since the penalized problem $\widetilde{SLPP}$ does not have the troublesome constraint $f(x, y) - V(x) = 0$, we may assume the constraint qualification of the type $M^0(\Sigma) = \{0\}$ holds and obtain from Theorem 2.4 the following Kuhn-Tucker condition for BLPP.

Let (x, y) be a solution to the penalized problem $\widetilde{SLPP}$, the abnormal mulitiplier set corresponding to (x, y) is

$$\widetilde{M}^0(x,y) = \{r \in R^m : r \geq 0, \nabla g(x,y)^T r = 0, \ \langle r, g(x,y)\rangle = 0\}.$$

Theorem 4.3 *Let (x, y) solves* BLPP *and* SLPP *is partially calm at (x, y). Suppose that $M_x^0(\Sigma_x) = \{0\}$. If $\widetilde{M}^0(x, y) = \{0\}$. Then the conclusion of Theorem 4.1 holds with $\lambda = 1$ and $s > 0$.*

Remark 4.1 If the Slater condition holds for the penalized problem, conditions $M_x^0(\Sigma_x) = \{0\}$ and $\widetilde{M}^0(x, y) = \{0\}$ are both satisfied automatically.

5 An illustrative example

In this section we give an example for which the partial calmness condition is satisfied.

Consider the problem

$$
\begin{array}{llll}
(P_1) & \min\limits_{-1 \le x \le 1} & x + y(x) & \text{where } y(x) \text{ solves} \\
 & \min\limits_{-1 \le y \le 1} & -2xy + y^2. &
\end{array}
$$

The solution of the lower level problem is

$$
\Sigma_x = \begin{cases} \{x\} & \text{if } -1 \le x \le 1 \\ \{1\} & \text{if } x \geq 1 \\ \{-1\} & \text{if } x \le -1. \end{cases}
$$

Its value function is

$$V(x) = \begin{cases} -x^2 & \text{if } -1 \le x \le 1 \\ (1-x)^2 - x^2 & \text{if } x \ge 1 \\ (1+x)^2 - x^2 & \text{if } x \le -1. \end{cases}$$

The equivalent formulation SLPP is

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

It is obviously that $(x, y) = (-1, -1)$ is an optimal solution.

Now we consider the penalized problem

$$\begin{aligned} (\tilde{P}_1) \qquad \min \quad & x + y + M(x-y)^2 \\ \text{s.t.} \quad & -1 \le x \le 1 \\ & -1 \le y \le 1 \end{aligned}$$

where $M > 0$.

Since the above problem is convex, the necessary and sufficient optimality condition for the above problem is the existence of $\alpha \ge 0, \beta \ge 0, \gamma \ge 0, \delta \ge 0$ such that

$$\begin{aligned} & 1 + 2k(x-y) + \alpha - \beta = 0 \\ & 1 + 2k(y-x) + \gamma - \delta = 0 \\ & \alpha(x-1) = 0 \\ & \beta(-x-1) = 0 \\ & \gamma(y-1) = 0 \\ & \delta(-y-1) = 0. \end{aligned}$$

$x = -1,\ y = -1,\ \gamma = \alpha = 1,\ \beta = \delta = 2$ is a solution of this system. Hence the above problem is partially calm at $(x, y) = (-1, -1)$.

6 Optimality conditions for the minmax problem and the linear bilevel programming problem

6.1 Minmax problem

The well studied minmax problem (*c.f.* Shimizu and Aiyoshi [13] and Schmitendorf [12]) is defined as follows

$$(P_2) \qquad \min_{x \in X} \quad \max_{y \in Y}\{\phi(x,y) : g(x,y) \leq 0\}$$
$$\text{s.t.} \quad g(x,y) \leq 0.$$

