

***GENERALIZED BILEVEL PROGRAMMING
PROBLEMS***

J.J. YE, D.L. ZHU & Q. ZHU

DMS-646-IR

September 1993

Generalized Bilevel Programming Problems

J.J. Ye*, D.L. Zhu and Q. Zhu†

September 8, 1993

Abstract

The generalized bilevel programming problem (GBLP) is a mathematical programming problem with variational inequality constraints. In this paper by error bound approach, we establish some equivalent single level formulation of the generalized bilevel programming problem. The necessary optimality condition of Kuhn-Tucker type are then given by using nonsmooth analysis.

Key words: Generalized bilevel programming problems, variational inequality, error bound, necessary optimality condition, nonsmooth analysis.

AMS(MOS) subject classification:49K99, 90C,90D65

*Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada V8W 3P4. This work was supported in part by NSERC under grant WFA 0123160

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

1 Introduction

In this paper we consider a *mathematical programming problem with variational inequality constraints* defined as follows:

$$\text{minimize } f(x, y) \quad \text{subject to } x \in X \text{ and } y \in S(x) \quad (1)$$

where $f : R^{n+m} \rightarrow R$, X is a nonempty and closed subset of R^n , and for each $x \in X$ $S(x)$ is the *solution set of a variational inequality parameterized in x* , i.e.

$$S(x) = \{y \in U(x) : \langle F(x, y), y - z \rangle \leq 0 \quad \forall z \in U(x)\}$$

where $U : X \rightarrow R^m$ is a set-valued mapping and $F : R^{n+m} \rightarrow R^m$ is a point-to-point mapping.

One can interpret the above problem by a *hierarchical decision process* where there are two decision makers and the upper level decision maker always have the first choice as follows: Given a decision vector x by the upper level decision maker (hereafter the *leader*), $S(x)$ can be considered to be the lower level decision maker's (hereafter the *follower*) decision set, i.e. the set of decision vectors that the follower agrees to use. Suppose that the game is *co-operative*, i.e. if the follower's decision set $S(x)$ is not a singleton, the follower will allow the leader to choose which of them is actually used. Having the complete knowledge of the possible reaction by the follower, the leader then select the decision vector $x \in X$ and $y \in S(x)$ to minimize his objective function $f(x, y)$.

If $F(x, y) = \nabla_y g(x, y)$, where $g : R^{n+m} \rightarrow R$ and if $U(x)$ is convex, then the variational inequality parameterized in x ,

$$\langle F(x, y), y - z \rangle \leq 0 \quad \forall z \in U(x) \quad (2)$$

is just a restatement of the first-order necessary conditions of optimality for the following optimization problem parameterized in x :

$$\text{minimize } g(x, y) \quad \text{subject to } y \in U(x). \quad (3)$$

Furthermore, if $g(x, y)$ is pseudo-convex in y (i.e., $\langle \nabla_y g(x, y), y - z \rangle \leq 0$ implies $g(x, y) \leq g(x, z)$ for all $y, z \in U(x)$), then a vector $y \in U(x)$ is a solution to (2) if and only if it is a global optimal solution of (3). In this case, the mathematical programming problem with variational inequality constraints (1) is the *classical bilevel*

programming problem, or so called Stackelberg game (see e.g. [1]). The above connection between the variational inequality and optimization problem breaks down if F is not a gradient mapping of a certain function in y . Since problem (1) includes problems which are not the classical bilevel programming problem, we shall call problem (1) the *generalized bilevel programming problem* (GBLP).

To be more specific, in this paper we will assume that

$$U(x) = \{y \in R^m \mid c(x, y) \leq 0, y \in Y\}$$

where Y is a nonempty and closed subset of R^m , and $c : R^{n+m} \rightarrow R^d$ is a point-to-point mapping. We will assume throughout the paper that f, c, F are continuous and that

$$\lim_{x \in X, y \in S(x), \|(x, y)\| \rightarrow \infty} f(x, y) = \infty.$$

Under these assumptions, problem GBLP is well defined and possesses at least one solution.

Reducing the (generalized or classical) bilevel optimization problem to a single level constrained optimization problem is an important strategy in studying the problem. There are various ways to implement this strategy, e.g., value function approach in [9], gap function approach in [8]. In this paper we will use the error bound of the lower level variable to the lower level solution set as a tool to establish some equivalent single level formulations of the bilevel problem. The necessary optimality conditions of Kuhn-Tucker type are also given.

The organization of the paper is as follows. In the next section we introduce the notion of error bounds and show that error bounds provide local exact penalty formulations of GBLP. In section 3, we show that the local exact penalty representations of error bounds are also global provided Lipschitz continuity and error bound assumptions are satisfied globally. In section 4, some useful error bounds are given. Necessary conditions for optimality of Kuhn-Tucker type for problem GBLP associated with various error bounds are given in section 5. In section 6, relations between various error bounds are given and some examples are given to show that various equivalent single level optimization formulations with error bounds and their corresponding necessary conditions for optimality complement with each other.

2 Partial calmness and local exact penalty representation

In this section we introduce the notion of error bounds and show that error bounds provide local exact penalty formulations of GBLP.

For any $x \in X, y \in Y$, denote $d_S(x, y) = \text{dist}(y, S(x)) := \inf\{\|y - z\| : z \in S(x)\}$ the distance from y to $S(x)$. Then the generalized bilevel programming problem GBLP is equivalent to the *mathematical programming problem involving a distance function constraint*:

$$\begin{array}{lll} \text{DP} & \text{minimize} & f(x, y) \\ & \text{subject to} & d_S(x, y) = 0 \\ & & c(x, y) \leq 0 \\ & & x \in X, y \in Y. \end{array}$$

Its *perturbed problem* is the following:

$$\begin{array}{lll} \text{DP}(\epsilon) & \text{minimize} & f(x, y) \\ & \text{subject to} & d_S(x, y) = \epsilon \\ & & c(x, y) \leq 0 \\ & & x \in X, y \in Y. \end{array}$$

The following definition is equivalent to that given in [9].

Definition 2.1 [Partial Calmness] Let (x^*, y^*) solve DP. Problem DP is *partially calm* at (x^*, y^*) provided that there exist $\delta > 0$ and $\mu > 0$ such that, for all $\epsilon \in [0, \delta)$ and all $(x, y) \in (x^*, y^*) + \delta B$ which are feasible for DP(ϵ), one has

$$f(x, y) - f(x^*, y^*) + \mu\epsilon \geq 0.$$

The above definition is different from that of [2] (see also Definition 5.1) in that instead of having all equality and inequality constraints perturbed, in our perturbed problem only the constraint $d_S(x, y) = 0$ is perturbed. We are particularly interested in perturbing the constraint $d_S(x, y) = 0$ because it is the essential constraint that reflects the bilevel nature of the problem.

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

Proposition 2.1 *Suppose (x^*, y^*) is a local solution of GBLP, then DP is partially calm at (x^*, y^*) if and only if (x^*, y^*) is a local solution of the following penalized problem:*

$$\begin{aligned} \text{DP}_\mu \quad & \text{minimize} && f(x, y) + \mu d_S(x, y) \\ & \text{subject to} && c(x, y) \leq 0 \\ & && x \in X, y \in Y. \end{aligned}$$

Proof. The conclusion follows easily from the fact that $d_S(x, y) = \epsilon$ for (x, y) which are feasible for $\text{DP}(\epsilon)$ and $d_S(x^*, y^*) = 0$. ■

Remark 2.1 Although Definition 2.1 and Proposition 2.1 are stated for the mathematical programming problem involving a distance function constraint, they can be stated as the same without change for any standard mathematical programming problem. Henceforth we will use Definition 2.1 and Proposition 2.1 for any standard mathematical programming problem.

The following result shows that problem DP is partially calm under local Lipschitz continuity assumption of $f(x, \cdot)$.

