Nonlinear Analysis 75 (2012) 1202–1218

Contents lists available at SciVerse ScienceDirect

Nonlinear Analysis

journal homepage: www.elsevier.com/locate/na

On the Karush–Kuhn–Tucker reformulation of the bilevel optimization problem

S. Dempe*, A.B. Zemkoho

Institute of Numerical Mathematics and Optimization, Technical University Bergakademie Freiberg, 09596 Freiberg, Germany

ARTICLE INFO

Communicated by Ravi Agarwal

Keywords:
Bilevel optimization problem
KKT reformulation
Set-valued mapping
Basic CQ
Optimality conditions

ABSTRACT

This paper is mainly concerned with the classical KKT reformulation and the primal KKT reformulation (also known as an optimization problem with generalized equation constraint (OPEC)) of the optimistic bilevel optimization problem. A generalization of the MFCQ to an optimization problem with operator constraint is applied to each of these reformulations, hence leading to new constraint qualifications (CQs) for the bilevel optimization problem. M- and S-type stationarity conditions tailored for the problem are derived as well. Considering the close link between the aforementioned reformulations, similarities and relationships between the corresponding CQs and optimality conditions are highlighted. In this paper, a concept of partial calmness known for the optimal value reformulation is also introduced for the primal KKT reformulation and used to recover the M-stationarity conditions.

1. Introduction

We consider the optimistic bilevel programming problem to

$$\text{minimize } F(x, y) \text{ subject to } x \in X \subseteq \mathbb{R}^n,\ y \in S(x), \tag{1.1}$$

also called the upper level problem, where $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a continuously differentiable function and the set-valued mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, describes the solution set of the following parametric optimization problem also known as the lower level problem:

$$\text{minimize } f(x, y) \text{ subject to } y \in K(x), \tag{1.2}$$

where $K(x)$ is a closed subset of $\mathbb{R}^m$, for all $x \in X$, and the function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is twice continuously differentiable. We assume that the upper and lower level feasible sets are given as

$$X := \{x \in \mathbb{R}^n | G(x) \le 0\} \quad \text{and} \quad K(x) := \{y \in \mathbb{R}^m | g(x, y) \le 0\} \quad \text{for all } x \in X, \tag{1.3}$$

respectively; the functions $G : \mathbb{R}^n \to \mathbb{R}^k$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ being continuously and twice continuously differentiable, respectively. Also, unless otherwise stated, the functions $f(x, .)$ and $g_i(x, .)$, $i = 1, \dots, p$ are assumed to be convex for all $x \in X$. It is well known from convex optimization that the lower level problem would be equivalent to the parametric generalized equation:

$$0 \in \nabla_y f(x, y) + N_{K(x)}(y), \tag{1.4}$$

* Corresponding author.
E-mail addresses: dempe@math.tu-freiberg.de (S. Dempe), zemkoho@student.tu-freiberg.de (A.B. Zemkoho).

0362-546X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.na.2011.05.097

where $N_{K(x)}(y)$ denotes the normal cone (in the sense of convex analysis) to $K(x)$ at y, provided $y \in K(x)$, and $N_{K(x)}(y) := \emptyset$, otherwise.

Hence, the following one level reformulation of the bilevel program that we call *primal KKT reformulation* (this terminology may be justified by the fact that the above generalized equation can be considered as a compact form of the KKT conditions of the lower level problem):

$$\text{minimize } F(x,y)$$
$$\text{subject to } \begin{cases} G(x) \leq 0 \\ 0 \in \nabla_y f(x,y) + N_{K(x)}(y). \end{cases} \tag{1.5}$$

This problem, also corresponding to an optimization problem with (parametric) generalized equation constraint (OPEC), where K is a moving set, has recently been studied in [1].

Theorem 1.1. *The point $(\bar{x},\bar{y})$ is a local (resp. global) optimal solution of (1.1) if and only if $(\bar{x},\bar{y})$ is a local (resp. global) optimal solution of (1.5).*

This complete equivalence between problem (1.1) and its primal KKT reformulation (1.5) is lost if one considers the following detailed form of the normal cone to $K(x)$ at y:

$$N_{K(x)}(y) = \{\nabla_y g(x,y)^\top u | u \geq 0, u^\top g(x,y) = 0\},$$

which holds under a certain CQ; cf. [2, Theorem 4.3]. In fact, the resulting problem is the so-called KKT reformulation of the bilevel optimization problem:

$$\text{minimize } F(x,y)$$
$$\text{subject to } \begin{cases} G(x) \leq 0, \quad \mathcal{L}(x,y,u) = 0 \\ u \geq 0,\ g(x,y) \leq 0, \quad u^\top g(x,y) = 0, \end{cases} \tag{1.6}$$

where $\mathcal{L}(x,y,u) := \nabla_y f(x,y) + \nabla_y g(x,y)^\top u$. The relationship between the latter problem, that we call *classical KKT reformulation* in the sequel, and the bilevel program, in terms of optimal solutions, has recently been studied in [3]. This link can be summarized in the following result where the fulfillment of a CQ at $(\bar{x},\bar{y})$, say

$$[\nabla_y g(\bar{x},\bar{y})^\top \beta = 0, \beta \geq 0, \beta^\top g(\bar{x},\bar{y}) = 0] \Longrightarrow \beta = 0, \tag{1.7}$$

is necessary. Furthermore, $\Lambda(\bar{x},\bar{y})$ will denote the set of Lagrange multipliers of the lower level problem, i.e. the set of all the vectors u satisfying: $u \geq 0$, $u^\top g(\bar{x},\bar{y}) = 0$ and $\mathcal{L}(\bar{x},\bar{y},u) = 0$.

Theorem 1.2. *Let $(\bar{x},\bar{y})$ be a global (resp. local) optimal solution of (1.1) and assume that CQ (1.7) is satisfied at $(\bar{x},\bar{y})$. Then, for each $\bar{u} \in \Lambda(\bar{x},\bar{y})$, the point $(\bar{x},\bar{y},\bar{u})$ is a global (resp. local) optimal solution of (1.6). Conversely, let CQ (1.7) be satisfied at (x,y), for all $y \in S(x)$, $x \in X$ (resp. at $(\bar{x},\bar{y})$). Assume that $(\bar{x},\bar{y},\bar{u})$ is a global optimal solution (resp. local optimal solution for all $\bar{u} \in \Lambda(\bar{x},\bar{y})$) of (1.6). Then, $(\bar{x},\bar{y})$ is a global (resp. local) optimal solution of (1.1).*

Clearly, for $(\bar{x},\bar{y})$ to be a local optimal solution of (1.1), one needs to make sure that $(\bar{x},\bar{y},\bar{u})$ is a local optimal solution of problem (1.6), for all $\bar{u} \in \Lambda(\bar{x},\bar{y})$. In fact, an example of bilevel program was provided in [3], where $(\bar{x},\bar{y})$ fails to solve (1.1) locally, whereas $(\bar{x},\bar{y},\bar{u})$ is a local solution of (1.6), for all but one $\bar{u} \in \Lambda(\bar{x},\bar{y})$. This fact has motivated the following definition for the notion of optimality conditions for the bilevel optimization problem from the perspective of the KKT reformulation. As usually done in the literature on MPCCs (mathematical programs with complementarity constraints), we partition the set of indices of the functions involved in the complementarity slackness as

$$\eta := \eta(\bar{x},\bar{y},\bar{u}) := \{i | \bar{u}_i = 0, g_i(\bar{x},\bar{y}) < 0\}$$
$$\mu := \mu(\bar{x},\bar{y},\bar{u}) := \{i | \bar{u}_i = 0, g_i(\bar{x},\bar{y}) = 0\}$$
$$\nu := \nu(\bar{x},\bar{y},\bar{u}) := \{i | \bar{u}_i > 0, g_i(\bar{x},\bar{y}) = 0\}.$$

