SIAM J. OPTIM.
Vol. 7, No. 2, pp. 481–507, May 1997

012

EXACT PENALIZATION AND NECESSARY OPTIMALITY CONDITIONS FOR GENERALIZED BILEVEL PROGRAMMING PROBLEMS*

J. J. YE†, D. L. ZHU‡, AND Q. J. ZHU§

Abstract. The generalized bilevel programming problem (GBLP) is a bilevel mathematical program where the lower level is a variational inequality. In this paper we prove that if the objective function of a GBLP is uniformly Lipschitz continuous in the lower level decision variable with respect to the upper level decision variable, then using certain uniform parametric error bounds as penalty functions gives single level problems equivalent to the GBLP. Several local and global uniform parametric error bounds are presented, and assumptions guaranteeing that they apply are discussed. We then derive Kuhn–Tucker-type necessary optimality conditions by using exact penalty formulations and nonsmooth analysis.

Key words. generalized bilevel programming problems, variational inequalities, exact penalty formulations, uniform parametric error bounds, necessary optimality conditions, nonsmooth analysis

AMS subject classifications. 49K99, 90C, 90D65

PII. S1052623493257344

1. Introduction.

We consider the following *mathematical programming problem with variational inequality constraints* (which is called the *generalized bilevel programming problem* (GBLP)):

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

where $f : R^{n+m} \to R$, 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 with parameter* x,

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

Here $U : X \to R^m$ is a set-valued map and $F : R^{n+m} \to R^m$ is a function. Throughout this paper, we make the blanket assumption that $\mathrm{Gr}S := \{(x,y) : x \in X, y \in S(x)\}$, the graph of S, is not empty.

One can interpret the above problem as a *hierarchical decision process* where there are two decision makers and the upper level decision maker always has the first choice as follows: given a decision vector x for the upper level decision maker (the *leader*), $S(x)$ is viewed as the lower level decision maker's (the *follower's*) decision set, i.e., the set of decision vectors that the follower may use. Assuming that the game is *cooperative* (i.e., the follower's decision set $S(x)$ is not a singleton), the follower allows the leader to choose the lower level decision from $S(x)$. Having complete knowledge of the follower's possible reactions, the leader selects decision vectors $x \in X$ and $y \in S(x)$, minimizing his objective function $f(x,y)$.

* Received by the editors October 18, 1993; accepted for publication (in revised form) December 15, 1995.
http://www.siam.org/journals/siopt/7-2/25734.html

† Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada V8W 3P4 (janeye@uvic.ca). The research of this author was supported by National Science and Engineering Research Council of Canada grant WFA 0123160.

‡ Center for Research on Transportation, University of Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec, Canada H3T 1V6 (daoli@crt.umontreal.ca).

§ Department of Mathematics and Statistics, Western Michigan University, Kalamazoo, MI 49008-5152 (zhu@math-stat.wmich.edu).

If $F(x, y)$ is the partial gradient of a real-valued differentiable function (i.e., $F(x, y) = -\nabla_y g(x, y)$, where $g : R^{n+m} \to R$ is differentiable in y and $U(x)$ is convex), then the variational inequality with parameter x,

$$\langle F(x, y), y - z\rangle \leq 0 \qquad \forall z \in U(x), \tag{2}$$

is the first-order necessary optimality condition for the following optimization problem with parameter x:

$$\mathrm{P}_x \qquad \text{minimize}_y\ g(x, y) \qquad \text{subject to } y \in U(x) \tag{3}$$

(see, e.g., [13]). Furthermore, if $g(x, y)$ is pseudoconvex 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 to (3). In this case, the mathematical programming problem with variational inequality constraints (1) is the *classical bilevel programming problem* (CBLP), or Stackelberg game (see, e.g., [1, 6, 17, 26, 29, 30, 31, 32]),

$$\text{CBLP} \qquad \text{minimize } f(x, y) \qquad \text{subject to } x \in X \text{ and } y \in \Sigma(x),$$

where $\Sigma(x)$ is the set of solutions for the problem P_x. The correspondence between lower level problems breaks down if F is not the partial gradient of a function with respect to y. Since problem (1) includes problems that are not classical bilevel programming problems, we call problem (1) a *generalized bilevel programming problem* (GBLP). The problem has been studied under the name "mathematical programs with equilibrium constraints" by other authors (see [12] and [19]).

In this paper we assume that

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

where $c : R^{n+m} \to R^d$ is a function. Throughout this paper we assume that f, c, and F are continuous. Under these assumptions, it is known [12, Lem. 1] that the solution set $S(x)$ of the variational inequality with parameter x is closed. Refer to [12] for the results on the existence of solutions for GBLP and CBLP.

Reducing a (generalized or classical) bilevel programming problem to a single level optimization problem is a useful strategy from both theoretical and computational points of view. There are several equivalent single level formulations for the GBLP. The Karush–Kuhn–Tucker (KKT) approach is to interpret the variational inequality constraint $y \in S(x)$ with y being a solution of the following optimization problem:

$$\text{minimize } \langle F(x, y), z\rangle \qquad \text{subject to } z \in U(x),$$

and to replace this minimization problem by its KKT necessary optimality conditions. These conditions are also sufficient if the feasible region $U(x)$ is convex. Assuming that $U(x)$ is convex, $c(x, y)$ is differentiable in y and one of the usual constraint qualifications, such as the Mangasarian–Fromowitz, condition is satisfied by the system of constraints $c(x, y) \leq 0$ in terms of variable y at a feasible point (x^*, y^*). Then (x^*, y^*) is a solution to the GBLP if and only if there exists $u^* \in R^d$ such that (x^*, y^*, u^*) is a solution to the following problem:

$$\begin{aligned} \text{KS} \qquad & \min f(x, y) \\ & \text{s.t. } F(x, y) + \nabla_y c(x, y)^T u = 0, \\ & \qquad \langle u, c(x, y)\rangle = 0, \\ & \qquad u \geq 0, c(x, y) \leq 0, \\ & \qquad x \in X, y \in R^m. \end{aligned} \tag{4}$$

To handle general GBLPs and CBLPs where $U(x)$ *may or may not be convex*, the value function and the gap function can be used to derive equivalent single level problems. Consider the CBLP. Define the *value function* $V(x) : X \to [-\infty, \infty]$ by

$$V(x) := \inf\{g(x, y) : y \in U(x)\}. \tag{5}$$

Then, for any $x \in X$, we have

$$g(x, y) - V(x) \geq 0 \ \forall y \in U(x), \text{ and } g(x, y) - V(x) = 0 \text{ if and only if } y \in \Sigma(x). \tag{6}$$

Thus, CBLP is equivalent to the following single level optimization problem:

$$\begin{aligned} \text{VS} \qquad & \min f(x, y) \\ & \text{s.t. } g(x, y) - V(x) = 0, \\ & \quad\;\; c(x, y) \leq 0, \\ & \quad\;\; x \in X, y \in R^m. \end{aligned} \tag{7}$$

Following [14] and [25], define the *gap function* $G_0(x, y) : X \times R^m \to [-\infty, \infty]$ by

$$G_0(x, y) := \sup\{\langle F(x, y), y - z\rangle : z \in U(x)\}. \tag{8}$$

It is easy to see that, for any $x \in X$,

$$G_0(x, y) \geq 0 \, \forall y \in U(x) \text{ and } G_0(x, y) = 0 \text{ if and only if } y \in S(x). \tag{9}$$

Hence GBLP is equivalent to the following single level optimization problem:

$$\begin{aligned} \text{GS} \qquad & \min f(x, y) \\ & \text{s.t. } G_0(x, y) = 0, \\ & \quad\;\; c(x, y) \leq 0, \\ & \quad\;\; x \in X, y \in R^m, \end{aligned} \tag{10}$$

Using the single level equivalent formulations KS, VS, and GS (see (4), (7), and (10)), one can derive Fritz John-type necessary optimality conditions for the original GBLP or CBLP. (See, e.g., [30] for the derivation of Fritz John-type necessary optimality conditions for CBLP.) In deriving Kuhn–Tucker-type necessary optimality conditions, however, we need to find constraint qualifications. Unfortunately, the usual constraint qualifications such as the Mangasarian–Fromowitz condition, never hold for problems VS and GS. To see this, for convenience, we assume that $U(x) = R^m$, $X = R^n$ and that $g(x, y), V(x)$, and $G_0(x, y)$ are Lipschitz continuous. Now suppose that (x^*, y^*) is a solution of GBLP. Then (6) and (9) imply the inclusions $0 \in \partial(g(x^*, y^*) - V(x^*))$ and $0 \in \partial G_0(x^*, y^*)$, respectively. These imply that there always exist abnormal multipliers for problems VS and GS. This is equivalent to saying that the Mangasarian–Fromowitz condition will never hold (see, e.g., [30, Prop. 3.1] for the equivalence). This phenomenon is intrinsic in bilevel problems. Even when using the KKT approach, the usual constraint qualifications will never hold for KS as long as the lower level problem is constrained. The following is a precise statement of this fact.