Theorem 2.1 *Let (x^*, y^*) be a local solution of problem DP. Assume that f is locally Lipschitz near y^* uniformly in x in a neighborhood of x^* . Then problem DP is partially calm at (x^*, y^*) .*

Proof. Let $\delta > 0$ be such that (x^*, y^*) is a local solution of DP in $(x^*, y^*) + 2\delta B \subset X \times Y$. For any $0 \leq \epsilon < \delta$, let $(x, y) \in (x^*, y^*) + \delta B$ be feasible for DP_ϵ , i.e. $d_S(x, y) = \epsilon$ and $c(x, y) \leq 0$, $(x, y) \in (x^*, y^*) + \delta B$. Since $S(x)$ is closed, one can choose a $y' \in S(x)$ such that $\|y' - y\| = \epsilon$. Since (x, y') is feasible for DP and

$$\begin{aligned} \|(x, y') - (x^*, y^*)\| &\leq \|(x, y') - (x, y)\| + \|(x, y) - (x^*, y^*)\| \\ &\leq \epsilon + \delta < 2\delta, \end{aligned}$$

we have

$$f(x, y') \geq f(x^*, y^*). \tag{4}$$

Local Lipschitz continuity of $f(x, \cdot)$ near y^* implies that,

$$f(x, y) - f(x, y') \geq -L\epsilon. \tag{5}$$

Combining (4) and (5) one has

$$f(x, y) - f(x^*, y^*) + L\epsilon \geq 0$$

i.e., DP is partially calm at (x^*, y^*) . ■

The above theorem shows that problem GBLP formulated in terms of the mathematical programming problem involving a distance function constraint DP is partially calm under very mild condition. Therefore by equivalence to the penalized problem DP_μ (see Proposition 2.1), if functions $f(x, y), d_S(x, y), c(x, y)$ are Lipschitz continuous, theoretically, one may use the generalized Lagrange multiplier rule of Clarke (see Proposition 6.4.4 of [2]) to derive a Kuhn-Tucker condition under appropriate constraint qualifications on constraints $c(x, y) \leq 0, x \in X, y \in Y$. Unfortunately, the Clarke generalized gradient of the distance function $d_S(x, y)$ is very difficult to study and one may not be able to derive a meaningful Kuhn-Tucker condition. On the other hand, using the gap function defined by

$$G_0(x, y) := \max\{\langle F(x, y), y - z \rangle : z \in U(x)\} \quad (6)$$

in place of $d_S(x, y)$ in problem DP as in [8] and using $g(x, y) - V(x)$ where $V(x)$ is the value function defined by

$$V(x) := \inf\{g(x, y) : y \in U(x)\} \quad (7)$$

in place of $d_S(x, y)$ in problem DP for the classical bilevel programming problem as in [9], one could obtain the Lagrange multiplier rule since both gap functions and the value functions are Lipschitz continuous and their Clarke generalized gradients are known to contain in the convex hull of normal Lagrange multipliers under mild conditions. As illustrated in [9], however, for the mathematical programming formulation of bilevel programming problems in terms of the value functions, the usual constraint qualifications (in the sense of Remark 5.1) do not hold. Similarly, one can show that for the mathematical programming formulation of GBLP in terms of gap functions the usual constraint qualifications also do not hold. Therefore unlike in the case of the mathematical programming formulation in terms of the distance function DP, calmness (or partially calmness) conditions do not hold in general for the mathematical programming formulations in terms of either gap function or value function.

To solve this dilemma, we shall use the distance function $d_S(x, y)$ as a tool to establish some equivalent formulations of GBLP which on one hand are partially

calm and on the other hand provide a meaningful Kuhn-Tucker condition. For this purpose, we call a nonnegative function $r(x, y) : X \times Y \rightarrow R^+$ an *error bound* of the inclusion $y \in S(x)$ with modulus $\delta > 0$ in set $Q \subset X \times Y$ provided

$$d_S(x, y) \leq \delta r(x, y) \quad \forall (x, y) \in Q. \quad (8)$$

We are particularly interested in error bounds that have the following property:

$$(A) \quad r(x, y) = 0, y \in Y, c(x, y) \leq 0 \text{ if and only if } y \in S(x).$$

An error bound with Property (A) provides the following equivalent formulation of GBLP:

$$\begin{array}{ll} \text{RP} & \text{minimize} \quad f(x, y) \\ & \text{subject to} \quad r(x, y) = 0 \\ & \quad \quad \quad c(x, y) \leq 0 \\ & \quad \quad \quad x \in X, y \in Y. \end{array}$$

Its penalized problem is:

$$\begin{array}{ll} \text{RP}_\mu & \text{minimize} \quad f(x, y) + \mu r(x, y) \\ & \text{subject to} \quad c(x, y) \leq 0 \\ & \quad \quad \quad x \in X, y \in Y. \end{array}$$

In the following theorem we show that problem RP is partially calm if $r(x, y)$ is an error bound with property (A) and f is locally Lipschitz near y^* uniformly in x .

Theorem 2.2 *Let (x^*, y^*) be a local solution of problem RP. Assume that r is an error bound in a neighborhood $N(x^*, y^*)$ of (x^*, y^*) with Property (A) and that f is locally Lipschitz near y^* uniformly in x in a neighborhood of x^* . Then problem RP is partially calm at (x^*, y^*) .*

Proof. Let $\delta > 0$ be the modulus of error bound $r(x, y)$. (x^*, y^*) being an optimal solution of RP is also a solution of DP. Since DP is partially calm, by Proposition 2.1 there exists $\mu/\delta > 0$ such that (x^*, y^*) is also a solution to $\text{DP}_{\mu/\delta}$. Let $(x, y) \in \text{Gr}U \cap N(x^*, y^*)$, where $\text{Gr}U$ denotes the graph of the set-valued mapping $U(x)$, i.e.,

$$\text{Gr}U := \{(x, y) : x \in X, y \in U(x)\}.$$

Notice that $r(x^*, y^*) = d_S(x^*, y^*) = 0$. We have

$$\begin{aligned} f(x^*, y^*) + \mu \cdot r(x^*, y^*) &= f(x^*, y^*) + \mu/\delta \cdot d_S(x^*, y^*) \\ &\leq f(x, y) + \mu/\delta \cdot d_S(x, y) \quad \text{since } (x^*, y^*) \text{ solves } \text{DP}_{\mu/\delta} \\ &\leq f(x, y) + \mu \cdot r(x, y) \quad \text{by inequality (8)}. \end{aligned}$$

Therefore, (x^*, y^*) is also a local solution of RP_μ , i.e., RP is partially calm at (x^*, y^*) . ■

An error bound with Property (A) also provides the following equivalent formulation of the GBLP:

$$\begin{array}{ll} \text{RSP} & \text{minimize} \quad f(x, y) \\ & \text{subject to} \quad r^2(x, y) = 0 \\ & \quad \quad \quad c(x, y) \leq 0 \\ & \quad \quad \quad x \in X, y \in Y. \end{array}$$

Its penalized problem is

$$\begin{array}{ll} \text{RSP}_\mu & \text{minimize} \quad f(x, y) + \mu r^2(x, y) \\ & \text{subject to} \quad c(x, y) \leq 0 \\ & \quad \quad \quad x \in X, y \in Y. \end{array}$$

To establish the partial calmness condition for problem RSP. Lipschitz continuity of f in y is not enough and one needs the following definition.

Definition 2.2 Let $x_0 \in X$. The mapping $f(x, y) : R^n \times R^m \rightarrow R$ is *upper Hölder continuous with exponent 2 near every $y \in S(x)$ uniformly for x in a neighborhood $N(x_0)$ of x_0* provided there exists $L > 0$ such that

$$f(x, y') - f(x, y) \geq -L \|y' - y\|^2, \quad \forall y' \in N(y), y \in S(x), x \in N(x_0).$$

The constant L is called the modulus.

In the following theorem we show that problem RSP is partially calm if $r(x, y)$ is an error bound with property (A) and f is upper Hölder continuous with exponent 2 near every $y \in S(x)$ uniformly in x in a neighborhood of x^* .

Theorem 2.3 *Let (x^*, y^*) be a local solution of problem RSP. Assume that r is an error bound in a neighborhood of (x^*, y^*) with Property (A) and that f is upper Hölder continuous with exponent 2 and modulus $L > 0$ near every $y \in S(x)$ uniformly in x in a neighborhood of x^* . Then problem RSP is partially calm at (x^*, y^*) .*

Proof. Let $\delta > 0$ be the modulus of the error bound $r(x, y)$. Let $\alpha > 0$ be such that (x^*, y^*) is a local solution of RSP in $(x^*, y^*) + \alpha(\delta + 1)B \subset X \times Y$. For any $0 \leq \epsilon < \alpha$, let $(x, y) \in (x^*, y^*) + \alpha B$ be such that $r^2(x, y) = \epsilon, c(x, y) \leq 0$. Since $S(x)$ is closed, one can choose $y'(x) \in S(x)$ such that $\|y - y'(x)\| = d_S(x, y) \leq \delta r(x, y) = \delta\epsilon$. Since $(x, y'(x))$ is feasible for RSP and

$$\begin{aligned} \|(x, y'(x)) - (x^*, y^*)\| &\leq \|(x, y'(x)) - (x, y)\| + \|(x, y) - (x^*, y^*)\| \\ &\leq \delta\epsilon + \alpha < \alpha(\delta + 1), \end{aligned}$$

we have

$$f(x, y'(x)) \geq f(x^*, y^*).$$

Therefore we have

$$\begin{aligned} f(x, y) - f(x^*, y^*) &\geq f(x, y) - f(x, y'(x)) \quad \text{by optimality of } (x^*, y^*) \\ &\geq -L\|y - y'(x)\|^2 \quad \text{by upper Hölder continuity of } f \\ &= -Ld_S^2(x, y) \\ &\geq -L\delta^2 r^2(x, y) \quad \text{since } r(x, y) \text{ is an error bound} \\ &= -L\delta^2 \epsilon, \end{aligned}$$

i.e., RSP is partially calm at (x^*, y^*) . ■

3 Global exact penalty representation

In this section we show that the local exact penalty representations in Theorem 2.1 and 2.2 are also global provided Lipschitz continuity and error bound assumptions are satisfied globally. First we study the global exact penalty representations for DP.