Definition 1.3. A point $(\bar{x},\bar{y})$ will be said to be *M-stationary* for the bilevel optimization problem (1.1) if there exists $(\alpha,\beta,\gamma) \in \mathbb{R}^{k+p+m}$ such that $\forall \bar{u} \in \Lambda(\bar{x},\bar{y})$:

$$\nabla_x F(\bar{x},\bar{y}) + \nabla G(\bar{x})^\top \alpha + \nabla_x g(\bar{x},\bar{y})^\top \beta + \nabla_x \mathcal{L}(\bar{x},\bar{y},\bar{u})^\top \gamma = 0 \tag{1.8}$$

$$\nabla_y F(\bar{x},\bar{y}) + \nabla_y g(\bar{x},\bar{y})^\top \beta + \nabla_y \mathcal{L}(\bar{x},\bar{y},\bar{u})^\top \gamma = 0 \tag{1.9}$$

$$\alpha \geq 0, \quad \alpha^\top G(\bar{x}) = 0 \tag{1.10}$$

$$\nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0, \qquad \beta_\eta = 0 \tag{1.11}$$

$$\forall i \in \mu, \quad (\beta_i > 0 \wedge \nabla_y g_i(\bar{x},\bar{y})\gamma > 0) \vee \beta_i(\nabla_y g_i(\bar{x},\bar{y})\gamma) = 0. \tag{1.12}$$

Conditions (1.8)–(1.12) are called the M-stationarity conditions for problem (1.1).

Definition 1.4. A point $(\bar{x}, \bar{y})$ will be said to be *S-stationary* for the bilevel optimization problem (1.1) if there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ such that $\forall \bar{u} \in \Lambda(\bar{x}, \bar{y})$: (1.8)–(1.11) and

$$\forall i \in \mu, \quad \beta_i \geq 0 \wedge \nabla_y g_i(\bar{x}, \bar{y})\gamma \geq 0. \tag{1.13}$$

Conditions (1.8)–(1.11) and (1.13) are called the S-stationarity conditions for (1.1).

Similarly, surrogates of other well-known types of stationarity concepts could also be defined for the bilevel optimization problem. In this paper though, we will focus our attention only on the above M- and S-types, since they are the most important ones. For the other ones, the interested reader is referred, for example, to [4].

A third possibility to write problem (1.1) as a one level optimization problem is the following optimal value reformulation:

$$\begin{aligned}&\text{minimize } F(x, y)\\ &\text{subject to } \begin{cases} f(x, y) \leq \varphi(x) \\ G(x) \leq 0, \qquad g(x, y) \leq 0 \end{cases}\end{aligned} \tag{1.14}$$

where φ is the optimal value function of the lower level problem, defined as

$$\varphi(x) := \min\{f(x, y) | y \in K(x)\}.$$

The latter problem is globally and locally equivalent to the bilevel programming problem (1.1). For an extensive review on constraint qualifications and optimality conditions for the optimal value reformulation, the interested reader is referred to [5]. As far as an algorithmic approach is concerned, see [6]. It is worth mentioning that, in general, the optimality conditions of the bilevel program obtained via the optimal value reformulation and those derived via the KKT reformulation are not related [5,7].

Our main concerns in this paper are the classical and primal KKT reformulations of the bilevel optimization problem. First of all, let us recall that problem (1.6) is a special class of the mathematical programming problem with complementarity constraints (MPCC). Hence, in the literature, the bilevel programming problem has generally been considered as embedded in the MPCC family. But recently, as mentioned above (see Theorem 1.2), it was shown in [3] that the bilevel programming problem is not a special case of the MPCC. The main reason is, that a local optimal solution of problem (1.6) may happen not to be a local optimal solution of the initial bilevel programming problem [3].

One of the consequences of this widespread idea, that the bilevel optimization problem is a special case of the MPCC, is that, in the literature, very little attention has been given to dual optimality conditions for problem (1.6). Although, a great amount of work has been devoted to dual optimality conditions for the MPCC; see e.g. [4,8,9]. With the observation in [3], it seems necessary that a specific study be devoted to (1.6), taking into account its particularities. This is one of the interests of the current paper.

As far as primal optimality conditions for problem (1.6) are concerned, a number of interesting results can be found in [10–13]. In particular, for the special class of linear bilevel programming problem, polynomial-time verifiable primal optimality conditions of problem (1.6) were given in [13], something which is unusual in bilevel programming [11]. Studying dual form or KKT-type optimality conditions for problem (1.6), and for the MPCC, in general, the main concern raised in the literature is the failure of most of the well-known CQs [4,8,14,15].

The first step of our approach consists of rewriting the feasible set of (1.6) as an *operator constraint* (this appellation is borrowed from [16]). The basic CQ, which can be seen as an extension of the MFCQ to a certain geometric constraint, is then applied to the *new* problem, and this leads to new CQs and the M- and S-type stationarity conditions (1.8)–(1.12) and (1.8)–(1.11), (1.13) respectively, for the problem (1.6). The M- and S-type stationarity for a local optimal solution of the bilevel optimization problem (1.1) are then deduced by means of Theorem 1.2. The first major advantage offered by the operator constraint formulation is the flexibility in choosing Ω, Λ, and ψ (see (3.1)), which helps circumvent the failure of the usual MFCQ.

Another interesting thing about the operator constraint formulation is that, it could allow one to study the (one level) reformulations of the bilevel optimization problem in a unified manner. This has already been the case for the optimal value reformulation, where it helped design new CQs [5]. In this paper, the primal KKT reformulation (1.5) has also been studied from this perspective. Considering the close link between the primal and the classical KKT reformulations (in the sense that when the normal cone in (1.5) is replaced by its expression given above, one obtains problem (1.6)), one of our concerns is to compare the two approaches, in terms of CQs and optimality conditions. Furthermore, we have introduced a concept of partial calmness for problem (1.5), substantially weaker than the initial basic-type CQ. This has also inspired new ideas of CQs for problem (1.5), and hence for the bilevel optimization problem.

In the next section of the paper, we present the basic tools (i.e. the Mordukhovich normal cone, subdifferential and coderivative) and their relevant properties. Some Lipschitzian properties of set-valued mappings will also be recalled. Section 3 is concerned with the optimization problem with operator constraint. We derive KKT-type optimality conditions for this problem, from a perspective different from the one already known in [2,16,17]. The partial calmness concept will also be introduced and characterized in this section. Sections 4 and 5 are mainly concerned with the applications of the results from Section 3. This has led to new CQs and M- and S-type optimality conditions for the bilevel programming problem.

We now introduce some notations that will simplify the presentation of the paper. Let $\mathbb{N}^*$ be the set of positive integers. For $l \in \mathbb{N}^*$, $\mathbb{R}^{n_1+\cdots+n_l} := \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_l}$ and $0_{n_1+\cdots+n_l}$ is the origin of the space $\mathbb{R}^{n_1+\cdots+n_l}$. For vectors $a^i \in \mathbb{R}^{n_i}$, $i = 1, \ldots, l$, the joint vector $(a^1, \ldots, a^{n_l})$ may be used instead of its transposed vector $(a^1, \ldots, a^{n_l})^\top$. Let $a^i \in \mathbb{R}^{n_i}$, $i = 1, 2$; $(a^1)^\top a^2 = \langle a^1, a^2\rangle$ is used for the inner product of a^1 and a^2. For two properties a and b, property $a \vee b$ means that either a or b is satisfied, whereas $a \wedge b$ denotes the fulfillment of a and b simultaneously. For $A \subseteq \mathbb{R}^l$, bd A denotes the topological boundary of A, while $d_A(a) = d(a, A)$ is the distance from the point a to A.