PROPOSITION 1.1. *Let (x^*, y^*, u^*) be a solution of KS. Suppose that $I := \{0 \leq i \leq d : c_i(x^*, y^*) = 0\} \neq \emptyset$. Then the Mangasarian–Fromowitz condition does not hold at (x^*, y^*, u^*).*

Proof. The complementary slackness condition (4) implies that $u_i^* = 0\ \forall i \in I^c := \{0 \le i \le d : c_i(x^*, y^*) \neq 0\}$. So

$$c_i(x^*, y^*) = 0, \qquad i \in I$$

and

$$-u_i^* = 0, \qquad i \in I^c$$

are active constraints for KS at (x^*, y^*, u^*). Set

$$\widetilde{c}_i(x, y, u) := c_i(x, y),$$

$$\widehat{c}_i(x, y, u) := -u_i,$$

and

$$h(x, y, u) := \langle u, c(x, y) \rangle.$$

Suppose that there exists a vector $v \in R^{n+m+d}$ such that

$$\langle v, \nabla \widetilde{c}_i(x^*, y^*, u^*) \rangle = \sum_{j=1}^{n+m} v_j \nabla_j c_i(x^*, y^*) < 0\ \forall i \in I$$

and

$$\langle v, \nabla \widehat{c}_i(x^*, y^*, u^*) \rangle = -v_{n+m+i} < 0\ \forall i \in I^c,$$

where $\nabla_j c_i(x, y)$ denotes the gradient of c_i with respect to the jth component of the vector (x, y). Then

$$\begin{aligned} &\langle v, \nabla h(x^*, y^*, u^*) \rangle \\ &= \sum_{i \in I} u_i \sum_{j=1}^{n+m} v_j \nabla_j c_i(x^*, y^*) + \sum_{i \in I^c} v_{n+m+i} c_i(x^*, y^*) < 0. \end{aligned}$$

Thus, the Mangasarian–Fromowitz condition cannot hold at (x^*, y^*, u^*). ☐

The difficulty here is obviously due to the equality constraints (4), (7), and (10), which reflect the bilevel nature of the problem.

The partial calmness condition is identified in [30] as an appropriate constraint qualification for problem VS. It is also proved that the existence of a uniformly weak sharp minimum is a sufficient condition for partial calmness, and a parametric linear lower level problem is always partially calm.

Recently, using the theory of exact penalization for mathematical programming problems with subanalytic constraints and the theory of error bounds for quadratic inequality systems, Luo et al. [19] successfully derived various penalty functions for the single level equivalent mathematical programming problem KS. By using the theory of parametric normal equations, Luo et al. [19] also obtained some necessary and sufficient stationary point conditions for GBLP.

In this paper we use the *uniform parametric error bound* as a tool to establish (local or global) exact penalty formulations of several single level mathematical programming problems (including KS, VS, and GS) that are equivalent to GBLP. Since the exact penalty formulations move the troublesome equality constraints (4), (7), and (10) to the objective function, we can get Kuhn–Tucker-type necessary optimality conditions under the usual constraint qualifications. The concept of a uniform parametric error bound generalizes the uniformly weak sharp minimum defined in [30]. Thus, the uniform parametric error bounds derived in this paper provide many more exact penalty formulations than those in [30] for VS. Using the uniform parametric error bound as a tool, the conditions we derived in this paper are very general and distinct (cf. Theorem 6.5) from the ones derived in [19].

The paper is arranged as follows. In the next section we introduce uniform parametric error bounds and show that they provide local and global exact penalty formulations of GBLP. In section 3, we discuss several useful uniform parametric error bounds. Kuhn–Tucker-type necessary optimality conditions for problem GBLP associated with various uniform parametric error bounds are derived in section 4. In section 5, the relationships between various uniform parametric error bounds are discussed and some examples are given showing that the various equivalent single level optimization formulations with uniform parametric error bounds and their corresponding necessary optimality conditions complement each other. In section 6, we show that uniform parametric error bounds can be used to derive exact penalty formulations for KS.

2. Partial calmness and exact penalization. In this section we introduce uniform parametric error bounds and show that they are useful in deriving exact penalty formulations for GBLP.

Consider the following mathematical programming problem:

$$\begin{array}{lll} \text{MP} & \text{minimize} & f(x) \\ & \text{subject to} & h(x) = 0, \\ & & g(x) \le 0, \\ & & x \in C, \end{array}$$

where $f : R^n \to R$, $h : R^n \to R$, $g : R^n \to R^m$ are lower semicontinuous and C is a closed subset in R^n. The corresponding *perturbed problem* is

$$\begin{array}{lll} \text{MP}(\epsilon) & \text{minimize} & f(x) \\ & \text{subject to} & h(x) = \epsilon, \\ & & g(x) \le 0, \\ & & x \in C, \end{array}$$

where $\epsilon \in R$. The following definition was introduced in [30].

DEFINITION 2.1 (partial calmness). *Let x^* solve* MP. *The problem* MP *is said to be* partially calm *at x^* provided that there exist constants $\mu > 0, \delta > 0$ such that, for all $\epsilon \in \delta B$ and all $x \in x^* + \delta B$ that are feasible for* MP*(ϵ), one has*

$$f(x) - f(x^*) + \mu|h(x)| \ge 0.$$

Here B denotes the open unit ball in R^n. The constants μ and δ are called the modulus and radius, respectively.

The partial calmness condition is similar to, but different from, the calmness condition introduced by Clarke and Rockafellar (see, e.g., [5]; see also Definition 4.1) in that only the equality constraint $h(x) = 0$ is perturbed.

The concept of calmness was shown to be closely related to "exact penalization" in [5, Prop. 6.4.3]. More precisely, if x^* is a local solution of MP and the problem MP is calm at x^*, then x^* is a local solution for a penalized problem. In the following proposition we show that the concept of partial calmness is equivalent to local exact penalization.

Proposition 2.2. *Assume that f is continuous. Suppose x^* is a local minimum of* MP *and* MP *is partially calm at x^*. Then there exists $\mu^* > 0$ such that x^* is a local minimum of the following penalized problems for all $\mu \geq \mu^*$:*

$$\begin{array}{lll} \mathrm{MP}_\mu \qquad & \textit{minimize} & f(x) + \mu|h(x)| \\ & \textit{subject to} & g(x) \leq 0, \\ & & x \in C. \end{array}$$

Any local minima of MP_μ *with $\mu > \mu^*$ with respect to the neighborhood of x^* in which x^* is a local minimum are also local minima of* MP.

Proof. Suppose that x^* is a local minimum of MP but not MP_μ for any $\mu > 0$. Then, for each positive integer k, there exists a point $x_k \in x^* + (1/k)B \subset C$ and $g(x_k) \leq 0$ such that

$$f(x_k) + k|h(x_k)| < f(x^*). \tag{11}$$

Since x^* is a local minimum of MP, the above inequality implies that $|h(x_k)| > 0$. Therefore,

$$0 < |h(x_k)| < \frac{f(x^*) - f(x_k)}{k}. \tag{12}$$

Taking the limit as k goes to infinity in (12), one has

$$|h(x_k)| \to 0 \text{ as } k \to \infty.$$

But then the inequality (11) contradicts the hypothesis that MP is partially calm at x^*. Thus for some $\mu^* > 0$, x^* must be a local minimum of MP_{μ^*}.

It is obvious that a local minimum of MP_{μ^*} must be a local minimum for MP_μ whenever $\mu \geq \mu^*$.

Conversely, let $\mu > \mu^*$ and x_μ be a local minimum of MP_μ in the neighborhood of x^* in which x^* is a local minimum. Then

$$\begin{aligned} f(x_\mu) + \mu|h(x_\mu)| = f(x^*) \qquad & \text{since } x^* \text{ is a local minimum of } \mathrm{MP}_\mu, \\ \leq f(x_\mu) + \frac{1}{2}(\mu + \mu^*)|h(x_\mu)| \qquad & \text{since } \frac{1}{2}(\mu + \mu^*) > \mu^*, \end{aligned}$$

which implies that

$$(\mu - \mu^*)|h(x_\mu)| \leq 0.$$

Therefore, $h(x_\mu) = 0$, which implies that x_μ is also a local minimum of MP. □

Remark 2.3. Notice that in the above result, no continuity assumption is required for the function $h(x)$. When the function h is continuous, it is easy to see that if MP is partially calm at a solution x^* of MP with modulus μ and radius ϵ, then there exists a $\hat{\delta} \leq \delta$ such that x^* is a $\hat{\delta}$-local solution to the penalized problem MP_μ; i.e.,

$$f(x) + \mu|h(x)| \geq f(x^*) \quad \forall x \in C \text{ s.t. } g(x) \leq 0, x \in x^* + \hat{\delta}B.$$

Therefore, in our definition of partial calmness, the restriction on the size of perturbation $\epsilon \in \delta B$ can be removed when h is continuous, and it then corresponds to the definition of calmness given by Burke [2]. Furthermore, the infimum of μ^* in Proposition 2.2 can be taken as the modulus of partial calmness.