Theorem 3.1 *Let $(x^*, y^*) \in \text{Gr}U$ and $f(x, \cdot)$ be Lipschitz continuous with constant $L > 0$ uniformly for all $x \in X$. Then for any μ larger than L , (x^*, y^*) is a global solution of DP_μ if and only if it is also a global solution of GBLP.*

Proof. Let $\mu > L$ and $(x(\mu), y(\mu))$ be a global solution of DP_μ . Then

$$f(x(\mu), y(\mu)) + \mu d_S(x(\mu), y(\mu)) \leq f(x(\mu), y) \quad \forall y \in S(x(\mu)).$$

If $y(\mu) \in S(x(\mu))$, by virtue of the above inequality and the Lipschitz continuity of $f(x, \cdot)$, one has

$$\begin{aligned} \mu d_S(x(\mu), y(\mu)) &\leq f(x(\mu), y) - f(x(\mu), y(\mu)) \\ &\leq L \|y - y(\mu)\| \quad \forall y \in S(x(\mu)). \end{aligned}$$

Taking the infimum on the right side over $y \in S(x(\mu))$ yields

$$0 < (\mu - L) d_S(x(\mu), y(\mu)) \leq 0$$

which is a contradiction. Therefore $y(\mu) \in S(x(\mu))$, i.e., $d_S(x(\mu), y(\mu)) = 0$. Thus, for any solution (\hat{x}, \hat{y}) to the GBLP, one has

$$f(x(\mu), y(\mu)) = f(x(\mu), y(\mu)) + \mu d_S(x(\mu), y(\mu)) \leq f(\hat{x}, \hat{y}) + \mu d_S(\hat{x}, \hat{y}) = f(\hat{x}, \hat{y})$$

which implies that $(x(\mu), y(\mu))$ is a solution of GBLP.

Conversely, let (x^*, y^*) be a global solution of the GBLP and consider a solution $(x(\mu), y(\mu))$ of DP_μ with $\mu > L$. The same argument as in the previous paragraph yields $d_S(x(\mu), y(\mu)) = 0$. Hence one has

$$\begin{aligned} f(x^*, y^*) + \mu d_S(x^*, y^*) &= f(x^*, y^*) \\ &\leq f(x(\mu), y(\mu)) = f(x(\mu), y(\mu)) + \mu d_S(x(\mu), y(\mu)). \end{aligned}$$

Therefore (x^*, y^*) is a solution of DP_μ , as claimed. ■

We now give a global exact penalty representation for RP.

Theorem 3.2 *Assume that $f(x, \cdot)$ is Lipschitz continuous with constant $L > 0$ uniformly in $x \in X$. Let r be an error bound satisfying inequality (8) for all $(x, y) \in \text{Gr}U$ and condition (A). Then for any $\mu > L$, (x^*, y^*) is a global solution of $RP_{\delta\mu}$ if and only if it is also a global solution of GBLP.*

Proof. Let (x^*, y^*) be a global solution of GBLP. Then $r(x^*, y^*) = d_S(x^*, y^*) = 0$ and it is also a global solution to DP_μ for any $\mu \geq L$ by virtue of Theorem 3.1. Let $(\bar{x}(\mu), \bar{y}(\mu))$ be a global solution of the penalized problem $RP_{\delta\mu}$. For any $\mu > L$ since (x^*, y^*) is a global solution to DP_μ and inequality (8) holds at every point of $\text{Gr}U$ by assumption, we have

$$\begin{aligned} f(x^*, y^*) + \delta\mu r(x^*, y^*) &= f(x^*, y^*) + \mu d_S(x^*, y^*) \\ &\leq f(\bar{x}(\mu), \bar{y}(\mu)) + \mu d_S(\bar{x}(\mu), \bar{y}(\mu)) \\ &\leq f(\bar{x}(\mu), \bar{y}(\mu)) + \delta\mu r(\bar{x}(\mu), \bar{y}(\mu)), \end{aligned}$$

which implies that (x^*, y^*) is a global solution of $\text{RP}_{\delta\mu}$ for $\mu \geq L$.

Conversely, let $((x(\mu), y(\mu)))$ be a global solution of $\text{RP}_{\delta\mu}$ where $\mu > L$. Let (x^*, y^*) be a global solution of GBLP. Then by Theorem 3.1 it is also a global solution of DP_μ for $\mu > L$. We have

$$\begin{aligned}
& f(x(\mu), y(\mu)) + \mu d_S(x(\mu), y(\mu)) \\
& \leq f(x(\mu), y(\mu)) + \delta\mu r(x(\mu), y(\mu)) \quad \text{by inequality (8)} \\
& \leq f(x^*, y^*) + \delta\mu r(x^*, y^*) \quad \text{by optimality of } (x(\mu), y(\mu)) \text{ in } \text{RP}_{\delta\mu} \\
& = f(x^*, y^*) + \mu d_S(x^*, y^*) \quad \text{since } r(x^*, y^*) = d_S(x^*, y^*) = 0,
\end{aligned}$$

which implies $(x(\mu), y(\mu))$ is a global solution of DP_μ . By Theorem 3.1 it is also a solution of GBLP and the proof is completed. \blacksquare

4 Some useful error bounds

Up to now we have seen that the error bounds of the inclusion $y \in S(x)$ with Property (A) can serve as a tool to study GBLP. In this section we shall give some such useful error bounds.

Definition 4.1 Let $\Omega \subset X$. The mapping $F : R^n \times R^m \rightarrow R^m$ is *strongly monotone with respect to y uniformly in $x \in \Omega$* if there exists $\mu > 0$ such that

$$\langle F(x, y_2) - F(x, y_1), y_2 - y_1 \rangle \geq \mu \|y_2 - y_1\| \quad \forall y_1, y_2 \in U(x), x \in \Omega.$$

Definition 4.2 Let $\Omega \subset X$. $F(x, y)$ is *pseudo strongly monotone with respect to y uniformly in $x \in \Omega$* if there exists $\mu > 0$ such that

$$\langle F(x, y), z - y \rangle \geq 0 \quad \text{implies } \langle F(x, z), z - y \rangle \geq \mu \|z - y\|^2 \quad \forall y, z \in U(x), x \in \Omega.$$

1. Standard gap bound.

Consider a parametric variational inequality with *nonseparable and linear constraints*, i.e.

$$U(x) = \{y \in R^m \mid c(x, y) = Ax + By - b \leq 0\}$$

where A and B are $d \times n$ and $d \times m$ matrices respectively and $b \in R^d$. In this case, $\forall x_0 \in X$, a solution y_0 to the variational inequality (2) is characterized by the

following *complementarity system*:

$$\begin{aligned} F(x_0, y_0) + B^t \lambda_0 &= 0 \\ (Ax_0 + By_0 - b)^t \lambda_0 &= 0 \\ \lambda_0 &\geq 0 \end{aligned}$$

where λ_0 is a vector in R^d . This characterization means that $y_0 \in U(x_0)$ solves the variational inequality parameterized in x_0 (2) if and only if there exists $\lambda_0 \in R^d$ such that (x_0, y_0, λ_0) satisfies the above complementarity system. If *gradients of binding constraints* in the variational inequality (2) at (x_0, y_0) , i.e., $\nabla_y c_j(x_0, y_0)$ where j are such that $c_j(x_0, y_0) = 0, j \in \{1, 2, \dots, d\}$, are *linearly independent*, and the *strict complementarity condition*

$$\lambda_{0i} > 0 \iff c_i(x_0, y_0) = 0, \quad \forall i \in \{1, 2, \dots, d\} \quad (9)$$

holds, then the variational inequality parameterized in x (2) has a unique solution $y(x)$ for x in a neighborhood of x_0 and the above complementarity system has a unique solution $(y(x), \lambda(x))$ for x in a neighborhood of x_0 . Furthermore the functions $y(x)$ and $\lambda(x)$ are Lipschitz continuous and the strict complementarity condition (9) is satisfied in a neighborhood of x_0 (see e.g. Friesz et al. [5]).

The following result due to Marcotte and Zhu [8] shows that the gap function defined by (6) can serve as an useful error bound under certain conditions.