2. Basic definitions and background material

We first consider the Fréchet normal cone to a closed set $A \subseteq \mathbb{R}^l$ at some point $\bar{z} \in A$

$$\widehat{N}_A(\bar{z}) := \{z^* \in \mathbb{R}^l | \langle z^*, z - \bar{z}\rangle \leq o(\|z - \bar{z}\|)\ \forall z \in A\}.$$

This cone is known (see [18, Theorem 6.28]) to be the polar of the Bouligand tangent cone

$$T_A(\bar{z}) := \{z \in \mathbb{R}^l | \exists t_k \downarrow 0, z_k \to z : \bar{z} + t_k z_k \in A\}.$$

The Mordukhovich normal cone to A at $\bar{z}$ is the Kuratowski–Painlevé upper limit [16] of the Fréchet normal cone, i.e.

$$N_A(\bar{z}) := \{z^* \in \mathbb{R}^l | \exists z_k^* \to z^*, z_k \to \bar{z}(z_k \in A) : z_k^* \in \widehat{N}_A(z_k)\}.$$

The Mordukhovich subdifferential can also be defined from the Fréchet subdifferential as in the case of the normal cone. But in order to save some space, we simply consider the well-known interplay between most of the normal cone and subdifferential objects in the literature. That is, for a lower semicontinuous function $f : \mathbb{R}^l \to \overline{\mathbb{R}}$, the Mordukhovich subdifferential of f at some point $\bar{z} \in \operatorname{dom} f$ is

$$\partial f(\bar{z}) := \{z^* \in \mathbb{R}^l | (z^*, -1) \in N_{\operatorname{epi} f}(\bar{z}, f(\bar{z}))\},$$

where epi f denotes the epigraph of f. It is important to mention that if f is a continuously differentiable function, then ∂f coincides with the gradient of f. Considering two functions, the sum and chain rules are obtained respectively as:

Theorem 2.1 *([19, Corollary 4.6]). Let the functions $f, g : \mathbb{R}^l \to \mathbb{R}$ be locally Lipschitz continuous around $\bar{z}$. Then*

$$\partial(f + g)(\bar{z}) \subseteq \partial f(\bar{z}) + \partial g(\bar{z}).$$

Equality holds if f or g is continuously differentiable.

Theorem 2.2 *([20, Proposition 2.10]). Let $f : \mathbb{R}^{l_1} \to \mathbb{R}^{l_2}$ be Lipschitz continuous around $\bar{z}$, and $g : \mathbb{R}^{l_2} \to \overline{\mathbb{R}}$ Lipschitz continuous around $\overline{w} = f(\bar{z}) \in \operatorname{dom} g$. Then*

$$\partial(gof)(\bar{z}) \subseteq \bigcup[\partial\langle w^*, f(\bar{z})\rangle : w^* \in \partial g(\overline{w})].$$

In the next result, we recall the necessary optimality condition for a Lipschitz optimization problem with geometric constraint.

Theorem 2.3 *([18, Theorem 8.15]). We let $f : \mathbb{R}^l \to \mathbb{R}$ be a locally Lipschitz continuous function and $A \subseteq \mathbb{R}^l$, a closed set. For $\bar{z}$ to be a local minimizer of f on A, it is necessary that*

$$0 \in \partial f(\bar{z}) + N_A(\bar{z}).$$

For a set-valued mapping $\Phi : \mathbb{R}^{l_1} \rightrightarrows \mathbb{R}^{l_2}$, a derivative-like object, called coderivative, and introduced by Mordukhovich (see [16]), can also be defined. Let $(\bar{u}, \bar{z}) \in \operatorname{gph} \Phi$, the coderivative of Φ at $(\bar{u}, \bar{z})$ is a positively homogeneous set-valued mapping $D^*\Phi(\bar{u}, \bar{z}) : \mathbb{R}^{l_2} \rightrightarrows \mathbb{R}^{l_1}$, such that for any $z^* \in \mathbb{R}^{l_2}$, we have

$$D^*\Phi(\bar{u}, \bar{z})(z^*) := \{u^* \in \mathbb{R}^{l_1} | (u^*, -z^*) \in N_{\operatorname{gph}\Phi}(\bar{u}, \bar{z})\}.$$

If Φ reduces to a single-valued Lipschitz continuous function, then $\bar{z}$ can be omitted and the coderivative of Φ reduces to $D^*\Phi(\bar{u})(z^*) = \partial\langle z^*, \Phi\rangle(\bar{u})$, for all $z^* \in \mathbb{R}^{l_2}$; with $\langle z^*, \Phi\rangle(\bar{u}) := \langle z^*, \Phi(\bar{u})\rangle$ and ∂ being the basic subdifferential defined above. It clearly follows that, in case Φ is single-valued and continuously differentiable, then $D^*\Phi(\bar{u})(z^*) = \{\nabla\Phi(\bar{u})^\top z^*\}$, for all $z^* \in \mathbb{R}^{l_2}$; where $\nabla\Phi(\bar{u})$ denotes the Jacobian matrix of Φ.

The set-valued mapping Φ is said to be Lipschitz-like at $(\bar{u}, \bar{z}) \in \operatorname{gph} \Phi$, if there exist neighborhoods U of $\bar{u}$, V of $\bar{z}$, and a number $L > 0$ such that

$$\Phi(u) \cap V \subseteq \Phi(u') + L\|u - u'\|\mathbb{B}, \quad \forall u, u' \in U. \tag{2.1}$$

This property, often called the Aubin property, was introduced by Aubin [21] and also studied for example in [16,18,20]. It is worth mentioning that the Aubin property is a natural extension of the Lipschitz continuity known for a single-valued function. If we fix $u' := \bar{u}$ in (2.1), then we obtain the following inclusion

$$\Phi(u) \cap V \subseteq \Phi(\bar{u}) + L\|u - \bar{u}\|\mathbb{B}, \quad \forall u \in U, \tag{2.2}$$

which defines the calmness or upper pseudo-Lipschitz continuity of the set-valued mapping Φ. Hence, it is obvious that the Aubin property implies the calmness property. As shown in the next theorem, the calmness property may be very useful in computing the normal cone to a subset of $\mathbb{R}^l$ defined by finitely many inequalities and equalities.

Theorem 2.4 (*[22, Theorem 5]*). *We consider the set*

$$A := \{z \in \mathbb{R}^l | g(z) \leq 0, h(z) = 0\},$$

where $g : \mathbb{R}^l \to \mathbb{R}^{l_1}$ and $h : \mathbb{R}^l \to \mathbb{R}^{l_2}$ are continuously differentiable functions. Then

$$N_A(\bar{z}) = \{\nabla g(\bar{z})^\top \lambda + \nabla h(\bar{z})^\top \mu | \lambda \geq 0, \lambda^\top g(\bar{z}) = 0\},$$

for $\bar{z} \in A$, provided the following set-valued mapping is calm at $(0, 0, \bar{z})$

$$\mathcal{M}(t_1, t_2) := \{z \in \mathbb{R}^l | g(z) + t_1 \leq 0, h(z) + t_2 = 0\}.$$

3. Optimization problem with operator constraint

The optimization problem with operator constraint, which may be seen as a special optimization problem with geometric constraint is

$$\text{minimize } F(z) \text{ subject to } z \in \Omega \cap \psi^{-1}(\Lambda), \tag{3.1}$$

where $F : \mathbb{R}^l \to \mathbb{R}$ and $\psi : \mathbb{R}^l \to \mathbb{R}^{l_1}$ are locally Lipschitz continuous functions, and the sets $\Omega \subseteq \mathbb{R}^l$, $\Lambda \subseteq \mathbb{R}^{l_1}$ are closed. To derive KKT-type dual optimality conditions for problem (3.1), we consider the basic CQ. Let $\bar{z}$ be a feasible point of problem (3.1); the basic CQ, which may have been introduced in [23] and studied for example in [2,18,24], is said to be satisfied at $\bar{z}$ if

$$\left.\begin{array}{l} 0 \in \partial\langle u^*, \psi\rangle(\bar{z}) + N_\Omega(\bar{z}) \\ u^* \in N_\Lambda(\psi(\bar{z})) \end{array}\right\} \Longrightarrow u^* = 0. \tag{3.2}$$

Some CQs closely related to the basic CQ have also been studied in [25].