For any $x \in X, y \in R^m$, define the parametric distance function

$$d_{S(x)}(y) := \inf\{\|y - z\| : z \in S(x)\}$$

to be the distance from the point y to the set $S(x)$. The GBLP is equivalent to a *mathematical programming problem involving a parametric distance function constraint*:

$$\begin{array}{lll} \text{DP} & \text{minimize} & f(x,y) \\ & \text{subject to} & d_{S(x)}(y) = 0, \\ & & c(x,y) \le 0, \\ & & x \in X, y \in R^m. \end{array}$$

It is known (see [5, Prop. 2.4.3]) that if the objective function of a constrained optimization problem is Lipschitz continuous then the distance function is an exact penalty term. In what follows, we extend this result to the mathematical programming problem with variational inequality constraints, GBLP. The constraint implied in the parametric distance function is, in fact, in the lower level decision variable. It is natural that we only need to assume that the objective function is locally Lipschitz in the lower level decision variable uniformly in the upper level decision variable to prove the exact penalty property of the parametric distance function. We need the following definition.

From now on we shall use $N(z)$ to denote a neighborhood of z.

DEFINITION 2.4. *Let $(x^*, y^*) \in R^{n+m}$. The function $f(x,y)$ is said to be locally Lipschitz near y^* uniformly in $x \in N(x^*)$ if there exists $L > 0$ and a neighborhood $N(y^*)$ of y^* such that*

$$|f(x,y') - f(x,y)| \le L|y' - y| \qquad \forall y', y \in N(y^*), x \in N(x^*).$$

The following result generalizes Proposition 2.4.3 of Clarke [5] to GBLP. We omit the proof of the global result, since it is essentially the same as the local one and the converse part of the proof in Proposition 2.2.

THEOREM 2.5. *Let (x^*, y^*) be a local solution of problem* DP. *Assume that f is locally Lipschitz near y^* uniformly in x on a neighborhood of x^* with constant L. Then problem* DP *is partially calm at (x^*, y^*) with modulus L.*

Furthermore, let (x^*, y^*) be a global solution of GBLP and assume that $f(x,\cdot)$ is Lipschitz continuous in y with constant $L > 0$ uniformly for all $x \in X$. Then (x^*, y^*) is a global solution of the penalized problem

$$\begin{array}{lll} \text{DP}_\mu & \textit{minimize} & f(x,y) + \mu d_{S(x)}(y) \\ & \textit{subject to} & c(x,y) \le 0, \\ & & x \in X, y \in R^m \end{array}$$

for any $\mu \ge L$, and any other global solution of DP_μ *for any $\mu > L$ is also a global solution of GBLP.*

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 $0 \le \epsilon < \delta$, let $(x,y) \in (x^*, y^*) + \delta B$ be feasible for DP_ϵ; i.e., $d_{S(x)}(y) = \epsilon$ and $c(x,y) \le 0$, $(x,y) \in (x^*, y^*) + \delta B$. 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

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

we have

$$f(x,y') \ge f(x^*, y^*). \tag{13}$$

Since $f(x,\cdot)$ is locally Lipschitz near y^*,

$$f(x,y) - f(x,y') \geq -L\epsilon. \tag{14}$$

Combining (13) and (14) yields

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

i.e., DP is partially calm at (x^*,y^*) with modulus L. ∎

Theorem 2.5 shows that the distance function provides an exact penalty equivalent formulation for GBLP under very mild conditions. However, the parametric distance function is usually an implicit nonsmooth function of the data in the original problem. It is difficult to compute or estimate its Clarke generalized gradient.

To overcome this difficulty, we shall use the parametric distance function $d_{S(x)}(y)$ establishing some equivalent exact penalty formulations of GBLP. These equivalent formulations have penalty functions with computable Clarke generalized gradients.

We call a function $r(x,y) : R^{n+m} \to R$ a *merit function* provided

$$r(x,y) \geq 0 \quad \forall (x,y) \in \mathrm{Gr}U \text{ and } r(x,y) = 0 \text{ if and only if } (x,y) \in \mathrm{Gr}S. \tag{15}$$

A merit function is called a *uniform parametric error bound* for the inclusion $y \in S(x)$ with modulus $\delta > 0$ in the set $Q \subset \mathrm{Gr}U$ if it satisfies

$$d_{S(x)}(y) \leq \delta r(x,y) \qquad \forall (x,y) \in Q. \tag{16}$$

A merit function provides the following equivalent formulation of GBLP:

$$\begin{array}{lll} \text{RP} & \text{minimize} & f(x,y) \\ & \text{subject to} & r(x,y) = 0, \\ & & c(x,y) \leq 0, \\ & & x \in X, y \in R^m. \end{array}$$

Its corresponding penalized problem is

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

Next we show that if $r(x,y)$ is a uniform parametric error bound and f is Lipschitz near y^* uniformly in x, then there exists $\mu > 0$ such that the problem RP_μ is an exact penalty equivalence of RP. As in Theorem 2.5 we omit the proof for the global result.

THEOREM 2.6. *Let (x^*,y^*) be a local solution of problem* GBLP *and r be a uniform parametric error bound with modulus $\delta > 0$ in a neighborhood of (x^*,y^*). Suppose that f is locally Lipschitz near y^* uniformly for all x in a neighborhood of x^*. Then there exists $\mu^* > 0$ such that (x^*,y^*) is a local solution of the penalized problem* RP_μ *for all $\mu \geq \delta\mu^*$ and any local solution to* RP_μ *with $\mu > \delta\mu^*$ with respect to the neighborhood of (x^*,y^*) is also a local solution to* RP.

Furthermore, let (x^*,y^*) be a global solution of GBLP *and r be a uniform parametric error bound in* $\mathrm{Gr}U$. *Assume that $f(x,\cdot)$ is Lipschitz continuous with constant $L > 0$ uniformly for all $x \in X$. Then (x^*,y^*) is a global solution of* RP_μ *for all $\mu \geq \delta L$, and any other global solution of* RP_μ *for all $\mu > \delta L$ is also a global solution of* GBLP.

Proof. Being a local solution of GBLP, (x^*,y^*) is also a local solution of DP. DP is partially calm by Theorem 2.5. Thus, by Proposition 2.2, there exists a $\mu^* > 0$

such that (x^*, y^*) is also a solution to DP_{μ^*}. Hence, for all (x, y) in a neighborhood of (x^*, y^*) which are feasible for $\mathrm{RP}_{\delta\mu^*}$, one has

$$\begin{aligned} f(x^*, y^*) + \delta\mu^* \cdot r(x^*, y^*) &= f(x^*, y^*) + \mu^* \cdot d_{S(x^*)}(y^*) && \text{since } y^* \in S(x^*), \\ &\leq f(x, y) + \mu^* \cdot d_{S(x)}(y) && \text{since } (x^*, y^*) \text{ solves } \mathrm{DP}_{\mu^*}, \\ &\leq f(x, y) + \delta\mu^* \cdot r(x, y) && \text{by inequality (16).} \end{aligned}$$

Therefore, (x^*, y^*) is also a local solution of $\mathrm{RP}_{\delta\mu^*}$. The proof for the converse is similar to that of the converse part of Proposition 2.2. □

Remark 2.7. As in Remark 2.3 when the uniform parametric error bound r is continuous, the constant μ^* in Theorem 2.6 can be taken as the modulus of partial calmness, which is the Lipschitz constant of $f(x, \cdot)$ by virtue of Theorem 2.5.

Sometimes a uniform parametric error bound is not nicely behaved but its square is; e.g., $\sqrt{|x|}$ is not Lispchitz continuous near 0 but $|x|$ is. Therefore, we are interested in the following formulations which are equivalent to GBLP when $r(x, y)$ is a merit function.

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

Its penalized problem is

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

Although the penalty term $r^2(x, y)$ might be better behaved, it is smaller than $r(x, y)$ for all (x, y) that are close to (x^*, y^*). Hence, to formulate an equivalent exact penalty formulation for the problem RSP, one needs to impose a stronger condition on f. The following definition gives such a condition.

DEFINITION 2.8. *Let* $x_0 \in X$. *The mapping* $f(x, y) : R^n \times R^m \to R$ *is* upper Hölder continuous with exponent 2 near every $y \in S(x)$ uniformly for x in a neighborhood of x_0 *provided there exists* $L > 0$ *such that*

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

The constant L *is called the modulus.*

We prove that $r^2(x, y)$ provides an exact penalty formulation for GBLP if $r(x, y)$ is a uniform parametric error bound and f is upper Hölder continuous with exponent 2 near every $y \in S(x)$ uniformly in x in a neighborhood of x^*.

THEOREM 2.9. *Let* (x^*, y^*) *be a local solution of the problem* RSP. *Assume that* r *is a uniform parametric error bound with modulus* δ *in a neighborhood of* (x^*, y^*) *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* (x^*, y^*) *is a local solution of the penalized problem* RSP_μ *for all* $\mu \geq \delta^2 L$, *and any local solution to* RSP_μ *with* $\mu > \delta^2\mu^*$ *in the neighborhood of* (x^*, y^*) *is also a local solution to* RSP.