Proposition 4.1 *Assume that X is a compact, convex subset of R^n . Let the mapping F be strongly monotone with respect to y uniformly in $x \in X$ and $\nabla_y F$ be Lipschitz continuous in y uniformly in x . Let $x_0 \in X$. If the linear independence and strict complementarity conditions hold at $y_0 = y(x_0)$, then there exists a constant $\delta > 0$ and a neighborhood $N(x_0, y_0)$ of (x_0, y_0) such that*

$$d_S(x, y) \leq \delta G_0(x, y) \quad \forall (x, y) \in \text{Gr}U \cap N(x_0, y_0).$$

Note that $G_0(x, y)$ satisfies condition (A). So $G_0(x, y)$ is a useful error bound.

Now we consider a parametric variational inequality with separable and linear constraints, i.e. $U(x) = Y$ and Y is a convex polyhedron. In this case we can weaken the assumptions of Proposition 4.1.

We will need the following definition due to Dussault and Marcotte [3].

Definition 4.3 Let F be a continuous, monotone mapping from a convex polyhedron $X \subset \mathbb{R}^n$ into \mathbb{R}^n and denote by $\text{VIP}(X, F)$ the variational inequality problem associated with X and F , i.e: Find x^* in X such that

$$\text{VIP}(F, X) : \quad \langle F(x^*), x^* - x \rangle \leq 0 \quad \text{for all } x \text{ in } X.$$

We say that $\text{VIP}(F, X)$ is *geometrically stable* if, for any solution x^* of the variational inequality, $\langle F(x^*), x^* - x \rangle = 0$ implies that x lies on the optimal face, i.e. the minimal face of X containing the (convex) solution set to $\text{VIP}(F, X)$.

The following result due to Marcotte and Zhu [8] gives a useful error bound.

Proposition 4.2 *Let $U(x) = Y$ be a compact convex polyhedron. Let the mapping F be strongly monotone with respect to y uniformly in $x \in X$. Let $x_0 \in X$ and assume that there exists a neighborhood of x_0 such that $\text{VIP}(F(x, \cdot), Y)$ is geometrically stable inside that neighborhood. Then there exist some neighborhood $N(x_0)$ of x_0 and a positive number $\delta > 0$ such that*

$$d_S(x, y) \leq \delta G_0(x, y), \quad \forall x \in N(x_0), y \in U(x).$$

2. Square root standard gap bound

The following result gives an error bound in terms of square root of the gap function G_0 .

Proposition 4.3 *Assume that the mapping F is pseudo strongly monotone with respect to y uniformly in $x \in N(x_0)$. Then there exists $\delta > 0$ such that:*

$$d_S(x, y) \leq \delta \sqrt{G_0(x, y)}, \quad \forall x \in N(x_0), y \in U(x).$$

Proof. Let $y(x) \in S(x)$. Then by definition of $S(x)$ one has

$$\langle F(x, y(x)), y - y(x) \rangle \geq 0 \quad \forall y \in U(x).$$

Since $y(x) \in U(x)$, it follows from the pseudo strong monotonicity of F and definition of G_0 that for all $x \in N(x_0)$ and $y \in U(x)$ one has

$$\mu \|y(x) - y\|^2 \leq \langle F(x, y), y - y(x) \rangle \leq G_0(x, y),$$

from which the result follows readily. ■

3. Square root differentiable gap bound

Recently, Fukushima [6] gave an optimization formulation of variational inequality based the *differentiable gap function* defined as

$$G_\alpha(x, y) = \max_{z \in U(x)} \{ \langle F(x, y), y - z \rangle - \frac{1}{2\alpha} \|y - z\|_M^2 \}. \quad (10)$$

where $\|\cdot\|_M$ denote the norm in R^m defined by $\|z\|_M = \langle z, Mz \rangle^{\frac{1}{2}}$ and M is a symmetric positive definite matrix. $G_\alpha(x, y)$ also has Property (A). Now we give an error bound based on $\sqrt{G_\alpha}$.

Proposition 4.4 *Suppose $U(x)$ is convex and $x_0 \in X$. Let the mapping F be pseudo strongly monotone with respect to y uniformly in $x \in N(x_0)$. Then there exists $\delta > 0$ such that*

$$d_S(x, y) \leq \delta \sqrt{G_\alpha(x, y)}, \quad \forall x \in N(x_0), y \in U(x).$$

Proof. Let $y(x) \in S(x)$. Then by definition of $S(x)$ one has

$$\langle F(x, y(x)), y - y(x) \rangle \geq 0 \quad \forall y \in U(x).$$

Since $y(x) \in U(x)$, it follows from the pseudo strong monotonicity of F that for every $x \in N(x_0)$ and $y \in U(x)$ we have

$$\langle F(x, y), y - y(x) \rangle \geq \mu \|y - y(x)\|^2.$$

Let $y_t = y + t(y(x) - y)$ for $t \in [0, 1]$. Then by convexity of $U(x)$, $y_t \in U(x)$ for any $y \in U(x)$. From definition of $G_\alpha(x, y)$ (see (10)), we have

$$\begin{aligned} G_\alpha(x, y) &\geq \langle F(x, y), y - y_t \rangle - \frac{1}{2\alpha} \|y - y_t\|_M^2 \\ &= t \langle F(x, y), y - y(x) \rangle - \frac{t^2}{2\alpha} \|y - y(x)\|_M^2 \\ &\geq \left(t\mu - \frac{t^2 \|M\|}{2\alpha} \right) \|y - y(x)\|^2. \end{aligned}$$

Let $t = \min\{1, \frac{\alpha\mu}{\|M\|}\}$; this gives

$$G_\alpha(x, y) \geq \sigma \|y - y(x)\|^2$$

where

$$\sigma = \begin{cases} \left(\mu - \frac{\|M\|}{2\alpha} \right) & \text{if } \mu \geq \frac{\|M\|}{\alpha} \\ \frac{\alpha\mu^2}{2\|M\|} & \text{if } \mu \leq \frac{\|M\|}{\alpha} \end{cases}$$

which proves the result. ■

4. Projection bound

The proof of the following lemma is straightforward.

Lemma 4.1 *An arbitrary vector $y \in Y$ is a solution of the variational inequality with parameter x if and only if it satisfies*

$$h(x, y) = y - \text{proj}_{U(x)}(y - F(x, y)) = 0$$

where $\text{proj}_{U(x)}(z)$ is the orthogonal projection of a vector z onto the set $U(x)$.

It follows from the above lemma that $h(x, y)$ has Property (A) and any vector norm of $h(x, y)$ can be used as such measure of errors of the lower level problem. We choose the euclidean norm $\|\cdot\|$.

Proposition 4.5 *Let $x_0 \in X$. Assume the the mapping F is strongly monotone with respect to y uniformly in $N(x_0)$ and F is Lipschitz continuous in y with constant $L > 0$ uniformly in $x \in N(x_0)$. Then we have*

$$d_S(x, y) \leq ((L + 1)/\mu)\|h(x, y)\| \quad \forall x \in N(x_0), y \in N(x). \quad (11)$$

Proof. Since $y(x) \in S(x)$ and $y - h(x, y) = \text{proj}_{U(x)}(y - F(x, y)) \in U(x)$, by definition of $S(x)$ we have

$$\langle F(x, y(x)), y - h(x, y) - y(x) \rangle \geq 0.$$

On the other hand, since $y(x) \in U(x)$ and $y - h(x, y)$ is the projection of $y - F(x, y)$ onto $U(x)$, we obtain

$$\begin{aligned} & \langle (y - F(x, y)) - (y - h(x, y)), y(x) - (y - h(x, y)) \rangle \\ &= \langle F(x, y) - h(x, y), y - h(x, y) - y(x) \rangle \leq 0. \end{aligned}$$

Adding the above two inequalities, we deduce

$$\begin{aligned} 0 &\leq \langle F(x, y(x)) - F(x, y) + h(x, y), y - h(x, y) - y(x) \rangle \\ &= \langle F(x, y(x)) - F(x, y), y - h(x, y) - y(x) \rangle + \langle h(x, y), y - y(x) \rangle - \|h(x, y)\|^2 \\ &= \langle F(x, y(x)) - F(x, y), y - y(x) \rangle \\ &\quad + \langle F(x, y) - F(x, y(x)) + y - y(x), h(x, y) \rangle - \|h(x, y)\|^2, \end{aligned}$$

i.e.,

$$\langle F(x, y) - F(x, y(x)), y - y(x) \rangle \leq \langle F(x, y) - F(x, y(x)) + y - y(x), h(x, y) \rangle - \|h(x, y)\|^2.$$

By strongly monotonicity and Lipschitz continuity of F in y uniformly in $x \in N(x_0)$ and Cauchy-Schwartz inequality, it follows that

$$\mu \|y - y(x)\|^2 \leq (L + 1) \|h(x, y)\| \|y - y(x)\|, \quad \forall x \in N(x_0), y \in U(x)$$

from which (11) follows. ■

5. Uniformly weak sharp minimum for lower level optimization problem

Definition 4.4 Let objective function $g(x, y)$ of the lower level problem (3) be convex in y and $U(x)$ is convex and let the solution set of the lower level problem $\Sigma_x = \arg \min_{y \in U(x)} g(x, y)$ be nonempty and closed. Problem (3) is said to have a *uniformly weak sharp minimum* (following a definition of a weak sharp minimum by Ferris and Mangasarian [4]) if there exists a positive constant α such that

$$d_S(x, y) := \text{dist}(y, \Sigma_x) \leq \alpha(g(x, y) - V(x)). \quad (12)$$

It is obvious that the uniformly weak sharp minimum can serve as an error bound.