Remark 3.1. If $\Omega := \mathbb{R}^l$, $\Lambda := \mathbb{R}^{l_2}_- \times \{0_{l_1 - l_2}\}$ and ψ is a continuously differentiable function, then the basic CQ coincides with the dual form of the well-known MFCQ; cf. [26]. Hence, the basic CQ is a generalization of the MFCQ to the optimization problem with operator constraint.

We consider the following perturbation map of the operator constraint:

$$\Psi(u) := \{z \in \Omega | \psi(z) + u \in \Lambda\}. \tag{3.3}$$

The next lemma, which is a consequence of [19, Theorem 6.10] and [24, Corollary 4.2], shows that this set-valued mapping is Lipschitz-like at $(0, \bar{z}) \in \operatorname{gph} \Psi$, if the basic CQ is satisfied at $\bar{z}$.

Lemma 3.2. *Assume that $\bar{z} \in \Psi(0)$. Then*

$$D^*\Psi(0, \bar{z})(z^*) \subseteq \{u^* \in N_\Lambda(\psi(\bar{z})) | -z^* \in \partial\langle u^*, \psi\rangle(\bar{z}) + N_\Omega(\bar{z})\}.$$

If in addition, the basic CQ is satisfied at $\bar{z}$, then Ψ is Lipschitz-like at $(0, \bar{z})$.

Since we are only interested in designing optimality conditions for local optimal solutions, it is necessary to show that the Lipschitz-like and calmness properties defined in the previous section are locally preserved for the set-valued mapping Ψ.

Lemma 3.3. *Let $\bar{z} \in \Psi(\bar{u})$ and let V be a neighborhood of $\bar{z}$. If Ψ is calm (resp. Lipschitz-like) at $(\bar{u}, \bar{z})$, then the set-valued mapping*

$$\Psi_V(u) := \{z \in \Omega \cap V | \psi(z) + u \in \Lambda\}$$

is also calm (resp. Lipschitz-like) at $(\bar{u}, \bar{z})$.

Proof. Let us set $\psi_u(z) := \psi(z) + u$. Then we have

$$\Psi(u) = \Omega \cap \psi_u^{-1}(\Lambda) \quad \text{and} \quad \Psi_V(u) = V \cap \Omega \cap \psi_u^{-1}(\Lambda),$$

and the result follows from the definition of calmness (2.2) (resp. Lipschitz-like property (2.1)), by noting that for $A, B, C \subseteq \mathbb{R}^l$, $A \subseteq B$ implies $A \cap C \subseteq B \cap C$. $\square$

We are now ready to state a KKT-type optimality condition for problem (3.1) under the basic CQ. The technique utilized in the proof is inspired by [27, Lemma 3.1], in the framework of OPECs. For the latter class of optimization problems, this approach has also been used in [16,28]. The statement of this result and many proofs exist in the literature (cf. [2,16–18,29]), but we were unable to find any reference where the multiplier u^* is bounded, except as already mentioned, for the OPEC. Hence, the reason why we include the proof here.

Proposition 3.1. *Let $\bar{z}$ be a local optimal solution of problem (3.1). Assume that the basic CQ is satisfied at $\bar{z}$. Then, there exists $\mu > 0$ such that for any $r \geq \mu$, one can find $u^* \in r\mathbb{B} \cap N_\Lambda(\psi(\bar{z}))$ such that*

$$0 \in \partial F(\bar{z}) + \partial\langle u^*, \psi\rangle(\bar{z}) + N_\Omega(\bar{z}). \tag{3.4}$$

Proof. Let $\bar{z}$ be an optimal solution of problem (3.1) in the closed neighborhood V of $\bar{z}$. It follows from Lemmas 3.2 and 3.3 that Ψ_V is Lipschitz-like at $(0, \bar{z})$, under the basic CQ (3.2). Denote by L_V and L_F the Lipschitz modulus of Ψ_V and F, respectively. We claim that for any $r \geq \mu := L_V L_F$, the point $(0, \bar{z})$ is a local optimal solution of problem

$$\text{minimize } F_r(u, z) \text{ subject to } (u, z) \in \operatorname{gph} \Psi_V, \tag{3.5}$$

with $F_r(u, z) := F(z) + r\|u\|$.

In fact, Ψ_V being Lipschitz-like at $(0, \bar{z})$ implies, by definition, that there exist neighborhoods U_1 of 0 and V_1 of $\bar{z}$ such that $\forall u \in U_1,\ \forall z \in \Psi(u) \cap V_1$, there exists $z^1 \in \Psi_V(0)$ with

$$\|z - z^1\| \leq L_V\|u\|. \tag{3.6}$$

In addition, $F(\bar{z}) \leq F(z^1)$, given that $z^1 \in \Psi_V(0) = \Omega \cap V \cap \psi^{-1}(\Lambda)$. Now, let $r \geq \mu$ and $(u, z) \in \operatorname{gph} \Psi_V \cap (U_1 \times V_1)$, then we have

$$\begin{aligned} F_r(0, \bar{z}) &= F(\bar{z}) \leq [F(z^1) - F(z)] + F(z) \\ &\leq L_F\|z - z^1\| + F(z) \quad \text{(cf. Lipschitz continuity of } F) \\ &\leq L_V L_F\|u\| + F(z) \leq F_r(u, z) \quad \text{(cf. inequality (3.6))}. \end{aligned}$$

Applying Theorem 2.3 to problem (3.5), we have

$$(0, 0) \in r\mathbb{B} \times \partial F(\bar{z}) + N_{\operatorname{gph} \Psi_V}(0, \bar{z}).$$

Hence, there exist $u^* \in r\mathbb{B}$ and $z^* \in \partial F(\bar{z})$ such that $(-u^*, -z^*) \in N_{\operatorname{gph} \Psi_V}(0, \bar{z})$. It follows from the definition of the coderivative, and the inclusion of Lemma 3.2, that

$$-u^* \in N_\Lambda(\psi(\bar{z})) \quad \text{and} \quad -z^* \in \partial\langle -u^*, \psi\rangle(\bar{z}) + N_\Omega(\bar{z}),$$

which concludes the proof. □

Remark 3.4. Under the setting of Remark 3.1, the basic CQ coincides with the usual MFCQ. Further, it is well known that under the MFCQ, the set of Lagrange multipliers is bounded. But the bound is not usually provided with the classical technique to derive KKT conditions via the MFCQ. Hence, the interesting feature of the approach in Proposition 3.1.