Proof. 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 ε, $0 \leq \epsilon^{\frac{1}{2}} < \alpha$, 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 $\|y - y'(x)\| = d_{S(x)}(y) \leq \delta r(x, y) = \delta\epsilon^{\frac{1}{2}}$. Since $(x, y'(x))$ is feasible for RSP and

$$\begin{aligned} \|(x, y'(x)) - (x^*, y^*)\| &\leq \|(x, y'(x)) - (x, y)\| + \|(x, y) - (x^*, y^*)\| \\ &\leq \delta\epsilon^{\frac{1}{2}} + \alpha < \alpha(\delta + 1), \end{aligned}$$

we have

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

Therefore

$$\begin{aligned}
&f(x,y) - f(x^*, y^*)\\
&\geq f(x,y) - f(x, y'(x)) \qquad \text{by optimality of } (x^*, y^*),\\
&\geq -L\|y - y'(x)\|^2 \qquad \text{by upper Hölder continuity of } f,\\
&= -L(d_{S(x)}(y))^2,\\
&\geq -L\delta^2 r^2(x,y) \qquad \text{since } r(x,y) \text{ is a uniform parametric error bound},\\
&= -L\delta^2\epsilon;
\end{aligned}$$

i.e., RSP is partially calm at (x^*, y^*) with modulus $\delta^2 L$. The rest of the proof is similar to the converse part of Proposition 2.2. □

3. Some uniform parametric error bounds. In this section we discuss some useful uniform parametric error bounds. We start with two definitions.

DEFINITION 3.1. *Let $\Omega \subset R^n$. A mapping $F(x, y) : R^n \times R^m \to R^m$ is called* strongly monotone with respect to y uniformly *in $x \in \Omega$ with modulus $\mu > 0$ provided*

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

DEFINITION 3.2. *Let $\Omega \subset R^n$. The mapping $F(x, y) : R^n \times R^m \to R^m$ is called* pseudostrongly monotone with respect to y uniformly in $x \in \Omega$ *with modulus $\mu > 0$ provided*

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

3.1. Uniformly weak sharp minima for the lower level optimization problem.

DEFINITION 3.3 (see [30]). *A family of parametric mathematical programming problems $\{(P_x) : x \in X\}$ as defined in (3) is said to have uniformly weak sharp minima in $\Omega \subset GrU$ if there exists an $\delta > 0$ such that*

$$d_{\Sigma(x)}(y) \leq \delta(g(x,y) - V(x)) \quad \forall (x,y) \in \Omega, \tag{17}$$

where $\Sigma(x)$ is the solution set of the lower level optimization problem P_x. The constant δ is called the modulus of the uniformly weak sharp minima.

By virtue of (9), $g(x, y) - V(x)$ is a merit function. When $\Sigma(x) = S(x)$ (e.g., when $U(x)$ is convex, $g(x, y)$ is pseudoconvex and differentiable in y), $g(x, y) - V(x)$ is obviously a uniform parametric error bound.

The next result follows easily from a result about regular points due to Ioffe (Theorem 1 and Corollary 1.1 of [8]).

PROPOSITION 3.4. *Let (x^*, y^*) be an optimal solution of the CBLP. Suppose that $g(x, y)$ is Lipschitz continuous in y uniformly in $x \in X$ with constant $L_g > 0$. Assume that there exist $\sigma > 0$ such that for any $(x, y) \in \mathrm{Gr}U$ satisfying $y \notin S(x)$ and any $\xi \in \partial_y g(x, y)$, $\eta \in (L_g + 1)\partial d_{S(x)}(y)$ (or $\eta \in N_{S(x)}(y)$),*

$$\|\xi + \eta\| \geq \sigma.$$

Then

$$d_{S(x)}(y) \le (1/\sigma)(g(x,y) - V(x)) \ \forall (x,y) \in \mathrm{Gr}U.$$

Consider the bilevel programming problem where the lower level problem is the following parametric quadratic programming problem:

$$\begin{aligned} \mathrm{QP}_x \qquad & \min g(x,y) := \langle y, Px\rangle + \frac{1}{2}\langle y, Qy\rangle + p^t x + q^t y \\ & \text{s.t. } y \in \Omega_x := \{y \in Y : Ax + By - b \le 0\}. \end{aligned}$$

Here $Q \in R^{m\times m}$ is a symmetric and positive semidefinite matrix, $p \in R^n$, $q \in R^m$, $P \in R^{m\times n}$; A and B are $d \times n$ and $d \times m$ matrices, respectively, and $b \in R^d$.

The next proposition gives a sufficient condition for the family of parametric quadratic programming problems $\{\mathrm{QP}_x : x \in R^n\}$ to have uniformly weak sharp minima.

PROPOSITION 3.5. *Assume that there exists a constant $M > 0$ such that for all $(x,y) \in \mathrm{Gr}S$, every element z of $(N(y,\Omega_x) + \mathrm{span}(\nabla_y g(x,\bar{y}))) \cap B$ can be expressed as*

$$z = \eta \nabla_y g(x,\bar{y}) + \xi,$$

where $|\eta| \le M$ and $\xi \in N(y,\Omega_x)$. Assume

$$\ker(\nabla_y^2 g(x,\bar{y}))^\perp \subset \mathrm{span}(\nabla_y g(x,\bar{y})) + N(y,\Omega_x) \ \ \forall\ (x,y) \in \mathrm{Gr}S \tag{18}$$

or, equivalently,

$$(\nabla_y g(x,\bar{y}))^\perp \cap T(y,\Omega_x) \subset \ker(\nabla_y^2 g(x,\bar{y})) \ \ \forall\ (x,y) \in \mathrm{Gr}S,$$

where $\bar{y}$ is any element in $S(x)$, $A^\perp := \{y \in R^m : \langle y,x\rangle = 0\ \forall x \in A\}$ denotes the subspace perpendicular to A, span(d) *represents the subspace generated by the vector d, $T(y,C)$ is the tangent cone to the set C at y, and* $\ker(A)$ *is the nullspace of the matrix A. Then $\{QP_x : x \in X\}$ has uniformly weak sharp minima.*

Before proving the above result we first state the following description of the solution set of a convex program given in Mangasarian [21].

LEMMA 3.6. *Let S be the set of solutions to the problem* $\min\{g(y) : y \in \Omega\}$ *where $g : R^n \to R$ is a twice continuously differentiable convex function and Ω is a convex subset of R^n. Let $\bar{y} \in S$. Then*

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

It follows that for QP_x, the solution set $S(x)$ is

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

Since Ω_x is a polyhedral one has

$$T(y, S(x)) = T(y,\Omega_x) \cap (\nabla_y g(x,\bar{y}))^\perp \cap \ker(\nabla_y^2 g(x,\bar{y})) \tag{19}$$

by virtue of Corollaries 16.4.2 and 23.8.1 of Rockafellar [28].

Proof of Proposition 3.5. By virtue of Theorem 2.6 of Burke and Ferris [4], it suffices to show that for all $x \in X, y \in S(x)$, there exists an $\alpha > 0$ such that

$$g_2'(x,y;d) \geq \alpha \|d\| \quad \forall d \in T(y,\Omega_x) \cap N(y,S(x)),$$

where $g_2'(x,y;d)$ is the directional derivative of g with respect to y in the direction d. Note that (19) and (18) imply that

$$\begin{aligned} N(y,S(x)) &= N(y,\Omega_x) + \mathrm{span}(\nabla_y g(x,\bar{y})) + \ker(\nabla_y^2 g(x,\bar{y}))^\perp \\ &= N(y,\Omega_x) + \mathrm{span}(\nabla_y g(x,\bar{y})). \end{aligned}$$

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

$$\begin{aligned} \|d\| &= \sup\{\langle z(x), d\rangle : z(x) \in B \cap N(y,S(x))\} \\ &\leq \sup\{\langle \eta \nabla_y g(x,\bar{y}) + \xi, d\rangle : |\eta| < M, \xi \in N(y,\Omega_x)\} \\ &\leq M\langle \nabla_y g(x,\bar{y}), d\rangle = M\langle \nabla_y g(x,y), d\rangle = M g_2'(x,y;d). \end{aligned}$$

The first inequality follows from the assumption, and the second equality follows from Lemma 3.6. Setting $\alpha = 1/M$ completes the proof. □

The following bilinear programming problem with parameter x is a special case of QP_x.

$$\begin{array}{lll} \mathrm{BLP}_x & & \min\langle y, Px\rangle + p^t x + q^t y \\ & & \text{s.t. } Ax + By - b \leq 0, \\ & & \quad y \in R^m. \end{array}$$

Proposition 3.5 has the following simple consequence.

COROLLARY 3.7. *The bilinear programming problem* BLP_x *has a uniformly weak sharp minima if there exists a constant* $M > 0$ *such that for all* $(x,y) \in \mathrm{Gr}S$*, every element* z *of* $(N(y,\Omega_x) + \mathrm{span}(Px+q)) \cap B$ *can be expressed as*

$$z = \eta(Px+q) + \xi$$

where $|\eta| \leq M$ *and* $\xi \in N(y,\Omega_x)$.

The following example shows that the assumption in Corollary 3.7 cannot be omitted.

Example 3.8. Consider the problem

$$\begin{array}{l} \min x + y \\ \text{s.t. } 0 \leq x \leq 1, y \in \arg\min\{-xy : x + y - 1 \leq 0, y \geq 0\}. \end{array}$$

The solution set of the lower level problem is

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

The value function of the lower problem is

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

It is easy to check that the assumption in Corollary 3.7 is not satisfied and there is no uniformly weak sharp minimum. In fact, if we replace the constraint $0 \leq x \leq 1$ by $0 < \epsilon \leq x \leq 1$, then the assumption in Corollary 3.7 is satisfied, and uniformly weak sharp minima exist.