Assume that $\phi(x,y) : R^n \to R, g(x,y) : R^n \to R^m$ are C^1, the solutions to

$$\max_{y \in Y}\{\phi(x,y) : g(x,y) \leq 0\}$$

exist and the solutions of (P_2) lie in the interior of $X \times Y$. As it was mentioned in section 1, the minmax problem is a special case of BLPP with $-f(x,y) = F(x,y) = \phi(x,y)$. Let $V(x) := \min\{-\phi(x,y) : g(x,y) \leq 0,\ y \in Y\}$. Then the minmax problem is equivalent to the following problem

$$\begin{aligned} (P_2) \qquad \min \quad & -V(x) \\ \text{s.t.} \quad & g(x,y) \leq 0 \\ & x \in X, y \in Y \end{aligned}$$

which can be rewritten as

$$\begin{aligned} (\tilde{P}_2) \qquad \min \quad & F(x,y) + (f(x,y) - V(x)) \\ \text{s.t.} \quad & g(x,y) \leq 0 \\ & x \in X, y \in Y. \end{aligned}$$

Therefore, the minmax problem is partially calm at any solution (x,y).

Theorem 4.3 for the minmax problem can be stated as follows:

Theorem 6.1 *Let (x,y) solves the minmax problem. Suppose that $M_x^0(\Sigma_x) = \{0\}$ and $M^0(x,y) = \{0\}$. Then there exists $r \geq 0$, integers I, J, $\lambda_{ij} \geq 0$, $\sum_{i=1}^{I}\sum_{j=1}^{J}\lambda_{ij} = 1$, $y_i \in \Sigma_x, \pi_{ij} \geq 0$ such that*

$$0 = \nabla_x g(x,y)^\top r + \sum_{ij}\lambda_{ij}(\nabla_x\phi(x,y_i) - \nabla_x g(x,y_i)^\top \pi_{ij})$$

$$0 = \nabla_y g(x, y)^\top r$$

$$0 = \nabla_y \phi(x, y_i) - \nabla_y g(x, y_i)^\top \pi_{ij}$$

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

$$\langle \pi_{ij}, g(x, y_i) \rangle = 0.$$

6.2 The linear bilevel programming problem

The linear bilevel programming problem is formulated as follows.

$$\begin{array}{lll} \text{LBLPP} & \min\limits_{x \in R^{n_1}} & F(x, y(x)) = ax + by(x) \text{ where } y(x) \text{ solves} \\ & \min\limits_{y \in R^{n_2}} & f(x, y) = cx + dy \\ & \text{s.t.} & g(x, y) = Cx + Dy - q \leq 0, \end{array}$$

where $a \in R^{n_1}, b \in R^{n_2}$, $C \in R^{m \times n_1}, D \in R^{m \times n_2}$, $q \in R^m$.

We make the following usual assumptions: the set

$$S = \{(x, y) \in R^n : Cx + Dy \leq q\}$$

is bounded and has nonempty interior.

The equivalent formulation of problem LBLPP is

$$\begin{array}{lll} \text{SLBLPP} & \min & ax + by \\ & \text{s.t.} & cx + dy - V(x) = 0 \\ & & Cx + Dy - q \leq 0 \\ & & x \in R^{n_1},\ y \in R^{n_2}, \end{array}$$

where $V(x) := \min\{cx + dy : Cx + Dy - q \leq 0, x \in R^{n_1},\ y \in R^{n_2}\}$.

Proposition 6.1 Let (x, y) solves the LBLPP, then SLBLPP is partially calm at (x, y).

Proof.

Step 1. First we prove that for sufficiently large $M > 0$, any solution of the following problem

$$\begin{array}{lll} SL\widetilde{BL}PP & \min & F(x, y) + M(f(x, y) - V(x)) \\ & \text{s.t.} & g(x, y) \leq 0 \\ & & x \in R^{n_1}, y \in R^{n_2} \end{array}$$

must satisfy $f(x, y) = V(x)$.

Indeed, suppose (x^*, y^*) is a solution of the original problem LBLPP, then $f(x^*, y^*) = V(x^*)$. Therefore, for any solution (x, y) to problem $SL\widetilde{BL}PP$, we have

$$F(x,y) + M(f(x,y) - V(x)) \leq F(x^*, y^*).$$

Since S is bounded, we have

$$f(x,y) - V(x) \leq \frac{F(x^*, y^*) - F(x,y)}{M} \leq \frac{K}{M}$$

where K is some positive constant. Thus as $M \to \infty$, we have $f(x, y) - V(x) \to 0$. However, since the function $F(x, y) + M(f(x, y) - V(x))$ is concave in (x, y) , the solutions of problem $SL\widetilde{BL}PP$ occur at the extreme points of polytope S. Hence, for some large finite value of M, say M^*, we have $f(x, y) = V(x)$.