Remark 3.5. One can easily check that the above result remains valid if the basic CQ is replaced by the weaker calmness of the set-valued mapping Ψ. The optimality condition (3.4) also follows from [27, Theorem 3.1], where one has to set $\Phi(z) := -\psi(z) + \Lambda$. However, the approach in [27] does not allow us to detect the fact that u^* also belongs to $N_\Lambda(\psi(\bar{z}))$, which is an important component of Proposition 3.1, regarding the structure of problem (3.1). Furthermore, as it will be clear in the next sections of the paper, the inclusion $u^* \in N_\Lambda(\psi(\bar{z}))$ plays an important role in the applications.

For the rest of this section, we assume that ψ is a real-valued function and $\Lambda = \mathbb{R}_-$. Then the optimization problem with operator constraint takes the form

$$\text{minimize } F(z) \text{ subject to } z \in \Omega, \qquad \psi(z) \leq 0. \tag{3.7}$$

Here, $\Omega := \{z | g(z) \leq 0, h(z) = 0\}$, with g and h being some given continuous functions. Next, we recall the definition of the concept of *partial calmness* for problem (3.7), as introduced in [30] in the framework of the optimal value reformulation of the bilevel program. The term *partial* as opposed to the stronger notion of calmness by Clarke [31, Definition 6.4.1] refers to the fact that only some of the constraints (ψ in our case) are perturbed. Since we will be dealing with the classical and primal KKT reformulations of (1.1), we will use the terminology of *ψ-partial calmness* in order to differentiate between the two values to be taken by ψ in the corresponding reformulation.

Definition 3.6. Let $\bar{z}$ be a local optimal solution of problem (3.7). Problem (3.7) is ψ-partially calm at $\bar{z}$ if there is a neighborhood U of $(0, \bar{z})$ and a number $\lambda > 0$ such that

$$F(z) - F(\bar{z}) + \lambda|t| \geq 0 \quad \forall (t, z) \in U : \ z \in \Omega, \qquad \psi(z) + t \leq 0.$$

Broadly speaking, the latter concept was tailored to move a *disturbing constraint* (in the sense of leading to the failure of a CQ) to the objective function. This corresponds in the case of problem (3.7), to the following exact penalization, where $z \to \psi(z)_+$ (with $\psi(z)_+ = \max\{0, \psi(z)\}$) represents the penalty function and λ the penalty coefficient.

Theorem 3.7 (*[15, Proposition 2.2]*). *Let $\bar{z}$ be a local optimal solution of problem (3.7). Problem (3.7) is ψ-partially calm at $\bar{z}$ if and only if there exists a number $\lambda > 0$ such that $\bar{z}$ is a local optimal solution of the problem to*

$$\text{minimize } F(z) + \lambda\psi(z)_+ \text{ subject to } z \in \Omega.$$

To conclude this section, let us mention a sufficient condition for problem (3.7) to be ψ-partially calm. The proof can be found in [31] or [5, Theorem 4.10].

Proposition 3.2. *Let $\bar{z}$ be a local optimal solution of problem (3.7) such that the set-valued mapping Ψ in (3.3) (with Ω and Λ defined in (3.7)) is calm at $(0, \bar{z})$. Then, problem (3.7) is ψ-partially calm at $\bar{z}$.*

4. The classical KKT reformulation

4.1. M-type optimality conditions

Deriving the classical KKT reformulation from the generalized equation (1.4), it is clear that if $K(x) := \mathbb{R}^m$, the MFCQ remains applicable. Otherwise, if problem (1.6) is considered as a usual nonlinear optimization problem, it would fail at any feasible point; cf. [8,4,14,15]. However, the basic CQ, which can be seen as a generalization of the MFCQ (cf. Remark 3.1) may well still be applied provided the feasible set is written differently. To motivate our discussion, we recall that the failure of the MFCQ is due to the following complementarity system

$$u \geq 0,\ g(x, y) \leq 0, \quad u^\top g(x, y) = 0. \tag{4.1}$$

We show in the next example that the basic CQ is applicable to problem (1.6), if we assume that the function g is linear in (x, y), and the feasible set is reformulated as an operator constraint; with $\psi(x, y, u) = (G(x), \mathcal{L}(x, y, u))$, $\Lambda = \mathbb{R}^k_- \times \{0_m\}$ and Ω denoting the set of (x, y, u) solving the complementarity problem (4.1).

Example 4.1. We consider the bilevel optimization problem to

$$\text{minimize } x^2 + y^2 \text{ subject to } x \geq 0,\ y \in S(x) := \arg\min\{xy + y | y \geq 0\}.$$

One can easily check that $(0, 0)$ is the optimal solution of the above problem. The classical KKT reformulation of this problem is:

$$\text{minimize } x^2 + y^2$$
$$\text{subject to } \begin{cases} x \geq 0, \quad x - u + 1 = 0 \\ u \geq 0,\ y \geq 0, \quad uy = 0. \end{cases}$$

It is obvious that the lower level multiplier corresponding to the optimal solution is $\bar{u} = 1$; and hence that the MFCQ fails to hold at $(0, 0, 1)$. We are now going to show that the basic CQ is satisfied if we set $\psi(x, y, u) = (-x, x - u + 1)$, $\Lambda = \mathbb{R}_- \times \{0\}$ and $\Omega = \{(x, y, u) \in \mathbb{R}^3 | y \geq 0, u \geq 0, yu = 0\}$. For some point $(\alpha, \beta) \in N_\Lambda(\psi(0, 0, 1))$, i.e. $(\alpha, \beta) \in \mathbb{R}_+ \times \mathbb{R}$, $(0, 0, 0) \in \langle \nabla\psi(0, 0, 1), (\alpha, \beta)\rangle + N_\Omega(0, 0, 1)$ if and only if $\alpha - \beta = 0$ and $(0, -\beta) \in N_\Theta(0, 1)$ (with $\Theta := \{(y, u) \in \mathbb{R}^2 | y \geq 0, u \geq 0, yu = 0\}$). It follows from Lemma 4.1 that $\beta = 0$ and hence that $\alpha = 0$. This shows that the basic CQ holds at $(0, 0, 1)$.

By setting $v := -g(x, y)$ and hence introducing a new (dummy) variable in the problem, the idea in the above example can be extended to the more general problem (1.6). The technicality behind this is that the new constraint $g(x, y) + v = 0$ is moved to the function ψ and thus allowing just the computation of the normal cone to the polyhedral set

$$\Theta := \{(u, v) \in \mathbb{R}^{2p} | u \geq 0, v \geq 0, u^\top v = 0\}, \tag{4.2}$$

which is possible without any qualification condition.

Lemma 4.1 (*[32, Proposition 2.1]*). *Let $(\bar{u}, \bar{v}) \in \Theta$, then*

$$N_\Theta(\bar{u}, \bar{v}) = \left\{ (u^*, v^*) \in \mathbb{R}^{2p} : \begin{array}{ll} u_i^* = 0 & \forall i : \bar{u}_i > 0 = \bar{v}_i \\ v_i^* = 0 & \forall i : \bar{u}_i = 0 < \bar{v}_i \\ (u_i^* < 0 \wedge v_i^* < 0) \vee u_i^* v_i^* = 0 & \forall i : \bar{u}_i = 0 = \bar{v}_i \end{array} \right\}.$$

Thanks to the aforementioned transformation, M-type stationarity conditions can be derived for problem (1.6).