3.2. A standard gap bound. Consider a parametric variational inequality with *nonseparable and linear constraints*, i.e.,

$$U(x) = \{y \in R^m | c(x,y) = Ax + By - b \leq 0\}, \tag{20}$$

where A and B are $d \times n$ and $d \times m$ matrices, respectively, and $b \in R^d$. In this case, $\forall x_0 \in X$, $y_0 \in U(x_0)$ solve the variational inequality with parameter x_0 (see (2)) if and only if there exists $\lambda_0 \in R^d$ such that (x_0, y_0, λ_0) satisfies the following *complementarity system*:

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

If the *gradients of the binding constraints* in the variational inequality (2) at (x_0, y_0), i.e., those $\nabla_y c_j(x_0, y_0)$ such that $c_j(x_0, y_0) = 0, j \in \{1, 2, \ldots, d\}$, are *linearly independent*, and the *strict complementarity condition*

$$\lambda_{0i} > 0 \iff c_i(x_0, y_0) = 0, \qquad \forall i \in \{1, 2, \ldots, d\} \tag{21}$$

holds, then the variational inequality (2) with parameter x has a unique solution $y(x)$ for all x in a neighborhood of x_0, and the above complementarity system has a unique solution $(y(x), \lambda(x))$ for all x in a neighborhood of x_0. Furthermore, the functions $y(x)$ and $\lambda(x)$ are Lipschitz continuous, and the strict complementarity condition (21) is satisfied in a neighborhood of x_0 (see, e.g., Friesz et al. [10]).

The following result due to Marcotte and Zhu [25] shows that the gap function defined by (8) can serve as a uniform parametric error bound under certain conditions.

PROPOSITION 3.9. *Assume that X is a compact, convex subset of R^n and $U(x)$ defined as in* (20) *is compact. Let the mapping F be strongly monotone with respect to y uniformly in $x \in X$, and let $\nabla_y F$ be Lipschitz continuous in y uniformly in x. Suppose $x_0 \in X$. 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 of (x_0, y_0) such that*

$$d_{S(x)}(y) \leq \delta G_0(x, y) \qquad \forall (x, y) \in \mathrm{Gr} U \cap N(x_0, y_0).$$

Now we consider a parametric variational inequality with separable and linear constraints; i.e., $U(x) = \{y \in R^m | By \leq b\}$ is a convex polyhedron. In this case we can weaken the assumptions of Proposition 3.9.

We need the following definition due to Dussault and Marcotte [7].

DEFINITION 3.10. *Let F be a continuous, monotone mapping from a convex polyhedron $X \subset R^n$ into R^n and denote by* VIP*(X, F) the variational inequality problem associated with X and F; i.e., find x^* in X such that*

$$\mathrm{VIP}(F, X) \qquad\qquad \langle F(x^*), x^* - x \rangle \leq 0 \qquad \textit{for all } x \textit{ in } X.$$

We say that 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* VIP*(F, X).*

The following result due to Marcotte and Zhu [25] gives a useful error bound.

PROPOSITION 3.11. *Assume that X is a convex polyhedron, $U(x) = \{y : By - b \leq 0\}$ is compact, and the mapping F is 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* $\mathrm{VIP}(F(x,\cdot), Y)$ *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*

$$d_{S(x)}(y) \leq \delta G_0(x,y) \qquad \forall y \in U(x), x \in N(x_0).$$

3.3. A square root standard gap bound. The following result gives a uniform parametric error bound in terms of the square root of the gap function G_0.

PROPOSITION 3.12. *Assume that the mapping F is pseudostrongly monotone with respect to y uniformly in $x \in N(x_0)$ with modulus μ. Then one has*

$$d_{S(x)}(y) \leq \frac{\sqrt{\mu}}{\mu}\sqrt{G_0(x,y)} \qquad \forall y \in U(x), x \in N(x_0).$$

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

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

Since $y(x) \in U(x)$, it follows from the pseudostrong monotonity of F and the 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. □

3.4. A square root differentiable gap bound. Recently, Fukushima [11] gave an optimization formulation of a variational inequality based on the *differentiable gap function* defined as

$$G_\alpha(x,y) = \max_{z\in U(x)}\left\{\langle F(x,y), y - z\rangle - \frac{1}{2\alpha}\|y - z\|_M^2\right\}, \tag{22}$$

where $\alpha > 0$ is a given constant, $\|\cdot\|_M$ denotes the elliptic norm in R^m defined by $\|z\|_M = \langle z, Mz\rangle^{\frac{1}{2}}$, and M is a symmetric positive definite matrix. It is easy to see that the differentiable gap function G_α satisfies condition (15). The following result gives a uniform parametric error bound based on $\sqrt{G_\alpha}$.

PROPOSITION 3.13. *Suppose $U(x)$ is convex and $x_0 \in X$. Let the mapping F be pseudostrongly monotone with respect to y uniformly in $x \in N(x_0)$. Then there exists $\delta > 0$ such that*

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

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

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

Since $y(x) \in U(x)$, it follows from the pseudostrong monotonity of F that, for every $x \in N(x_0)$ and $y \in U(x)$,

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

Let $y_t = y + t(y(x) - y)$ for $t \in [0, 1]$. By the convexity of $U(x)$, $y_t \in U(x)$ for any $y \in U(x)$. It follows from the definition of $G_\alpha(x, y)$ (see (22)) that

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

Letting $t = \min\{1, \frac{\alpha\mu}{\|M\|}\}$ gives

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

where

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

This proves the result. $\square$

3.5. A projection bound. The following projection characterization of $y \in S(x)$ is well known (see, e.g., [15]).

LEMMA 3.14. *An arbitrary vector $y \in Y$ is a solution of the variational inequality with parameter x if and only if it satisfies*

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

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

It follows from the above lemma that any vector norm of $h(x, y)$ satisfies condition (15). The following result is a parametric version of [27, Thm. 3.1]. The proof is omitted since it is essentially the same as that of [27, Thm. 3.1].

PROPOSITION 3.15. *Let $x_0 \in X$. Assume that the mapping F is strongly monotone with respect to y uniformly in $N(x_0)$ with modulus μ, and F is Lipschitz continuous in y with constant $L_F > 0$ uniformly in $x \in N(x_0)$. Then we have*

$$d_{S(x)}(y) \leq ((L_F + 1)/\mu)\|h(x,y)\| \qquad \forall y \in U(x), x \in N(x_0). \tag{23}$$

Remark 3.16. An important special case of GBLP is one where $F(x, y) = Qx + My + q$ and $U(x) = R^m_+$, the nonnegative orthant in R^m. In this case, finding a solution $y \in R^m$ to the parametric variational inequality (1) reduces to the parametric linear complementarity problem of finding a $y \in R^m$ satisfying

$$y \geq 0, Qx + My + q \geq 0, \langle y, Qx + My + q\rangle = 0.$$

The uniform projection error bound holds when M is a P-matrix (see Mathias and Pang [24]) and when M is an R_0-matrix. (See Mangasarian and Ren [23] and Luo and Tseng [18].)

4. Kuhn–Tucker-type necessary optimality conditions. In this section we derive Kuhn–Tucker-type necessary optimality conditions for 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.

First we give a concise review of the material on nonsmooth analysis. Our reference is Clarke [5].

Consider the following mathematical programming problem:

P	minimize	$\phi(x, y)$
	subject to	$c(x, y) \leq 0,$
		$x \in X, y \in R^m.$

The corresponding perturbed problem is

P(α)	minimize	$\phi(x, y)$
	subject to	$c(x, y) + \alpha \leq 0,$
		$x \in X, y \in R^m,$

where $\phi(x, y) : R^{n+m} \to R$ and $c(x, y) : R^{n+m} \to R^d$ are locally Lipschitz near the points of interest.

Definition 4.1 (calmness). *Let (x^*, y^*) solve* 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 P(α), one has*

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

Definition 4.2 (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 4.3. A sufficient condition for P to be calm at (x^*, y^*) is $M^0(x^*, y^*) = \{0\}$. $M^0(x^*, y^*) = \{0\}$ if and only if the Mangasarian–Fromowitz conditions are satisfied [30].

Proposition 4.4 (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*

$$0 \in \partial\phi(x^*, y^*) + \partial c(x^*, y^*)^\top s$$

and

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

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

Theorem 4.5. *Let (x^*, y^*) be a solution of problem* GBLP. *Let $r(x, y)$ be a uniform parametric error bound in a neighborhood of (x^*, y^*) 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 exists a nonzero vector $s \geq 0$ such that*

$$0 \in \partial f(x^*, y^*) + \mu\partial r(x^*, y^*) + \partial c(x^*, y^*)^\top s$$

and

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

Proof. By Theorem 2.6, (x^*, y^*) is also a solution of the associated penalized problem RP_μ. The result follows from Proposition 4.4. ◻

However, in many cases, uniform parametric error bounds are implicit functions of the original problem data. The useful uniform parametric 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:

$$\begin{array}{lll} \mathrm{P}_\alpha & \text{minimize} & \phi(\alpha, y) \\ & \text{subject to} & c(\alpha, y) \le 0, \\ & & y \in R^m. \end{array}$$

We assume that for problem P_α the functions ϕ and c are locally Lipschitz near the point of interest $y_0 \in R^m$. Let y be feasible for P_α. Define

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

Let $W(\alpha) = \inf\{\phi(\alpha, y) : c(\alpha, y) \le 0, y \in Y\}$. The following result is an easy consequence of Corollary 1 of Theorem 6.5.2 of Clarke [5].