Now we consider a linear lower level problem parameterized in x as follows:

$$\begin{aligned} \text{LP}_x \quad & \text{minimize} && px + qy \\ & \text{subject to} && Ax + By - b \leq 0 \\ & && y \in R^m, \end{aligned}$$

where $p \in R^n$, $q \in R^m$, A and B are $d \times n$ and $d \times m$ matrices respectively and $b \in R^d$. Suppose that the set $\{(x, y) : Ax + By - b \leq 0\}$ is nonempty.

The following result shows that the linear lower level problem LP_x always has a uniformly weak sharp minimum.

Proposition 4.6 *Problem LP_x has a uniformly weak sharp minimum globally, i.e. there exists $\alpha > 0$ such that*

$$d_S(x, y) \leq \alpha(px + qy - V(x)), \quad \forall (x, y) \in \{(x, y) : Ax + BY - b \leq 0, x \in R^n, y \in R^m\}.$$

For an ordinary (nonparameterized) linear programming problem, the above proposition is Lemma A.1 of [7]. Proposition 4.6 extends the result of Lemma A.1. of [7] to the parameterized linear programming problem LP_x . Since the proof of Proposition 4.6 is exactly similar to the proof of Lemma A.1 of [7], we will omit the proof here.

5 Necessary conditions for optimality

In this section we derive necessary conditions for optimality of Kuhn-Tucker type for problem GBLP.

Without loss of generality, we assume in this section that all solutions of the mathematical programming problems lie in the interior of their abstract constraint sets, i.e., $(x, y) \in \text{int } X \times Y$.

First we give a concise review of the material on nonsmooth analysis required in the sequel. Our reference is Clarke [2].

Consider the following mathematical programming problem:

$$\begin{array}{lll} \text{P} & \text{minimize} & \phi(x, y) \\ & \text{subject to} & c(x, y) \leq 0 \\ & & x \in X, y \in Y. \end{array}$$

Its perturbed problem is

$$\begin{array}{lll} \text{P}(\alpha) & \text{minimize} & \phi(x, y) \\ & \text{subject to} & c(x, y) + \alpha \leq 0 \\ & & x \in X, y \in Y. \end{array}$$

Definition 5.1 [Calmness] Let (x^*, y^*) solves P. Problem P is *calm* at (x^*, y^*) provided that there exist $\delta > 0$ and $\mu > 0$ such that for all $\alpha \in \delta B$, for all $(x, y) \in (x^*, y^*) + \delta B$ which are feasible for $\text{P}(\alpha)$, one has

$$f(x, y) - f(x^*, y^*) + \mu \|\alpha\| \geq 0.$$

Definition 5.2 [Abnormal and Normal Multipliers] Let (x, y) be feasible for P. Define $M^0(x, y)$, the set of *abnormal multipliers* corresponding to (x, y) as the set

$$M^0(x, y) := \{s \in R^d : 0 \in \partial c(x, y)^\top s, s \geq 0, \langle s, c(x, y) \rangle = 0\}.$$

Define $M^1(x, y)$, the set of *normal multipliers* corresponding to (x, y) as the set

$$M^1(x, y) := \{s \in R^d : 0 \in \partial \phi(x, y) + \partial c(x, y)^\top s, s \geq 0, \langle s, c(x, y) \rangle = 0\}.$$

Remark 5.1 A sufficient condition for P to be calm at (x^*, y^*) is $M^0(x^*, y^*) = \{0\}$. In the classical setting, a sufficient condition for $M^0(x^*, y^*) = \{0\}$ is that the classical constraint qualifications such as the Mangasarian-Fromowitz and Slater type condition are satisfied. Therefore we shall call condition $M^0(x^*, y^*) = \{0\}$ the *usual constraint qualification*.

Proposition 5.1 [Kuhn-Tucker Lagrange Multiplier Rule] *Let (x^*, y^*) solve P . Suppose ϕ, c are locally Lipschitz near (x^*, y^*) and problem P is calm at (x^*, y^*) . Then there exists $s \geq 0$ such that*

$$\begin{aligned} 0 &\in \partial\phi(x^*, y^*) + \partial c(x^*, y^*)^\top s \\ 0 &= \langle s, c(x^*, y^*) \rangle. \end{aligned}$$

The following theorem gives a necessary condition for optimality when the error bound $r(x, y)$ is explicitly known.

Theorem 5.1 *Let (x^*, y^*) be a solution of problem GBLP. Let r be an error bound in a neighborhood of (x^*, y^*) with Property (A) and RP_μ be the associated penalized problem of RP, where $\mu > 0$. Assume that f and r are locally Lipschitz near (x^*, y^*) and the associated penalized problem RP_μ is calm at (x^*, y^*) . Then there exist non-zero vector $s \geq 0$ such that*

$$\begin{aligned} 0 &\in \partial f(x^*, y^*) + \mu \partial r(x^*, y^*) + \partial c(x^*, y^*)^\top s \\ 0 &= \langle s, c(x^*, y^*) \rangle \end{aligned}$$

where ∂ signifies the (Clarke) generalized gradient.

Proof. By Theorem 2.2, (x^*, y^*) is also a solution of the associated penalized problem RP_μ . The result follows from Proposition 5.1. ■

However in many cases, the error bounds are only an implicit function of the original problem data. Those useful error bounds derived in Section 4 involve the class of marginal functions or value functions. In order to derive necessary conditions in these cases, one must first study the generalized differentiability of marginal functions.

Consider the following parametric mathematical programming problem:

$$\begin{aligned} P_\alpha \quad &\text{minimize} && \phi(\alpha, y) \\ &\text{subject to} && c(\alpha, y) \leq 0 \\ &&& y \in Y, \end{aligned}$$

where $Y \subset R^m$ is closed. We assume that for problem P_α functions ϕ, c are locally Lipschitz near a given point of interest $y_0 \in Y$. Let y be feasible for P_α . Define

$$\begin{aligned} M_\alpha^0(y) &:= \{ \pi \in R^d : 0 \in \partial_y c(\alpha, y)^\top \pi, \langle \pi, c(\alpha, y) \rangle = 0, \pi \geq 0 \} \\ M_\alpha^1(y) &:= \{ \pi \in R^d : 0 \in \partial_y \phi(\alpha, y) + \partial_y c(\alpha, y)^\top \pi, \langle \pi, c(\alpha, y) \rangle = 0, \pi \geq 0 \}. \end{aligned}$$

Let $W(\alpha) = \inf\{\phi(\alpha, y) : c(\alpha, y) \leq 0, y \in Y\}$. The following result is a easy consequence of Corollary 1 of Theorem 6.5.2 of Clarke [2].