Theorem 4.2. *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (1.6) and assume that the following CQ holds at $(\bar{x}, \bar{y}, \bar{u})$:*

$$\left.\begin{array}{l}\nabla G(\bar{x})^\top \alpha + \nabla_x g(\bar{x}, \bar{y})^\top \beta + \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0\\ \nabla_y g(\bar{x}, \bar{y})^\top \beta + \nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0\\ \alpha \geq 0, \quad \alpha^\top G(\bar{x}) = 0\\ \beta_\nu = 0, \qquad \nabla_y g_\eta(\bar{x}, \bar{y})\gamma = 0\\ \forall i \in \mu, \quad (\beta_i > 0 \wedge \nabla_y g_i(\bar{x}, \bar{y})\gamma > 0) \vee \beta_i(\nabla_y g_i(\bar{x}, \bar{y})\gamma) = 0\end{array}\right\} \Longrightarrow \begin{cases}\alpha = 0\\ \beta = 0\\ \gamma = 0.\end{cases} \tag{4.3}$$

Then, there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$, with $\|(\alpha, \beta, \gamma)\| \leq r$ (for some $r > 0$) such that the M-stationarity conditions are satisfied.

Proof. Let us set $\psi(x, y, u, v) = (G(x), g(x, y) + v, \mathcal{L}(x, y, u))$, $\Lambda = \mathbb{R}^k_- \times \{0_{p+m}\}$ and $\Omega = \mathbb{R}^{n+m} \times \Theta$. Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (1.6). One can easily verify that there is a vector $\bar{v}$ such that $(\bar{x}, \bar{y}, \bar{u}, \bar{v})$ is a local optimal solution of the problem to

$$\text{minimize } F(x, y) \text{ subject to } (x, y, u, v) \in \Omega \cap \psi^{-1}(\Lambda). \tag{4.4}$$

We have

$$N_\Omega(\bar{x}, \bar{y}, \bar{u}, \bar{v}) = \{0_{n+m}\} \times N_\Theta(\bar{u}, \bar{v}) \tag{4.5}$$

$$N_\Lambda(\psi(\bar{x}, \bar{y}, \bar{u}, \bar{v})) = \{(\alpha, \beta, \gamma) | \alpha \geq 0, \alpha^\top G(\bar{x}) = 0\} \tag{4.6}$$

$$\nabla \psi(x, y, u, v)^\top (\alpha, \beta, \gamma) = \begin{bmatrix} A(\alpha, \beta, \gamma)\\ \beta \end{bmatrix} \tag{4.7}$$

where

$$A(\alpha, \beta, \gamma) := \begin{bmatrix} \nabla G(\bar{x})^\top \alpha + \nabla_x g(\bar{x}, \bar{y})^\top \beta + \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma\\ \nabla_y g(\bar{x}, \bar{y})^\top \beta + \nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma\\ \nabla_y g(\bar{x}, \bar{y})\gamma \end{bmatrix}. \tag{4.8}$$

It follows from equalities (4.5)–(4.7) that the basic CQ applied to problem (4.4) at $(\bar{x}, \bar{y}, \bar{u}, \bar{v})$ can equivalently be formulated as follows: there is no nonzero vector $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ such that

$$\nabla G(\bar{x})^\top \alpha + \nabla_x g(\bar{x}, \bar{y})^\top \beta + \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \tag{4.9}$$

$$\nabla_y g(\bar{x}, \bar{y})^\top \beta + \nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \tag{4.10}$$

$$\alpha \geq 0, \quad \alpha^\top G(\bar{x}) = 0 \tag{4.11}$$

$$(-\nabla_y g(\bar{x}, \bar{y})\gamma, -\beta) \in N_\Theta(\bar{u}, \bar{v}). \tag{4.12}$$

By noting that $\bar{v}_i = -g_i(\bar{x}, \bar{y})$, for $i := 1, \ldots, p$, it follows from Lemma 4.1 that the basic CQ applied to problem (4.4) is equivalent to CQ (4.3). Hence, from Proposition 3.1 there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$, with $\|(\alpha, \beta, \gamma)\| \leq r$ (for some $r > 0$) such that (1.8)–(1.10) and (4.12) are satisfied, given that the objective function of problem (4.4) is independent of (u, v). The result then follows by interpreting inclusion (4.12), as already made above. □

The bound on the multiplier vector, usually neglected for MPCCs, can be explicitly given in terms of problem data; see the proof of Proposition 3.1. It may be important to mention that this bound can be very useful in developing an effective algorithm for problem (1.6), and hence for the bilevel optimization problem.

The technique used in the proof of Theorem 4.2, i.e. to transform the nonlinear complementarity problem in (4.1) into a linear one, has been used in various occasions, for the MPCC; see e.g. [9,32]. One can easily check that the M-stationarity conditions obtained here are identical to those in [27] or [33] under various CQs, among which CQ (b) of [27, Theorem 4.1] or (b) of [33, Theorem 5.1] coincides with the CQ in Theorem 4.2. But, it should be mentioned that in the latter case, this CQ is recovered from a perspective different from that of [27,33], where an enhanced generalized equation formulation of the KKT conditions of the lower level problem was used to design the CQ.

We now introduce a different way to choose ψ, Ω and Λ; that would lead to a new and weaker CQ allowing us to obtain the same optimality conditions as in Theorem 4.2. To proceed, let us recall that the complementarity system (4.1) is equivalent to

$$u_i \geq 0,\ g_i(x, y) \leq 0, \quad u_i g_i(x, y) = 0,\ i = 1, \ldots, p,$$

meaning that

$$(u_i, -g_i(x, y)) \in \Lambda_i := \{(a, b) \in \mathbb{R}^2 | a \geq 0, b \geq 0, ab = 0\}, \quad i = 1, \ldots, p.$$

Theorem 4.3. *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (1.6) and assume that the following assertions are satisfied:*

1. *The following set-valued mapping is calm at* $(0, 0, \bar{x}, \bar{y}, \bar{u})$
$$\mathcal{M}_1(t_1, t_2) := \{(x, y, u) | G(x) + t_1 \le 0, \mathcal{L}(x, y, u) + t_2 = 0\}$$
2. *The following implication holds at* $(\bar{x}, \bar{y}, \bar{u})$*:*
$$\left.\begin{array}{l} \nabla G(\bar{x})^\top \alpha + \nabla_x g(\bar{x}, \bar{y})^\top \beta + \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \\ \nabla_y g(\bar{x}, \bar{y})^\top \beta + \nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \\ \alpha \ge 0, \quad \alpha^\top G(\bar{x}) = 0 \\ \beta_\nu = 0, \qquad \nabla_y g_\eta(\bar{x}, \bar{y})\gamma = 0 \\ \forall i \in \mu, \quad (\beta_i > 0 \wedge \nabla_y g_i(\bar{x}, \bar{y})\gamma > 0) \vee \beta_i(\nabla_y g_i(\bar{x}, \bar{y})\gamma) = 0 \end{array}\right\} \Longrightarrow \begin{cases} \beta = 0 \\ \nabla_y g(\bar{x}, \bar{y})\gamma = 0. \end{cases} \tag{4.13}$$

Then, there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$*, with* $\|\beta\| \le r$ *(for some* $r > 0$*) such that the M-stationarity conditions are satisfied.*

Proof. We consider the set $\Omega = \{(x, y, u) | G(x) \le 0, \mathcal{L}(x, y, u) = 0\}$ and the function $\psi(x, y, u) = (u_i, -g_i(x, y))_{i=1,\ldots,p}$. If $(\bar{x}, \bar{y}, \bar{u})$ is a local optimal solution of problem (1.6), it means, in other words, that $(\bar{x}, \bar{y}, \bar{u})$ is a local optimal solution of the problem to

$$\text{minimize } F(x, y) \text{ subject to } (x, y, u) \in \Omega \cap \psi^{-1}(\Lambda), \tag{4.14}$$

where $\Lambda = \Lambda_1 \times \cdots \times \Lambda_p$, with $\Lambda_i = \{(a, b) \in \mathbb{R}^2 | a \ge 0, b \ge 0, ab = 0\}$ for $i = 1, \ldots, p$.