PROPOSITION 4.6 (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*

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

where clco$A$ *denotes the closed convex hull of the set A.*

Set $G_0(x, y) = -\min\{\langle F(x, y), z - y\rangle : c(x, z) \le 0, z \in R^m\}$. The parameter here is $\alpha = (x, y)$. Let $\Sigma_{(x,y)}$ denote the set of vectors at which $G_0(x, y)$ attains the maximum. By Proposition 4.6, one has the following result.

PROPOSITION 4.7. *Suppose $M_{(x^*,y^*)}^0(\Sigma_{(x^*,y^*)}) = \{0\}$. Assume that $f, F,$ and c are locally Lipschitz near (x^*, y^*) and that $\partial F(x^*, y^*) \subset \partial_x F(x^*, y^*) \times \partial_y F(x^*, y^*)$. Then $G_0(x, y)$ is Lipschitz near (x^*, y^*) and one has*

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

where

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

Combining Proposition 4.7, Remark 4.3, and Theorems 2.6, 2.9, and 4.5, one has the following result.

THEOREM 4.8. *Suppose $f, F,$ and c are C^1. Let (x^*, y^*) be a solution of* GBLP. *Assume either of the following assumptions is satisfied:*

- *$G_0(x,y)$ is a uniform parametric error bound in a neighborhood of (x^*,y^*).*
- *$\sqrt{G_0(x,y)}$ is a uniform parametric 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 $M^0_{(x^*,y^*)}(\Sigma_{(x^*,y^*)}) = \{0\}$. Then there exist $\mu > 0$, $s \in R^d$, positive integers I, J, $\lambda_{ij} \geq 0, \sum_{i=1}^{I}\sum_{j=1}^{J}\lambda_{ij} = 1, y_i \in \Sigma_{(x^*,y^*)}$, and $\pi_{ij} \in R^d$ such that

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

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

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

For $G_\alpha(x,y)$, the differentiable gap function defined in (22), since y is the unique solution in the right-hand side of (22), we have $\Sigma_{(x,y)} = \{y\}$. By Proposition 4.6, one has the following result.

PROPOSITION 4.9. *Suppose f, F, and c are locally Lipschitz near (x^*,y^*). Assume that $M^0_{(x^*,y^*)}(y^*) = \{0\}$. Then $G_\alpha(x,y)$ is Lipschitz near (x^*,y^*) and one has*

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

where

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

Furthermore, if c is a C^1 function and $M^1_{(x,y)}(y) = \{\pi\}$ 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 4.9, Remark 4.3, and Theorems 2.6, 2.9, and 4.5, one has the following result.

THEOREM 4.10. *Let (x^*,y^*) be a solution of* GBLP. *Suppose F is locally Lipschitz near (x^*,y^*) and f and c are C^1 functions. Assume that either of the following assumptions is satisfied:*

- *$G_\alpha(x,y)$ is a uniform parametric error bound in a neighborhood of (x^*,y^*).*
- *$\sqrt{G_\alpha(x,y)}$ is a uniform parametric 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 $M^0_{(x^*,y^*)}(y^*) = \{0\}$. Then there exist $\mu > 0$, $s \in R^d$, and $\pi \in R^d$ such that

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

Remark 4.11. To shorten the exposition, we have assumed in Theorems 4.8 and 4.10 that f, F, g, and c are C^1 functions. However, these theorems can also be stated without difficulty when f, F, g, and c are merely Lipschitz continuous.

5. Relationships between various uniform parametric error bounds. In this section, we study the relationships between various uniform parametric error bounds. Through illustrative examples we show that various equivalent single level optimization formulations with uniform parametric error bounds and their corresponding necessary optimality conditions complement each other.

The following result is easy to prove.

PROPOSITION 5.1. *Suppose that r_S and r_B are two merit functions that satisfy the following inequality:*

$$r_S(x,y) \le \delta r_B(x,y) \qquad \forall (x,y) \in \mathrm{Gr}U,$$

for a constant $\delta > 0$. If $r_S(x,y)$ is a uniform parametric error bound, then so is $r_B(x,y)$.

Motivated by the above result we now establish certain inequalities and equalities among various uniform parametric error bounds.

PROPOSITION 5.2.

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

$$g(x,y) - V(x) \le G_0(x,y). \tag{24}$$

Furthermore, if the lower level problem is linear, then

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

(2) *For* GBLP*, we have*

$$\sqrt{G_\alpha(x,y)} \le \sqrt{G_0(x,y)}. \tag{25}$$

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

$$G_0(x,y) \le \sqrt{G_0(x,y)}. \tag{26}$$

(4) *For* GBLP*, we have*

$$\|h(x,y)\| \le \sqrt{2G_0(x,y)}. \tag{27}$$

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

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

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

(2) This follows directly from the definitions 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. This implies the result.

(4) Taking $\alpha = 1$ and $M = I$ the identity matrix in the definition of G_α, we have

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

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

As shown in section 4, one of the major applications of the exact penalty formulation with uniform parametric error bounds is to derive Kuhn–Tucker-type necessary optimality conditions. For this purpose parametric error bounds must be Lipschitz continuous (see Theorem 4.5). Among the aforementioned error bounds, G_0, h, and $g - V$ are Lipschitz continuous under appropriate constraint qualifications on $U(x)$. The rest are generally not Lipschitz. By virtue of Proposition 5.1, if we have an exact penalty formulation with a given uniform parametric error bound then a similar exact penalty formulation is also valid, with that error bound replaced by a larger one. Smaller error bounds generally require stronger conditions. Hence, on one hand, error bounds G_0, h, and $\phi - V$ can be Lipschitz continuous but require stronger conditions. On the other hand, larger bounds such as $\sqrt{G_0}$ may not be Lipschitz continuous but require weaker conditions. In the case when uniform parametric error bounds are not Lipschitz continuous, Theorems 4.8 and 4.10 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 each other. The following are some illustrative examples.

Example 5.3. Consider the following classical bilevel programming problem:

$$\begin{aligned} \text{(P1)} \quad & \min x^2 - 2y \\ & \text{s.t. } x \in [0,2] \text{ and } y \in \arg\min\{y^2 - 2xy : y \in [0,2x]\}. \end{aligned}$$

It is easy to verify that $(1,1)$ is the unique solution of (P1) and assumption (18) does not hold. Therefore, Proposition 3.5 does not apply and one may suspect that (P1) does not have a uniformly weak sharp minimum. Indeed, direct calculation shows that the value function for the lower level problem is $V(x) = x^2$. Using the value function approach, problem (P1) is equivalent to the following problem:

$$\begin{aligned} & \min x^2 - 2y \\ & \text{s.t. } (y - x)^2 = 0, \\ & \qquad y \in [0,2x], x \in [0,2]. \end{aligned}$$

Here $(y-x)^2$ is not an exact penalty term for the above problem, since for any $\mu > 0$ $(1, y)$ where $y \in (1, \frac{2+\mu}{\mu})$ assigns a lower value to the objective function than $(1,1)$ in the penalized problem

$$\begin{aligned} & \min x^2 - 2y + \mu(y-x)^2 \\ & \text{s.t. } y \in [0,2x], x \in [0,2]. \end{aligned}$$

It is clear that the function $F(x,y) = \nabla_y g(x,y) = y - x$ is strongly monotone in y uniformly for $x \in R$. The standard gap function takes the form

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

The linear independence and the strict complementarity conditions can easily be verified at $(1,1)$. Hence, by Proposition 3.9, the gap function $G_0(x,y)$ is a uniform parametric error bound in a neighborhood of $(1,1)$. Indeed, it is easy to see that $(1,1)$ is also the unique solution of the penalized problem

$$\text{s.t. } y \in [0,2x], x \in [0,2]$$

for any $\mu > 0$.

We now slightly modify the above example to show that the strict complementarity conditions cannot be omitted from Proposition 3.9.

Example 5.4. Consider the same problem in Example 5.3 with constraints $y, z \in [0,2x]$ replaced by $y, z \in [0,x]$ and with $x \in [0,2]$ replaced by $x \in [0,\infty)$.

Again, one can check that $(1,1)$ is the only solution to the problem. However, the gap function is different. 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 involving the standard gap function is

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

Again, $(y-x)^2$ is not an exact penalty term. This is due to the fact that the strict complementarity condition does not hold at (1,1).

$F(x,y) = y - x$ is strongly monotone; therefore, it is pseudostrongly monotone with respect to y uniformly for all $x \in R^n$. Using Propositions 3.12, 3.13, and 3.15, the problem has the square root standard gap bound, the square root differentiable gap bound, and the projection bound. The differentiable gap function associated with $\alpha = 1$ and $M = I$ takes the form

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

The projection bound takes the form $|h(x,y)| = |y-x|$. Indeed, the original problem is equivalent to the following penalized problem:

$$\text{s.t. } y \in [0,x], x \in [0,\infty),$$

for all $\mu > 0$.