Proposition 5.2 [Generalized Differentiability of Marginal Functions] *Let Σ_{α_0} be the solution set to problem P_{α_0} and suppose it is nonempty. Suppose $M_{\alpha_0}^0(\Sigma_{\alpha_0}) = \{0\}$. Then $W(\alpha)$ is Lipschitz near α_0 and one has*

$$\partial W(\alpha_0) \subset \text{co}\{\partial_{\alpha}\phi(\alpha_0, y) + \partial_{\alpha}c(\alpha_0, y)^{\top}\pi : y \in \Sigma_{\alpha_0}, \pi \in M_{\alpha_0}^1(y)\}.$$

Rewrite $G_0(x, y) = -\min\{F(x, y), z - y : c(x, z) \leq 0, z \in Y\}$. Then the parameter here is $\alpha = (x, y)$. Let $\Sigma_{(x, y)}$ denote the set of vectors in which $G_0(x, y)$ attains the maximum. By Proposition 5.2, one has the following result:

Proposition 5.3 *Suppose $M_{(x^*, y^*)}^0(\Sigma_{(x^*, y^*)}) = \{0\}$. Assume that f, F, c are locally Lipschitz continuous near (x^*, y^*) and $\partial F(x^*, y^*) \subset \partial_x F(x^*, y^*) \times \partial_y F(x^*, y^*)$. Then $G_0(x, y)$ is Lipschitz near (x^*, y^*) and one has*

$$\begin{aligned} \partial G_0(x^*, y^*) \\ \subset \text{co}\{(\partial_x F(x^*, y^*)^{\top}(y^* - y) - \partial_x c(x^*, y)^{\top}\pi, \partial_y F(x^*, y^*)^{\top}(y^* - y) + F(x^*, y^*)) : \\ y \in \Sigma_{(x^*, y^*)}, \pi \in M_{(x^*, y^*)}^1(y)\}, \end{aligned}$$

where

$$\begin{aligned} M_{(x^*, y^*)}^0(y) &= \{\pi \in R^d : 0 \in \partial_y c(x^*, y)^{\top}\pi, \pi \geq 0, \langle \pi, c(x^*, y) \rangle = 0\} \\ M_{(x^*, y^*)}^1(y) &= \{\pi \in R^d : 0 \in F(x^*, y^*) + \partial_y c(x^*, y)^{\top}\pi, \pi \geq 0, \langle \pi, c(x^*, y) \rangle = 0\}. \end{aligned}$$

Combining Proposition 5.3, Remark 5.1 and Theorem 2.2, 2.3, 5.1, one has the following result:

Theorem 5.2 *Suppose f, F, c are C^1 . Let (x^*, y^*) be a solution of GBLP. Assume either of the following assumptions satisfied:*

- $G_0(x, y)$ is the error bound in a neighborhood of (x^*, y^*) .
- $\sqrt{G_0(x, y)}$ is the error bound in a neighborhood of (x^*, y^*) and f is upper Hölder continuous with exponent 2 near every $y \in S(x)$ uniformly in x in a neighborhood of x^* .

Suppose $M^0(x^*, y^*) = \{0\}$ and $M_{(x^*, y^*)}^0(\Sigma_{(x^*, y^*)}) = \{0\}$. Then there exist $\mu > 0$, $s \in R^d$, positive integers I, J , $\lambda_{ij} \geq 0$, $\sum_{i=1}^I \sum_{j=1}^J \lambda_{ij} = 1$, $y_i \in \Sigma_{(x^*, y^*)}$, $\pi_{ij} \in R^d$ such that

$$0 = \nabla_x f(x^*, y^*) + \nabla_x c(x^*, y^*)^\top s + \mu \sum_{ij} \lambda_{ij} \{ \nabla_x F(x^*, y^*)^\top (y^* - y_i) - \nabla_x c(x^*, y_i)^\top \pi_{ij} \}$$

$$0 = \nabla_y f(x^*, y^*) + \nabla_y c(x^*, y^*)^\top s + \mu \sum_{ij} \lambda_{ij} \{ \nabla_y F(x^*, y^*)^\top (y^* - y_i) + F(x^*, y^*) \}$$

$$0 = \langle s, c(x^*, y^*) \rangle, s \geq 0$$

$$0 = F(x^*, y^*) + \nabla_y c(x^*, y_i)^\top \pi_{ij}$$

$$0 = \langle \pi_{ij}, c(x^*, y_i) \rangle, \pi_{ij} \geq 0.$$

Note that a sufficient condition for both $M^0(x^*, y^*) = \{0\}$ and $M_{(x^*, y^*)}^0(\Sigma_{(x^*, y^*)}) = \{0\}$ to hold is that the Slater condition holds for $c(x, y) \leq 0$, $x \in X, y \in Y$.

For $G_\alpha(x, y)$, since y is the unique solution in the right hand side of (10), we have $\Sigma_{(x, y)} = \{y\}$. By Proposition 5.2, one has the following result:

Proposition 5.4 *Suppose f, F, c are locally Lipschitz continuous near (x^*, y^*) . Assume that $M_{(x^*, y^*)}^0(y^*) = \{0\}$. Then $G_\alpha(x, y)$ is Lipschitz near (x^*, y^*) and one has*

$$\partial G_\alpha(x^*, y^*) \subset \{(-\partial_x c(x^*, y^*)^\top \pi, F(x^*, y^*)) : \pi \in M_{(x^*, y^*)}^1(y^*)\},$$

where

$$M_{(x^*, y^*)}^1(y^*) = \{\pi \in R^d : 0 \in F(x^*, y^*) + \partial_y c(x^*, y^*)^\top \pi, \pi \geq 0, \langle \pi, c(x^*, y^*) \rangle = 0\}.$$

Furthermore if c is a C^1 function and $M_{(x, y)}^1(y) = \{\pi\}$ is a singleton, then $G_\alpha(x, y)$ is C^1 and one has

$$\nabla G_\alpha(x, y) = (-\nabla_x c(x, y)^\top \pi, F(x, y)).$$

Combining Proposition 5.4, Remark 5.1 and Theorem 2.2 2.3, 5.1, one has the following result:

Theorem 5.3 *Let (x^*, y^*) be a solution of GBLP. Suppose F is locally Lipschitz near (x^*, y^*) and f, c are C^1 functions. Assume that either of the following assumptions satisfied:*

- $G_\alpha(x, y)$ is the error bound in a neighborhood of (x^*, y^*) .

- $\sqrt{G_\alpha(x, y)}$ is the error bound in a neighborhood of (x^*, y^*) and f is upper Hölder continuous with exponent 2 near every $y \in S(x)$ uniformly in x in a neighborhood of x^* .

Suppose $M^0(x^*, y^*) = \{0\}$ and $M_{(x^*, y^*)}^0(y^*) = \{0\}$. Then there exist $\mu > 0$, $s \in R^d$, $\pi \in R^d$ such that

$$\begin{aligned} 0 &= \nabla_x f(x^*, y^*) + \nabla_x c(x^*, y^*)^\top s - \mu \nabla_x c(x^*, y^*)^\top \pi \\ 0 &= \nabla_y f(x^*, y^*) + \nabla_y c(x^*, y^*)^\top s + \mu F(x^*, y^*) \\ 0 &= \langle s, c(x^*, y^*) \rangle = 0, s \geq 0 \\ 0 &= F(x^*, y^*) + \nabla_y c(x^*, y^*)^\top \pi \\ 0 &= \langle \pi, c(x^*, y^*) \rangle, \pi \geq 0. \end{aligned}$$

Note that a sufficient condition for both $M^0(x^*, y^*) = \{0\}$ and $M_{(x^*, y^*)}^0(y^*) = \{0\}$ to hold is that the Slater condition holds for $c(x, y) \leq 0, x \in X, y \in Y$.

Now consider the classical bilevel programming problem. Let the solution set to the lower level optimization problem (3) be Σ_x . By Proposition 5.2, one has the following result:

Proposition 5.5 *Suppose g, c are locally Lipschitz continuous near (x^*, y) for all $y \in \Sigma_{x^*}$. Assume that Σ_{x^*} is nonempty and $M_{x^*}^0(\Sigma_{x^*}) = \{0\}$, then $V(x)$ is Lipschitz near x^* and one has*

$$\partial V(x^*) \subset \text{co}\{\partial_x g(x^*, y) + \partial_{c_x}(x^*, y)^\top \pi : y \in \Sigma_{x^*}, \pi \in M_{x^*}^1(y)\},$$

where $M_{x^*}^1(y) = \{\pi \in R^d : 0 \in \partial_y g(x^*, y) + \partial_y c(x^*, y)^\top \pi, \pi \geq 0, \langle \pi, c(x^*, y) \rangle = 0\}$.

Combining Proposition 5.5, Remark 5.1 and Theorem 2.2, 5.1, one has the following result:

Theorem 5.4 *Suppose f, g, c are C^1 . Let (x^*, y^*) be a solution of the classical bilevel programming problem. Assume $g(x, y) - V(x)$ is the error bound (i.e. an uniformly weak sharp minimum for the lower level problem) in a neighborhood of (x^*, y^*) . Suppose $M^0(x^*, y^*) = \{0\}$ and $M_{x^*}^0(\Sigma_{x^*}) = \{0\}$. Then there exist $\mu > 0$, $s \in R^d$, positive integer I, J , $\lambda_{ij} \geq 0, \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij} = 1, y_i \in \Sigma_{x^*}, \pi_{ij} \in R^d$ such that*

$$0 = \nabla_x f(x^*, y^*) + \nabla_x c(x^*, y^*)^\top s$$

$$\begin{aligned}
& +\mu \sum_{ij} \lambda_{ij} \{ \nabla_x g(x^*, y^*) - \nabla_x g(x^*, y_i) - \nabla_x c(x^*, y_i)^\top \pi_{ij} \} \\
0 & = \nabla_y f(x^*, y^*) + \nabla_y c(x^*, y^*)^\top s + \mu \nabla_y g(x^*, y^*) \\
0 & = \langle s, c(x^*, y^*) \rangle, s \geq 0 \\
0 & = \nabla_y g(x^*, y_i) + \nabla_y c(x^*, y_i)^\top \pi_{ij} \\
0 & = \langle \pi_{ij}, c(x^*, y_i) \rangle, \pi_{ij} \geq 0.
\end{aligned}$$

Note that a sufficient condition for both $M^0(x^*, y^*) = \{0\}$ and $M^0_{(x^*, y^*)}(\Sigma_{x^*}) = \{0\}$ to hold is that the Slater condition holds for $c(x, y) \leq 0, x \in X, y \in Y$.