Applying Proposition 3.1 to (4.14), there exists a vector $(\delta, \beta) \in \mathbb{R}^{2p}$ with $\|(\delta, \beta)\| \le r$ (for some $r > 0$) such that

$$(\delta_i, \beta_i) \in N_{\Lambda_i}(\psi_i(\bar{x}, \bar{y}, \bar{u})), \quad i := 1, \ldots, p \tag{4.15}$$

$$(0, 0) \in \begin{bmatrix} \nabla F(\bar{x}, \bar{y}) \\ 0 \end{bmatrix} + \begin{bmatrix} -\nabla g(\bar{x}, \bar{y})^\top \beta \\ \delta \end{bmatrix} + N_\Omega(\bar{x}, \bar{y}, \bar{u}), \tag{4.16}$$

provided there is no nonzero vector $(\delta, \beta) \in \mathbb{R}^{2p}$ such that

$$(\delta_i, \beta_i) \in N_{\Lambda_i}(\psi_i(\bar{x}, \bar{y}, \bar{u})), \quad i := 1, \ldots, p \tag{4.17}$$

$$(0, 0) \in \begin{bmatrix} -\nabla g(\bar{x}, \bar{y})^\top \beta \\ \delta \end{bmatrix} + N_\Omega(\bar{x}, \bar{y}, \bar{u}). \tag{4.18}$$

It follows, under assumption 1 (see Theorem 2.4), that

$$N_\Omega(\bar{x}, \bar{y}, \bar{u}) = \{A(\alpha, \beta, \gamma) - (\nabla_x g(\bar{x}, \bar{y})^\top \beta, \nabla_y g(\bar{x}, \bar{y})^\top \beta, 0)^\top | \alpha \ge 0, \alpha^\top G(\bar{x}) = 0\},$$

where $A(\alpha, \beta, \gamma)$ denotes the matrix given in (4.8). Hence, either from (4.16) or from (4.18), it follows that there exists $\gamma \in \mathbb{R}^m$ such that $\delta = -\nabla_y g(\bar{x}, \bar{y})\gamma$; à fortiori, either (4.15) or (4.17) implies that there exists $\gamma \in \mathbb{R}^m$ such that $(-\nabla_y g_i(\bar{x}, \bar{y})\gamma, \beta_i) \in N_{\Lambda_i}(\psi_i(\bar{x}, \bar{y}, \bar{u}))$, $i := 1, \ldots, p$. The result then follows by noting that $N_{\Lambda_i}(\psi_i(\bar{x}, \bar{y}, \bar{u}))$ is obtained from Lemma 4.1. □

The approach in the above result is similar to the one used in [34], for a mathematical program with vanishing constraints.

Remark 4.4. (i) Assumptions 1 and 2 in the previous result are satisfied, provided CQ (4.3) holds at $(\bar{x}, \bar{y}, \bar{u})$. In fact, it is obvious that CQ (4.3) implies assumption 2. On the other hand, CQ (4.3) can equivalently be written as

$$\begin{aligned} A(\bar{x}, \bar{y}, \bar{u}) := \{(\alpha, \beta, \gamma) | & \nabla G(\bar{x})^\top \alpha + \nabla_x g(\bar{x}, \bar{y})^\top \beta + \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \\ & \nabla_y g(\bar{x}, \bar{y})^\top \beta + \nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \; \alpha \ge 0, \alpha^\top G(\bar{x}) = 0 \\ & \beta_\nu = 0, \nabla_y g_\eta(\bar{x}, \bar{y})\gamma = 0 \; \forall i \in \mu, (\beta_i > 0 \wedge \nabla_y g_i(\bar{x}, \bar{y})\gamma > 0) \vee \beta_i(\nabla_y g_i(\bar{x}, \bar{y})\gamma) = 0\} \\ = \{(0, 0, 0)\}. & \end{aligned}$$

Furthermore, one has

$$A(\bar{x}, \bar{y}, \bar{u}) \supseteq \{(\alpha, 0, \gamma) | \nabla G(\bar{x})^\top \alpha + \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \; \nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \; \alpha \ge 0, \alpha^\top G(\bar{x}) = 0\} := B(\bar{x}, \bar{y}, \bar{u}),$$

which means that CQ (4.3) is also a sufficient condition for $B(\bar{x}, \bar{y}, \bar{u}) = \{(0, 0, 0)\}$. Following Lemma 3.2, the latter equality implies the fulfillment of assumption 1.

(ii) A second possibility to recover CQ (4.3) in the above theorem is to move the constraints defining Ω to the function ψ, i.e. to set $\Omega := \mathbb{R}^n \times \mathbb{R}^m$ and

$$\psi(x, y) := [G(x), \mathcal{L}(x, y, u), (u_i, -g_i(x, y))_{i=1,\ldots,p}].$$

This would also help recover the bound on all the multipliers.

To conclude this subsection, we now deduce the M-type optimality conditions for the bilevel optimization (1.1) from the above developments on the classical KKT reformulation (1.6).

Corollary 4.5. *Let $(\bar{x}, \bar{y})$ be a local optimal solution of problem (1.1). Assume that the following assertions hold:*

1. *CQ (1.7) holds at $(\bar{x}, \bar{y})$*
2. *CQ (4.3) holds at $(\bar{x}, \bar{y}, \bar{u})$, for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$.*

Then, $(\bar{x}, \bar{y})$ is M-stationary, with $\|(\alpha, \beta, \gamma)\| \leq r$ (for some $r > 0$).

This result can be restated with CQ (4.3) of assumption 2 replaced by the CQs in Theorem 4.3.

4.2. *S-type optimality conditions*

In the framework of MPCCs, the Guignard CQ has been shown to be one of the few CQs to be directly applicable to (1.6) considered as a usual nonlinear optimization problem; cf. [4]. In the next result, we derive the S-type stationarity conditions for problem (1.6) under the Guignard CQ, which holds at a point $(\bar{x}, \bar{y}, \bar{u}) \in \mathcal{C}$ ($\mathcal{C}$ being the feasible set of problem (1.6)) if the Fréchet normal cone to $\mathcal{C}$ at $\bar{z}$ takes the form

$$\widehat{N}_{\mathcal{C}}(\bar{x}, \bar{y}, \bar{u}) = \left\{ d := (d_x, d_y, d_u) : \begin{array}{ll} \nabla G_i(\bar{x})^\top d_x \geq 0, & \forall i : G_i(\bar{x}) = 0 \\ \nabla g_i(\bar{x}, \bar{y})^\top (d_x, d_y) \geq 0, & \forall i : g_i(\bar{x}, \bar{y}) = 0 \\ \nabla \mathcal{L}_i(\bar{x}, \bar{y}, \bar{u})^\top d = 0, & \forall i : i = 1, \ldots, m \\ \nabla \sigma(\bar{x}, \bar{y}, \bar{u})^\top d = 0 & \end{array} \right\}$$

where σ denotes the function defined as $\sigma(x, y, u) := -u^\top g(x, y)$.