Note that the uniform parametric error bounds for Example 5.4 are all Lipschitz continuous. We now give an example which has a square root standard gap bound that is not Lipschitz continuous.

Example 5.5. Consider the following classical bilevel programming problem:

$$\begin{aligned} &\min (x-1)^2 + x^2(y+1)^2 \\ &\text{s.t. } x \in [-1,1] \text{ and } y \in \arg\min \left\{ \left(\sin \frac{\pi}{2} x\right) y : y \in [-1,1] \right\}. \end{aligned}$$

Here $(1, -1)$ 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 for the problem is

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

Since $F(x, y) = \sin \frac{\pi}{2} x$ is independent of y, F is pseudostrongly monotone with respect to y uniformly for all x in a neighborhood of 1. By Proposition 3.12, $\sqrt{G_0(x, y)}$ 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 4.5 cannot be used.

We now verify that the assumptions of Theorem 4.8 are 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 a 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 4.8 implies that at $(x^*, y^*) = (1, -1)$, there must exist $\mu > 0$, $(s_1, s_2, s_3) \geq (0, 0, 0)$, an integer J, $\lambda_j \geq 0$, $\sum_{j=1}^{J} \lambda_j = 1$, and $\pi_j = (\pi_j^1, \pi_j^2, \pi_j^3) \in R^3$ such that

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

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

6. Exact penalty functions for the KKT formulation. In this section, we assume that $c(x, y)$ is convex and differentiable in y and that one of the usual constraint qualifications holds for the inequality system $c(x, y) \leq 0$ in terms of variable y. Under these assumptions, besides formulating GBLP as the single level equivalent problem GS or VS, one can also formulate GBLP as the equivalent single level problem KS. We will show that some of the uniform parametric error bounds such as $G_0(x, y)$, $\sqrt{G_0(x, y)}$, and $g(x, y) - V(x)$ can not only serve as exact penalty terms themselves, but can also play an important role in deriving equivalent exact penalty formulations for KS.

The following results establish the relationships among the KKT, the standard gap, and the value function formulations of GBLP.

PROPOSITION 6.1. *Suppose $c(x, y)$ is convex and differentiable in the y variable. Then*

$$G_0(x, y) \leq -\langle u, c(x, y)\rangle \text{ for all } (x, y, u) \in X \times R^m \times R^d \text{such that}$$
$$u \geq 0, c(x, y) \leq 0, F(x, y) + \nabla_y c(x, y)^t u = 0.$$

Proof. From mathematical programming weak duality (see, e.g., [20]), one has

$$\begin{aligned} G_0(x, y) &:= \sup\{\langle F(x, y), y - z\rangle : \forall z \in R^m \text{ s.t. } c(x, z) \leq 0\} \\ &= -\inf\{\langle F(x, y), z - y\rangle : \forall z \in R^m \text{ s.t. } c(x, z) \leq 0\} \\ &\leq -\sup\{\langle F(x, y), z - y\rangle + \langle u, c(x, y)\rangle : \forall (z, u) \in R^m \times R^d \text{ s.t.} \\ &\qquad\qquad u \geq 0, c(x, z) \leq 0, \\ &\qquad\qquad F(x, z) + \nabla_y c(x, z)^t u = 0\} \\ &\leq -\sup\{\langle u, c(x, y)\rangle : \forall (x, y, u) \in X \times R^m \times R^d \text{ s.t.} \\ &\qquad\qquad u \geq 0, c(x, y) \leq 0, \\ &\qquad\qquad F(x, y) + \nabla_y c(x, y)^t u = 0\}. \qquad \square \end{aligned}$$

Combining Proposition 6.1 and (1) from Proposition 5.2, we get the following result.

COROLLARY 6.2. *Assume that the objective function $g(x, y)$ for the lower level optimization problem* (3) *and $c(x, y)$ are convex and C^1 in y. Then*

$$g(x, y) - V(x) \leq -\langle u, c(x, y)\rangle \forall (x, y, u) \in X \times R^m \times R^d \text{ such that}$$
$$u \geq 0, c(x, y) \leq 0, \nabla_y g(x, y) + \nabla c(x, y)^t u = 0.$$

Remark 6.3. Propositions 5.1 and 6.1 and Corollary 6.2 show that any condition ensuring that the standard gap function or $g(x, y) - V(x)$ provides exact penalty terms for the equivalent single level problems GS and VS, respectively, ensure that $-\langle u, c(x, y)\rangle$ is an exact penalty function for the equivalent single level problem KS. The converse is not necessarily true.

Under assumptions involving continuous subanalytic functions, Luo et al. proved in [19] that there exists a constant $N > 0$ such that $(-\langle u, c(x, y)\rangle)^{1/N}$ is an exact penalty term for KS. Moreover, for the case where the mapping $F(x, y)$ is affine and the feasible region is compact, N can be taken as 1 or 2 depending on whether or not the strict complementarity condition is satisfied. To compare our results with those in [19], we summarize the related results in [19].

THEOREM 6.4 (see Theorems 4 and 6 of [19]). *Consider GBLP where the lower level problem is QP_x. Assume that $f(x, y)$ is Lipschitz continuous in both variables and that the set*

$$\{(x, y) \in X \times R^m : Ax + By - b \leq 0\}$$

is a compact polyhedron. Suppose GBLP has a solution. Then there exist positive scalars μ^* and β such that for all scalars $\mu \geq \mu^*$, any vector (x^*, y^*) solves GBLP *if and only if for some* $u^* \in R^d$*, the triple* (x^*, y^*, u^*) *solves the following penalized problem in the variables* (x, y, u)*:*

$$\begin{aligned}
&\min f(x,y) + \mu\sqrt{-\langle u, Ax+By-b\rangle}\\
&s.t.\ Px + Qy + q + B^T u = 0,\\
&\quad u \geq 0, \|u\| \leq \beta,\\
&\quad Ax + By - b \leq 0,\\
&\quad x \in X, y \in R^m.
\end{aligned}$$

Furthermore, if the strict complementarity condition is satisfied for all (x, y, u) *in the feasible region of KS, then we can remove the square root.*

In the following result we relax most of the assumptions of Theorem 6.4 but we require stronger conditions on $F(x, y)$.

THEOREM 6.5. *Consider GBLP where* $f(x, y)$ *is Lipschitz continuous in* y *uniformly in* $x \in R^n$ *with constant* L*, and* $c(x, y)$ *is convex and differentiable in* y*. Suppose that there exists a solution to GBLP.*

If $F(x, y)$ *is pseudostrongly monotone with respect to* y *uniformly in* $x \in X$ *with modulus* δ*, then any vector* (x^*, y^*) *is a global solution to GBLP if and only if for some* $u^* \in R^d$*, the triple* (x^*, y^*, u^*) *is a global solution to the following penalized problem in the variables* (x, y, u)*:*

$$\begin{aligned}
&\min f(x,y) + \mu\sqrt{-\langle u, c(x,y)\rangle}\\
&s.t.\ F(x,y) + \nabla_y c(x,y)^T u = 0,\\
&\quad u \geq 0, c(x,y) \leq 0,\\
&\quad x \in X, y \in R^m,
\end{aligned}$$

for all $\mu \geq \frac{\sqrt{\delta}}{\delta} L$*.*

Under the assumptions of Propositions 3.4 *and* 3.5*, any vector* (x^*, y^*) *is a global solution to GBLP if and only if for some* $u^* \in R^d$*, the triple* (x^*, y^*, u^*) *is a global solution to the following penalized problem in the variables* (x, y, u)*:*

$$\begin{aligned}
&\min f(x,y) - \delta\mu\langle u, c(x,y)\rangle\\
&s.t.\ F(x,y) + \nabla_y c(x,y)^T u = 0,\\
&\quad u \geq 0, c(x,y) \leq 0,\\
&\quad x \in X, y \in R^m,
\end{aligned}$$

for all $\mu \geq L$*, where* δ *is the modulus of the uniformly weak sharp minimum.*

Proof. We only prove the first assertion, since the proof of the second is similar. Assume (x^*, y^*, u^*) is a global solution of CS. Then (x^*, y^*) is a global solution of GBLP. By Proposition 3.12, $\sqrt{G_0(x, y)}$ is a uniform parametric error bound with modulus $\frac{\sqrt{\delta}}{\delta}$. Therefore, by Theorem 2.6, (x^*, y^*) is a global solution of $\mathrm{RP}_{\frac{\sqrt{\delta}}{\delta}\mu}$ with $r(x, y) = \sqrt{G_0(x, y)}$ for all $\mu \geq L$. Therefore,

$$\begin{aligned}
f(x^*, y^*) &\leq f(x,y) + \frac{\sqrt{\delta}}{\delta}\mu\sqrt{G_0(x,y)} \qquad \forall x, y \text{ s.t. } c(x,y) \leq 0,\\
&\leq f(x,y) + \frac{\sqrt{\delta}}{\delta}\mu\sqrt{-\langle u, c(x,y)\rangle} \qquad \forall (x,y,u) \text{ s.t.}\\
&\qquad u \geq 0, c(x,y) \leq 0, F(x,y) + \nabla_y c(x,y)^t u = 0,
\end{aligned}$$

where the last inequality follows from Proposition 6.1.