Combining Proposition 4.6 and Theorem 5.4, one concludes that a consequence of Theorem 5.4 is the following necessary condition for optimality for a classical bilevel programming problem with a linear lower level problem.

Corollary 5.1 *Let (x^*, y^*) be a solution of the classical bilevel programming problem with linear lower level problem LP_x . Assume that the objective function of the upper level problem $f(x, y)$ is C^1 . Then there exist $\mu > 0$, $s \in R^d$, positive integer I, J , $\lambda_{ij} \geq 0, \sum_{i=1}^I \sum_{j=1}^J \lambda_{ij} = 1, y_i \in \Sigma_{x^*}, \pi_{ij} \in R^d$ such that*

$$\begin{aligned}
0 & = \nabla_x f(x^*, y^*) + A^\top s - \mu \sum_{ij} \lambda_{ij} A^\top \pi_{ij} \\
0 & = \nabla_y f(x^*, y^*) + B^\top s + \mu q \\
0 & = \langle s, Ax^* + By^* - b \rangle, s \geq 0 \\
0 & = q + B^\top \pi_{ij} \\
0 & = \langle \pi_{ij}, Ax^* + By_i - b \rangle, \pi_{ij} \geq 0.
\end{aligned}$$

Remark 5.2 For a clear presentation, we have assumed in Theorem 5.2, 5.3 and 5.4 that f, F, g, c are C^1 functions. However the theorems can be also stated without difficulty when f, F, g and c are merely Lipschitz continuous.

Note that in all necessary conditions for optimality given in this section, not only the coefficient of f is one (the multiplier is normal) but also the coefficient μ of r is positive. Thus, the lower level constraint is essentially involved in the necessary condition.

6 Relationships between error bounds and Examples

In this section, we study relations between various error bounds defined in Section 4. by illustrative examples we show that various equivalent single level optimization formulations with error bounds and their corresponding necessary conditions for optimality complement with each other.

We establish first certain inequalities and equalities among various error bounds.

Proposition 6.1 *1. If the objective function $g(x, y)$ of the lower level optimization problem (3) is convex and differentiable in y then*

$$g(x, y) - V(x) \leq G_0(x, y). \quad (13)$$

Furthermore, if the lower level problem is linear, then

$$g(x, y) - V(x) = G_0(x, y).$$

2. For the generalized bilevel programming problem GBLP we have

$$\sqrt{G_\alpha(x, y)} \leq \sqrt{G_0(x, y)}. \quad (14)$$

3. For (x, y) in a neighborhood of the solution (x^, y^*) of problem GBLP*

$$G_0(x, y) \leq \sqrt{G_0(x, y)}. \quad (15)$$

4. For the generalized bilevel programming problem GBLP we have

$$\|h(x, y)\| \leq \sqrt{2G_0(x, y)}. \quad (16)$$

Proof. 1. Let $y \in \arg \min_{y \in U(x)} g(x, y)$. By convexity of $g(x, \cdot)$ and the definition of G_0 , we have

$$\begin{aligned} G_0(x, y) &\geq \langle \nabla_y g(x, y), y - y(x) \rangle \\ &\geq g(x, y) - g(x, y(x)) \\ &= g(x, y) - V(x). \end{aligned}$$

The second assertion follows from the definitions of $V(x)$ and $G_0(x, y)$.

2. From the definition of G_α and G_0 .

3. Since G_0 is continuous in (x, y) and $G_0(x^*, y^*) = 0$, $G_0(x, y) < 1$ in a neighborhood of the solution (x^*, y^*) of GBLP, which implies the result.
4. Taking $\alpha = 1$ and $M = I$ the identity matrix in the definition of G_α , we have

$$G_1(x, y) = \langle F(x, y), y - p(x, y) \rangle - \frac{1}{2} \|y - p(x, y)\|^2 \geq 0$$

where $p(x, y) = \text{Proj}_{U(x)}(y - F(x, y))$. Thus

$$\begin{aligned} G_0(x, y) &\geq \langle F(x, y), y - p(x, y) \rangle \\ &\geq \frac{1}{2} \|y - p(x, y)\|^2 = \frac{1}{2} \|h(x, y)\|^2. \end{aligned}$$

The proof is completed. ■

As shown in Section 5, one of the major functions of the exact penalty formulation with error bounds is for deriving necessary optimality conditions of Kuhn-Tucker type. For this purpose error bounds must have certain regularity conditions such as Lipschitz condition (see Theorem 5.1). Among the aforementioned error bounds, G_0 , h and $g - V$ are Lipschitz continuous under appropriate constraint qualifications on constraints $U(x)$ and the rests are generally not Lipschitz. If we have an exact penalty formulation with an given error bound satisfying condition (A) then a similar exact penalty formulation is also valid with that error bound replaced by a larger one satisfying condition (A) (the proof is similar to that of Theorem 2.2 in Section 2). Smaller error bounds in general require stronger conditions. Hence on one hand error bounds G_0 , h and $\phi - V$ can be Lipschitz continuous but requires stronger conditions. On the other hand, larger bounds such $\sqrt{G_0}$ may not be Lipschitz but require weaker conditions. In the case where error bounds are not Lipschitz continuous, Theorem 5.2 and 5.3 show that stronger assumptions such as upper Hölder continuity on the upper level objective functions may be required. Therefore various error bounds and their equivalent exact penalty representations complement with each other. The following are some illustrative examples.

Example 6.1 Consider the following generalized bilevel programming problem

$$\begin{aligned} \text{minimize} & \quad x^2 - 2y \\ \text{subject to} & \quad \langle y - x, y - z \rangle \leq 0 \\ & \quad z \in [0, 2x], y \in [0, 2x], x \geq 0. \end{aligned}$$

This problem is equivalent to

$$\begin{array}{ll}
\text{minimize} & x^2 - 2y \\
\text{subject to} & G_0(x, y) = 0 \\
& y \in [0, 2x], x \geq 0,
\end{array}$$

where

$$\begin{aligned}
G_0(x, y) &= \max_{z \in [0, 2x]} \langle y - x, y - z \rangle \\
&= y^2 - xy + \max_{z \in [0, 2x]} \langle y - x, -z \rangle \\
&= y^2 - 2xy - x[(y - x) - |y - x|] \\
&= (y - x)^2 + x|y - x|.
\end{aligned}$$

It is easy to see that $y = x$ is the solution to constraint $G_0(x, y) = 0$. Thus, $(1, 1)$ is the solution of the GBLP. We claim that $(1, 1)$ satisfies the conditions of Proposition 4.1. In fact, since $(1, 1)$ is not a binding point we do not need to check the linear independent condition. Since when (x, y) is close to $(1, 1)$ both $-y$ and $y - 2x$ are negative, λ_0 in the complement condition must be 0. Thus, the strict complementarity condition is also satisfied. Hence by Proposition 4.1, the gap function $G_0(x, y)$ is a error bound in neighborhood of $(1, 1)$. We now verify that $(1, 1)$ satisfies Theorem 5.2. Since the constraint set $y - 2x \leq 0, x \geq 0$ has an interior point, the Slater condition is satisfied. If Theorem 5.2 holds, at $(x^*, y^*) = (1, 1)$, there must exist $\mu > 0$, $(s_1, s_2) \geq (0, 0)$, integers I, J , $(\lambda_{ij}^1, \lambda_{ij}^2) \geq (0, 0)$ such that $\sum_{j=1}^J \sum_{i=1}^I \lambda_{ij} = 1$, $y_i \in [0, 2]$ and $(\pi_{ij}^1, \pi_{ij}^2) \geq (0, 0)$ such that

$$\begin{aligned}
0 &= 2x^* - 2s_1 - s_2 - \mu \sum_{ij} \lambda_{ij} [-(y^* - y_i) + 2\pi_{ij}^1 + \pi_{ij}^2] \\
0 &= -2 + s_1 + \mu \sum_{ij} \lambda_{ij} [(y^* - y_i) + (y^* - x^*)] \\
0 &= s_1(y^* - 2x^*) \\
0 &= s_2x^* \\
0 &= (y^* - x^*) + \pi_{ij}^1 \\
0 &= \pi_{ij}^1(y^* - 2x^*) \\
0 &= \pi_{ij}^2x^*,
\end{aligned}$$

which is indeed satisfied at $(1, 1)$ if we choose, say, $s_1 = s_2 = 0$, $\pi_{ij}^1 = \pi_{ij}^2 = 0$, $\mu = 2$ and $y_i = 1$. This verifies the validity of Proposition 4.1 and Theorem 5.2.