Step 2. We now show that a solution to the original problem LBLPP is a solution of the penalized problem $SL\widetilde{BL}PP$ for sufficiently large $M > 0$. Indeed, if a solution of the original problem LBLPP (x^*, y^*) is not a solution of the penalized problem for any $M > 0$. Then for $M > 0$, there exists a solution to the penalized problem (x_M, y_M) such that

$$F(x_M, y_M) + M(f(x_M, y_M) - V(x_M)) < F(x^*, y^*)$$

$$g(x_M, y_M) \leq 0$$

$$x_M \in R^{n_1}, \; y_M \in R^{n_2}.$$

From step 1, for a sufficiently large $\bar{M} > 0$ we have $f(x_{\bar{M}}, y_{\bar{M}}) = V(x_{\bar{M}})$ thus $F(x_{\bar{M}}, y_{\bar{M}}) < F(x^*, y^*)$. This contradicts with the fact that (x^*, y^*) is a solution of the original problem LBLPP. Thus problem LBLPP is partially calm. ■

We now present the Kuhn-Tucker necessary optimality condition (Theorem 4.2) for the LBLPP in the following form.

Theorem 6.2 *Let (x, y) be a solution to* LBLPP. *Then there exist $r \geq 0, M > 0$, integers I, J, $\lambda_{ij} \geq 0, \sum_{i=1}^{I} \sum_{j=1}^{J} \lambda_{ij} = 1$, $\pi_{ij} \geq 0$ $y_i \in \Sigma_x$ such that*

$$0 = a^\top + C^\top r - MC^\top \sum_{ij} \lambda_{ij} \pi_{ij}$$

$$0 = b^\top + D^\top r + Md^\top$$

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

$$0 = d^\top + D^\top \pi_{ij}$$

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

The following conterexample was given by Clark and Westerberg [4] to show that the necessary optimality conditions proposed by Bard [1] were not valid.

$$\begin{aligned} \min_{x \in R} \quad & x - 4y(x) \qquad \text{where } y(x) \text{ solves} \\ \min_{y \in R} \quad & y \\ \text{s.t.} \quad & -2x + y \leq 0 \\ & 2x + 5y - 108 \leq 0 \\ & 2x - 3y + 4 \leq 0 \end{aligned}$$

This problem is a linear bilevel programming problem, the optimal solution is $(19, 14)$. According to Theorem 6.2, at $(19, 14)$, there must exist $r \geq 0, \pi \geq 0$ such that

$$\begin{aligned} & 0 = 1 - 2r_1 + 2r_2 + 2r_3 - M[-2\pi_1 + 2\pi_2 + 2\pi_3] \\ & 0 = -4 + r_1 + 5r_2 - 3r_3 + M \\ & 0 = 1 + \pi_1 + 5\pi_2 - 3\pi_3 \\ & r_1 = 0, \ r_2 \geq 0, \ r_3 \geq 0, \ M > 0 \\ & \pi_1 = 0, \pi_2 \geq 0, \pi_3 \geq 0. \end{aligned}$$

The above system is indeed satisfied by $(x, y) = (19, 14)$, $M = 1/2$, $(r_1, r_2, r_3) = (0, 13/16, 3/16)$ and $(\pi_1, \pi_2, \pi_3) = (0, 1, 2)$.

Bard's necessary condition for the same problem is as follows

$$\begin{aligned} & 1 - 2r_1 + 2r_2 + 2r_3 = 0 \\ & -4 + r_1 + 5r_2 - 3r_3 + M = 0 \\ & r_1 = 0, \ r_2 \geq 0, \ r_3 \geq 0, \ M \geq 0. \end{aligned}$$

It is clear that it is impossible to find $r_1 = 0, \ r_2 \geq 0, \ r_3 \geq 0$ satisfying the above system.