The proof of the converse is similar to the converse part in the proof of Proposition 2.2. ∎

Even when $c(x,y)$ is convex and C^1 in y, the ranges of applications of Theorems 6.4 and 6.5 are different. Indeed, the following example, taken from [19], is a situation where Theorem 6.5 is applicable but Theorem 6.4 is not.

Example 6.6. Consider the problem:

$$\begin{aligned} \text{(P2)} \quad & \min x - y \\ & \text{s.t. } x \geq 0, \text{ and } y \in \arg\min \left\{ \frac{1}{2} y^2 : x + y \geq 0, y \geq 0 \right\}. \end{aligned}$$

In [19], by direct arguments, (P2) is shown to be equivalent to the penalized problem

$$\begin{aligned} \text{(P3)} \quad & \min x - y + \mu \sqrt{u(x+y)} \\ & \text{s.t. } y - u = 0, x + y \geq 0, \\ & (x, y, u) \geq 0, \end{aligned}$$

for any $\mu > 0$. Indeed, it is easy to see that $(0,0)$ is the unique solution to the problem (P2) and $(0,0,0)$ is the unique solution of the penalized problem (P3) for any $\mu > 0$. It is also observed in [19] that Theorem 6.4 is not applicable because the feasible region is not compact. On the other hand, since $F(x,y) = y$ is strongly monotone with respect to y for all $x \in R^n$, this example does satisfy all the conditions of Theorem 6.5. By Theorem 6.5 both $\sqrt{G_0(x,y)}$ and the square root of the complementarity term are exact penalty terms. The standard gap function in this case is $G_0(x,y) = y^2$ for all $x \geq 0$. Therefore, (P2) is equivalent to both (P3) and the following problem:

$$\begin{aligned} \text{(P4)} \quad & \min x - y + \mu |y| \\ & \text{s.t. } x + y \geq 0, x \geq 0, y \geq 0. \end{aligned}$$

Indeed, it is easy to see that $(0,0)$ is the unique solution of (P4).

Now we discuss an example to which both the KKT and the non-KKT approaches apply, but yield different equivalent single level problems.

Example 6.7. Consider the problem

$$\begin{aligned} \text{(P5)} \quad & \min x + y_1 + y_2 \\ & \text{s.t. } a \leq x \leq b, \\ & (y_1, y_2) \in \arg\min_{y_1, y_2} \left\{ \frac{1}{2} y_1^2 + x y_1 + y_2, : \frac{1}{2} y_1 + x \geq 0, y_1 \geq 0, y_2 \geq 0 \right\}, \end{aligned}$$

where a and b are positive constants. It is obvious that $S(x) = \{(0,0)\}$ and $V(x) = 0$ for all $x \geq 0$. $F(x,y) = (y_1 + x, 1)$ is not pseudostrongly monotone. Therefore, the assumptions of Propositions 3.9, 3.11, and 3.13 are not satisfied. However, one can verify that the assumptions of Proposition 3.5 are satisfied. Therefore, for any $\mu > 0$, (x^*, y^*) is a solution of the original problem (P5) if and only if it is the solution of the following problem (by the value function approach):

$$\begin{aligned} \text{(P6)} \quad & \min x + y_1 + y_2 + \mu \left(\frac{1}{2} y_1^2 + x y_1 + y_2 \right) \\ & \text{s.t. } \frac{1}{2} y_1 + x \geq 0, a \leq x \leq b, y_1 \geq 0, y_2 \geq 0. \end{aligned}$$

By Theorem 6.5, there exists $u^* \in R^3$ such that (x^*, y^*, u^*) is a solution of the following problem:

$$
\begin{aligned}
\text{(P7)}\quad & \min x + y_1 + y_2 + \mu\left(u_1\left(\frac{1}{2}y_1 + x\right) + u_2 y_1 + u_3 y_2\right) \\
& \text{s.t. } 0 = y_1 + x - \frac{1}{2}u_1 - u_2, \\
& \quad\quad 0 = 1 - u_3, \\
& \quad\quad a \le x \le b, \frac{1}{2}y_1 + x \ge 0, y_1 \ge 0, y_2 \ge 0, u_1 \ge 0, u_2 \ge 0
\end{aligned}
$$

for any $\mu > 0$. Clearly, (P5) and (P6) have a unique solution $(a, 0, 0)$, and (P7) has a unique solution $(a, 0, 0, 0, 0, 1)$. Note that the compactness of the feasible region and the strict complementarity assumptions of Theorem 6.4 fail for this example.

Examples 6.6 and 6.7 illustrate that both the KKT and the non-KKT approaches have their advantages and disadvantages. On one hand, by the KKT approach, the exact penalty term is an explicit function of the problem data, but the number of variables in the single level problem increases. On the other hand, by the non-KKT approach, although the number of variables stays the same in the equivalent single level problem, the exact penalty function needs to be computed.

Acknowledgment. We thank J. S. Treiman for his comments on an early version of this paper, which helped to improve the exposition.

REFERENCES

[1] G. Anandalingam and T. L. Friesz, eds., *Hierarchical optimization*, Ann. Oper. Res., 34 (1992).

[2] J. V. Burke, *Calmness and exact penalization*, SIAM J. Control Optim., 29 (1991), pp. 493–497.

[3] J. V. Burke, *An exact penalization viewpoint of constraint optimization*, SIAM J. Control Optim., 29 (1991), pp. 968–998.

[4] J. V. Burke and M. C. Ferris, *Weak sharp minima in mathematical programming*, SIAM J. Control Optim., 31 (1993), pp. 1340–1359.

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

[6] S. Dempe, *A necessary and sufficient optimality condition for bilevel programming problems*, Optimization, 25 (1992), pp. 341–354.

[7] J.-P. Dussault and P. Marcotte, *Conditions derégularité géométrique pour les inéquations variationnelles*, RAIRO Rech. Opér., 23 (1988), pp. 1–16.

[8] A. D. Ioffe, *Regular points of Lipschitz functions*, Trans. Amer. Math. Soc., 251 (1979), pp. 61–69.

[9] M. C. Ferris and O. L. Mangasarian, *Minimum principle sufficiency*, Math. Programming, 57 (1992), pp. 1–14.

[10] 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*, Math. Programming, 48 (1990), pp. 265–284.

[11] M. Fukushima, *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems*, Math. Programming, 53 (1992), pp. 99–110.

[12] P. T. Harker and J. S. Pang, *On the existence of optimal solutions to mathematical programs with equilibrium constraints*, Oper. Res. Lett., 7 (1988), pp. 61–64.

[13] P. T. Harker and J. S. Pang, *Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications*, Math. Programming, 48 (1990), pp. 161–220.

[14] D. W. Hearn, *The gap function of a convex program*, Oper. Res. Lett., 1 (1982), pp. 67–71.

[15] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*, Academic Press, New York, 1980.

[16] M. Labbé, P. Marcotte, and G. Savard, *A bilevel model of taxation and its application to optimal highway pricing*, preprint.

[17] P. Loridan and J. Morgan, *A theoretical approximation scheme for Stackelberg problems*, J. Optim. Theory Appl., 11 (1989), pp. 95–110.
[18] X. D. Luo and P. Tseng, *Conditions for a projection-type error bound for the linear complementarity problem to be global*, preprint.
[19] Z. Q. Luo, J. S. Pang, D. Ralph, and S.-Q. Wu, *Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints*, preprint.
[20] O. L. Mangasarian, *Nonlinear Programming*, McGraw-Hill, New York, 1969.
[21] O. L. Mangasarian, *A simple characterization of solution sets of convex programs*, Oper. Res. Lett., 7 (1988), pp. 21–26.
[22] O. L. Mangasarian and R. R. Meyer, *Nonlinear perturbation of linear programs*, SIAM J. Control Optim., 17 (1979), pp. 745–752.
[23] O. L. Mangasarian and J. Ren, *New improved error bounds for the linear complementarity problem*, Math. Programming, 66 (1994), pp. 241–255.
[24] R. Mathias and J. S. Pang, *Error bounds for the linear complementarity problem with a $P-$matrix*, Linear Algebra Appl., 132 (1990), pp. 123–136.
[25] 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, Canada, CRT-920, 1992.
[26] J. V. Outrata, *Necessary optimality conditions for Stackelberg problems*, J. Optim. Theory Appl., 76 (1993), pp. 305–320.
[27] J. S. Pang, *A posteriori error bounds for the linearly-constrained variational inequality problem*, Math. Oper. Res., 12 (1987), pp. 474–484.
[28] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[29] H. Von Stackelberg, *The Theory of the Market Economy*, Oxford University Press, Oxford, England, 1952.
[30] J. J. Ye and D. L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization, 33 (1995), pp. 9–27.
[31] R. Zhang, *Problems of hierarchical optimization: Nonsmoothness and analysis of solutions*, Ph.D. thesis, University of Washington, Seattle, 1990.
[32] R. Zhang, *Problems of hierachical optimization in finite dimensions*, SIAM J. Optim., 4 (1994), pp. 521–536.