It is worth noting that the GBLP considered in Example 6.1 is equivalent to the following classical bilevel programming problem

$$\begin{array}{ll}
\text{minimize} & x^2 - 2y \\
\text{subject to} & x \geq 0 \text{ and } y \text{ is a solution of the following problem} \\
& \text{minimize } y^2 - 2xy \\
& \text{subject to } y \in [0, 2x].
\end{array}$$

Direct calculation shows that the value function for the lower level problem is $V(x) = -x^2$. Therefore, according to [9], this problem is also equivalent to

$$\begin{array}{ll}
\text{minimize} & x^2 - 2y \\
\text{subject to} & y^2 - 2xy + x^2 = 0 \\
& y \in [0, 2x] \\
& x \geq 0.
\end{array}$$

However, it is easy to check above problem does not have any normal multiplier at $(1, 1)$. This fact shows that the gap function formulation is essentially different from the value function formulation.

We now change the above example a little bit to show that the linear independent and the strict complementarity conditions can not be omitted from Proposition 4.1.

Example 6.2 Consider the same problem in Example 6.1 with constraints $y, z \in [0, 2x]$ replaced by $y, z \in [0, x]$.

Again one can check $(1, 1)$ is the only solution to the problem. However, the gap function is different now. In fact, In this example

$$\begin{aligned}
G_0(x, y) &= \max_{z \in [0, x]} \langle y - x, y - z \rangle \\
&= y^2 - xy + \max_{z \in [0, x]} \langle y - x, -z \rangle \\
&= y^2 - xy + x^2 - xy \\
&= (y - x)^2.
\end{aligned}$$

Thus, the equivalent single level problem is

$$\begin{array}{ll}
\text{minimize} & x^2 - 2y \\
\text{subject to} & (y - x)^2 = 0 \\
& y \in [0, x], x \geq 0.
\end{array}$$

One can check directly that there is no normal multiplier at $(1, 1)$. This is due to the fact that the $(1, 1)$ is a binding point for the linear constraint $-y \leq 0$, $y - x \leq 0$ and the linear independent condition does not hold at this point.

On the other hand, we can derive a necessary condition by using a square root differentiable gap bound. Observe first that $\langle y - x, z - y \rangle \geq 0$ implies

$$\langle z - x, z - y \rangle = \langle z - y, z - y \rangle + \langle y - x, z - y \rangle \geq \|z - y\|^2.$$

The pseudo strong monotone condition of Proposition 4.4 is satisfied. Next we calculate

$$\begin{aligned} G_1(x, y) &= \max_{z \in [0, x]} \{ \langle y - x, y - z \rangle - \frac{1}{2}(y - z)^2 \} \\ &= \frac{1}{2}(y - x)^2. \end{aligned}$$

Since $\sqrt{G_1(x, y)} = \sqrt{1/2}|y - x|$ is Lipschitz continuous, Proposition 4.4 and Theorem 5.1 imply the necessary condition in Theorem 5.1 is valid for $r(x, y) = \sqrt{G_1(x, y)}$. To verify, observe that Theorem 5.1 holds, at $(x^*, y^*) = (1, 1)$, there must exist $\mu > 0$, $s_1 \geq 0$, $s_2 \geq 0$ such that

$$\begin{aligned} 0 &\in 2x^* - \mu[-1, 1] - 2s_1 - s_2 \\ 0 &\in -2 + \mu[-1, 1] + s_1 \\ 0 &= s_1(y^* - x^*) \\ 0 &= s_2x^*, \end{aligned}$$

which is indeed satisfied at $(1, 1)$ if we take $s_1 = s_2 = 0$ and $\mu = 2$.

Finally, we give an example which has a square root standard gap bound that is not Lipschitz continuous.

Example 6.3 Consider the following classical bilevel programming problem:

$$\begin{aligned} \text{minimize} & \quad (x - 1)^2 + x^2(y + 1) \\ \text{subject to} & \quad -1 \leq x \leq 1 \text{ and } y \text{ is a solution of the following problem} \\ & \quad \text{minimize } (\sin \frac{\pi}{2}x)y \\ & \quad \text{subject to } y \in [-1, 1]. \end{aligned}$$

$(1, -1)$ is the optimal solution of the problem and the solution set of the lower level problem is

$$S(x) = \begin{cases} \{1\} & \text{if } -1 \leq x < 0 \\ [-1, 1] & \text{if } x = 0 \\ \{-1\} & \text{if } 0 < x \leq 1. \end{cases}$$

The standard gap function of the problem is

$$\begin{aligned} G_0(x, y) &= \max\{\sin \frac{\pi}{2}x \cdot (y - z) : z \in [-1, 1]\} \\ &= \begin{cases} \sin \frac{\pi}{2}x \cdot (y - 1) & -1 \leq x < 0 \\ 0 & x = 0 \\ \sin \frac{\pi}{2}x \cdot (y + 1) & 0 < x \leq 1. \end{cases} \end{aligned}$$

Since $F(x, y) = \sin \frac{\pi}{2}x$ is independent of y , F is pseudo strongly monotone with respect to y . By Proposition 4.3, $\sqrt{G_0(x, y)}$ is an error bound in the neighborhood of $(1, -1)$. However $\sqrt{G_0(x, y)}$ is not Lipschitz continuous near $(x^*, y^*) = (1, -1)$. Theorem 5.1 can not be used. We now verify that the assumptions of Theorem 5.2 is satisfied. The objective function $f(x, y) = (x - 1)^2 + x^2(y + 1)^2$ is upper Hölder continuous near every $y \in S(x)$ uniformly for x in the neighborhood of -1 . Since the constraint set $-1 \leq x \leq 1, -1 \leq y \leq 1$ has an interior point, the Slater condition is satisfied. Theorem 5.2 implies that at $(x^*, y^*) = (1, -1)$, there must exist $\mu > 0$, $(s_1, s_2, s_3) \geq (0, 0, 0)$, integer J and $\lambda_j \geq 0$, $\sum_{j=1}^J \lambda_j = 1$ such that

$$\begin{aligned} 0 &= 2(x^* - 1) + 2x^*(y^* + 1)^2 + s_3 - \mu \sum_j \lambda_j \pi_j^3 \\ 0 &= 2x^{*2}(y^* + 1) + \mu \sin(\frac{\pi}{2}x^*) + s_1 - s_2 \\ 0 &= s_1(y^* - 1) \\ 0 &= s_2(-1 - y^*) \\ 0 &= s_3(x^* - 1) \\ 0 &= \sin(\frac{\pi}{2}x^*) + \pi_j^1 - \pi_j^2 \\ 0 &= \pi_j^1(y^* - 1) \\ 0 &= \pi_j^2(-1 - y^*) \\ 0 &= \pi_j^3(x^* - 1). \end{aligned}$$

Indeed, the above condition holds for $J = 1$, $\lambda_1 = 1$, $\mu = s_2 = \pi_1^2 = 1$ and $s_1 = s_3 = \pi_1^1 = \pi_1^3 = 0$.

References

- [1] G. Anandalingam and T.L. Friesz, Editors, Hierarchical Optimization, *Annals of Operations Research* **34**(1992).
- [2] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983.
- [3] J.-P. Dussault and P. Marcotte, Conditions de régularité géométrique pour les inéquations variationnelles , *RAIRO Rech. Opér* **23**(1988), 1-16.
- [4] M.C. Ferris and O.L. Mangasarian, Minimum principle sufficiency, *Mathematical Programming* **57**(1992), 1-14.
- [5] T.L. Friesz, R.T. Tobin, H-J Cho and N.J. Mehta, Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints, *Mathematical Programming* **48**(1990), 265-284.
- [6] M. Fukushima, Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems, *Mathematical Programming* **53**(1992), 99-110.
- [7] O.L. Mangasarian and R.R. Meyer, Nonlinear perturbation of linear programs, *SIAM J. Control and Optimization* **17**(1979), 745-752.
- [8] P. Marcotte and D.L. Zhu, Exact and inexact penalty methods for the generalized bilevel programming problem, Publication of Centre de recherche sur les transports, Université de Montréal, CRT-920 (1992).
- [9] J.J. Ye and D.L. Zhu, Optimality conditions for bilevel programming problems, Research report of Department of Mathematics and Statistics, University of Victoria, DMS-618-IR (1992).