7 Sufficient optimality conditions

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

$$
\begin{array}{lll}
(P_3) \quad \min_{x\in X} & F(x,y(x)) & \text{where } y(x) \text{ solves} \\
\min_{y\in Y} & f(x,y). &
\end{array}
$$

Note that (P_3) is a BLPP without inequality constraint $g(x,y) \leq 0$.

In addition to assumptions (A1) and (A2), we assume that F and f are C^2.

Theorem 7.1 *Let (x^*, y^*) be feasible for (P_3). If there exist $s \geq 0$, $\bar{y} \in \Sigma_{x^*}$ such that*

$$
\begin{aligned}
0 &= \nabla_x F(x^*,y^*) + s(\nabla_x f(x^*,y^*) - \nabla_x f(x^*,\bar{y})) && (9)\\
0 &= \nabla_y F(x^*,y^*) + s\nabla_y f(x^*,y^*) && (10)
\end{aligned}
$$

and the matrix

$$
\frac{\partial^2}{\partial(x,y)^2}[F(x,y) + s(f(x,y) - f(x,\bar{y}))]
$$

is positive semidefinite for any $x \in R^{n_1}, y \in R^{n_2}$, then (x^*, y^*) is optimal solution to (P_3).

Proof. Take any $x^0 \in X$ and $y^0 \in \Sigma_{x^0}$. By Taylor's theorem, we have

$$
\begin{aligned}
& F(x^0,y^0) + s(f(x^0,y^0) - f(x^0,\bar{y})) - F(x^*,y^*) - s(f(x^*,y^*) - f(x^*,\bar{y})) \\
&= [\nabla F(x^*,y^*) - s(\nabla f(x^*,y^*) - \nabla_x f(x^*,\bar{y}) \times \{0\})][(x^0,y^0) - (x^*,y^*)] \\
&\quad + \frac{1}{2}[(x^0,y^0) - (x^*,y^*)]^{-1} \frac{\partial^2}{\partial(x,y)^2}[F(x,y) + s(f(x,y) - f(x,\bar{y}))]_{(x,y)=(\tilde{x},\tilde{y})} \\
&\quad [(x^0,y^0) - (x^*,y^*)]
\end{aligned}
$$

where $(\tilde{x},\tilde{y}) = \rho(x^0,y^0) + (1-\rho)(x^*,y^*)$, $0 \leq \rho \leq 1$. Since the first term on the right hand side of the above equation is zero by virtue of (9) and (10) and the second term on the right hand side is greater and equal to zero by assumption, we have

$$
F(x^0,y^0) + s(f(x^0,y^0) - f(x^0,\bar{y})) \geq F(x^*,y^*) - s(f(x^*,y^*) - f(x^*,\bar{y})) \qquad (11)
$$

By definition of $\bar{y}$, we have

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

Therefore,

$$f(x^*, y^*) - f(x^*, \bar{y}) = 0. \tag{12}$$

Since

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

and $s \geq 0$, we have

$$s(f(x^0, y^0) - f(x^0, \bar{y})) \leq 0. \tag{13}$$

By virtue of (11), (12) and (13), we have

$$F(x^0, y^0) \geq F(x^0, y^0) + s(f(x^0, y^0) - f(x^0, \bar{y})) \geq F(x^*, y^*).$$

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

Theorem 7.2 *Let $x^* \in X$. Suppose that $\Sigma_{x^*} = \{y^*\}$ is a singleton. If*

$$0 = \nabla F(x^*, y^*) \tag{14}$$

and $\frac{\partial^2}{\partial (x,y)^2} F(x^*, y^*)$ is positive definite for (x^*, y^*), then (x^*, y^*) is a strict local minimum point of (P_3).

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,

$$F(x_k, y_k) \leq F(x^*, y^*). \tag{15}$$

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 Σ_{x^*} is a singleton and $y_k \in \Sigma_{x_k}$ 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 convegent to s^*. Now by Taylar's theorem, we have

$$F(x_k, y_k) - F(x^*, y^*) = \nabla F(x^*, y^*) \delta_k s_k + \frac{1}{2} \delta_2 s_k^{-1} \frac{\partial^2}{\partial (x,y)^2} F(\tilde{x}_k, \tilde{y}_k) s_k$$

where $(\tilde{x}_k, \tilde{y}_k) = \rho(x_k, y_k) + (1-\rho)(x^*, y^*), \quad 0 \leq \rho \leq 1$. By the first order condition (14) and the assumption (15), we have

$$\frac{1}{2} \delta_2 s_k^{-1} \frac{\partial^2}{\partial (x,y)^2} F(\tilde{x}_k, \tilde{y}_k) s_k \leq 0$$

which yields a contradiction as $k \to \infty$. Therefore (x^*, y^*) is a strict local minimum point of (P_3) and the proof of the theorem is complete. ■

8 Conclusion and future research

In this paper, we have identified the difficulty in deriving optimality conditions for the BLPP. For the neccessary conditions, we have shown that the usual constraint qualifications do not hold for BLPP and the right constriant qualification is the calmness condition. Moreover, we have shown that the calmness condition is satisfied automatically for the linear bilevel programming problem and the minmax problem. Since the calmness condition is not expressed explicitely in terms of the original problem data, it is highly desirable to search for conditions in terms of the original problem data which will lead to the calmness condition. It is known that the concept of calmness condition is closely related to the exact penality method. Recently Marcotte and Zhu [9] proposed the exact penalty method for general bilevel programming problem by using gap function approach. For the linear bilevel programming the difference $f(x, y) - V(x)$ coincides with the gap function of the lower problem $G(x, y)$ (see Hearn [7] for the definition). If $f(x, y)$ is convex in y for fixed x, the inequality $G(x, y) \geq f(x, y) - V(x)$ always holds. Hence the exact penalty method using the value function implies the exact penalty method using the gap function formulation for the convex case, and it is not true conversely. The search for sufficient conditions of BLPP with constrianted lower level problem is also a subject of the future study.

9 Acknowledgements

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

References

[1] J.F. Bard, "Optimality conditions for the bilevel programming problem", *Naval Research Logistics Quarterly*, **31**(1984), 13-26.

[2] M.S. Bazaraa, J.J. Goode and C.M. Shetty, "Constraint qualifications revisited", *Management Sicence* , **18**(1972), 567-573.

[3] Y. Chen and M. Florian, "The nonlinear bilevel programming problem: a general formulation and optimality conditions" , *Publication CRT-794, Université de Montréal, Canada*, (1991).

[4] P.A. Clark and A.W. Westerberg, A note on the optimality conditions for the bilevel programming problem", *Naval Research Logistics Quarterly*, **35**(1988), 414-418.

[5] F.H. Clarke, *Optimization and Nonsmooth Analysis*, (Wiley-Interscience, New York, 1983).

[6] F.H. Clarke, *Methods of Dynamic and Nonsmooth Optimization*, (SIAM CBMS-NSF, 1989).

[7] D.W. Hearn, "The gap function of a convex program", *Operations Research Letters* **1**(1981), 67-71.

[8] J.B. Hiriart-Urruty, "Refinements of necessary optimality conditions in nondifferentiable programming I" , *Appl. Math. Optim.*, **5**(1979), 63-82.

[9] P. Marcotte and D. Zhu, "The penalty function method and exact penalty method for generalized bi-level mathematical programming", Preprint, (1992)

[10] D.W. Peterson, "A review of constraint qualifications in finite-dimensional spaces", *SIAM Review*, **15**(1973), 639-655.

[11] R.T. Rockafellar, "Extensions of subgradient calculus with applications to optimization", *Nonlinear Analysis, Theory, Methods and Applications*, **9**(1985), 665-698.

[12] W.E. Schmitendorf, "Necessary conditions and sufficient conditions for static minmax problems", *Journal of Mathematical Analysis and Applications*, **57**(1977), 683-693.

[13] K. Shimizu and E. Aiyoshi, "Necessary conditions for minmax problems and algorithms by relaxation procedure" , *IEEE Trans. Automat. Contr.*, **AC-25**(1980), 62-66.