Theorem 4.6. *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (1.6) and assume that the Guignard CQ is satisfied at $(\bar{x}, \bar{y}, \bar{u})$. Then, there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ such that the S-stationarity conditions (in the sense of Definition 1.4) are satisfied.*

Proof. According to [35, Theorem 3.5], under the Guignard CQ, there exists $(\alpha, \beta, \gamma, \lambda) \in \mathbb{R}^{k+p+m+1}$ such that the following optimality conditions are satisfied:

$$\nabla_x F(\bar{x}, \bar{y}) + \nabla G(\bar{x})^\top \alpha + \nabla_x g(\bar{x}, \bar{y})^\top (\beta - \lambda \bar{u}) + \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \tag{4.19}$$

$$\nabla_y F(\bar{x}, \bar{y}) + \nabla_y g(\bar{x}, \bar{y})^\top (\beta - \lambda \bar{u}) + \nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \tag{4.20}$$

$$\alpha \geq 0, \quad \alpha^\top G(\bar{x}) = 0 \tag{4.21}$$

$$\beta \geq 0, \quad \beta^\top g(\bar{x}, \bar{y}) = 0 \tag{4.22}$$

$$\nabla_y g(\bar{x}, \bar{y})\gamma - \lambda g(\bar{x}, \bar{y}) \geq 0, \qquad \bar{u}^\top (\nabla_y g(\bar{x}, \bar{y})\gamma) = 0. \tag{4.23}$$

It suffices now to show that these conditions are equivalent to the S-stationarity conditions in the sense of Definition 1.4, i.e. for a Lagrange multiplier vector $\bar{v} = (\alpha, \beta, \gamma, \lambda)$, the triple $(\bar{x}, \bar{y}, \bar{v})$ satisfies (4.19)–(4.23) if and only if there exists $\bar{v}^* = (\alpha^*, \beta^*, \gamma^*)$ such that $(\bar{x}, \bar{y}, \bar{v}^*)$ satisfies (1.8)–(1.11) and (1.13).

The first implication follows trivially by means of the definitions of the index sets in the introduction of this paper. The converse follows by applying the same technique as in the proof of Proposition 4.2 in [36]. □

An example of bilevel optimization problem for which the Guignard CQ is satisfied can be found in [35]. For more on this CQ and its application to MPCCs, see [4].

For the rest of this section, we focus our attention to the concept of partial calmness. Precisely, we start by showing how a combination of partial calmness and basic CQ could lead to the S-stationarity conditions. The principle of this result is very simple. In fact, since the failure of the MFCQ is due to the complementarity system, then moving the function σ to the upper level objective function paves the way to the application of the same CQ.

Theorem 4.7. *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (1.6) and assume that the following assertions are satisfied:*

1. *Problem (1.6) is σ-partially calm at $(\bar{x}, \bar{y}, \bar{u})$*
2. *The following implication holds at $(\bar{x}, \bar{y}, \bar{u})$:*

$$\left. \begin{array}{l} \nabla G(\bar{x})^\top \alpha + \nabla_x g(\bar{x}, \bar{y})^\top \beta + \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \\ \nabla_y g(\bar{x}, \bar{y})^\top \beta + \nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0 \\ \alpha \geq 0, \quad \alpha^\top G(\bar{x}) = 0 \\ \beta \geq 0, \quad \beta^\top g(\bar{x}, \bar{y}) = 0 \\ \nabla_y g(\bar{x}, \bar{y})\gamma \geq 0, \qquad \bar{u}^\top (\nabla_y g(\bar{x}, \bar{y})\gamma) = 0 \end{array} \right\} \Longrightarrow \begin{cases} \alpha = 0 \\ \beta = 0 \\ \gamma = 0. \end{cases} \tag{4.24}$$

Then, there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$, with $\|(\alpha, \beta, \gamma)\| \leq r$ (for some $r > 0$) such that the S-stationarity conditions are satisfied.

Proof. Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (1.6) and let assumption 1 of the theorem be satisfied. Then, it follows from Theorem 3.7 that, there exists $\lambda > 0$ such that $(\bar{x}, \bar{y}, \bar{u})$ is also a local optimal solution of the problem

$$\text{minimize } F(x, y) - \lambda u^\top g(x, y) \text{ subject to } (x, y, u) \in \Omega \cap \psi^{-1}(\Lambda), \tag{4.25}$$

where $\Omega = \mathbb{R}^{n+m} \times \mathbb{R}^p_+$, $\Lambda = \mathbb{R}^{k+p}_- \times \{0_m\}$ and $\psi(x, y, u) = (G(x), g(x, y), \mathcal{L}(x, y, u))$. One can easily check that:

$$N_\Omega(\bar{x}, \bar{y}, \bar{u}) = \{0_{n+m}\} \times \{\eta \in \mathbb{R}^p | \eta \leq 0, \eta^\top \bar{u} = 0\} \tag{4.26}$$

$$N_\Lambda(\psi(\bar{x}, \bar{y}, \bar{u})) = \left\{(\alpha, \beta, \gamma) | \alpha \geq 0, \alpha^\top G(\bar{x}) = 0, \beta \geq 0, \beta^\top g(\bar{x}, \bar{y}) = 0\right\} \tag{4.27}$$

$$\nabla\psi(x, y, u)^\top(\alpha, \beta, \gamma) = A(\alpha, \beta, \gamma), \tag{4.28}$$

where $A(\alpha, \beta, \gamma)$ is the matrix in (4.8). It follows from equalities (4.26)–(4.28) that the basic CQ applied to problem (4.25) at $(\bar{x}, \bar{y}, \bar{u})$ can equivalently be formulated as: there is no nonzero vector $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ and a vector $\eta \in \mathbb{R}^p$ (dummy multiplier) such that the first four lines of the left hand side of implication (4.24) and the system

$$\nabla_y g(\bar{x}, \bar{y})\gamma + \eta = 0, \quad \eta \leq 0, \qquad \eta^\top \bar{u} = 0 \tag{4.29}$$

are satisfied, respectively. It clearly follows that assumption 2 corresponds to the basic CQ applied to problem (4.25), where the last line of the system in the left hand side of implication (4.24) is recovered from (4.29). Hence, applying Proposition 3.1 to problem (4.25), it also follows from (4.26)–(4.28) that there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$, with $\|(\alpha, \beta, \gamma)\| \leq r$ (for some $r > 0$) and $\lambda > 0$ such that (4.19)–(4.22) and

$$-\lambda g(\bar{x}, \bar{y}) + \nabla_y g(\bar{x}, \bar{y})\gamma + \eta = 0, \quad \eta \leq 0, \qquad \eta^\top \bar{u} = 0 \tag{4.30}$$

hold. Hence, (4.23) is regained from system (4.30) while noting that the feasibility of $(\bar{x}, \bar{y}, \bar{u})$ implies $\bar{u}^\top g(\bar{x}, \bar{y}) = 0$.

We have shown that there exists $(\alpha, \beta, \gamma, \lambda)$, with $\lambda > 0$, such that (4.19)–(4.23). The S-stationarity conditions (1.8)–(1.11) and (1.13) are then obtained as in the previous result. □

In the next result, we show that the CQ in assumption 2 of the previous theorem can be weakened, if the perturbation map of the joint upper and lower level feasible set is calm.

Theorem 4.8. *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (1.6) and assume that the following assertions are satisfied:*

1. *Problem (1.6) is σ-partially calm at $(\bar{x}, \bar{y}, \bar{u})$*
2. *The following set-valued mapping is calm at $(0, 0, \bar{x}, \bar{y})$*

$$\mathcal{M}_2(t_1, t_2) := \{(x, y) | G(x) + t_1 \leq 0, g(x, y) + t_2 \leq 0\}$$

3. *The following implication holds at $(\bar{x}, \bar{y}, \bar{u})$:*

$$\left.\begin{array}{l} \nabla G(\bar{x})^\top\alpha + \nabla_x g(\bar{x}, \bar{y})^\top\beta + \nabla_x\mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top\gamma = 0 \\ \nabla_y g(\bar{x}, \bar{y})^\top\beta + \nabla_y\mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top\gamma = 0 \\ \alpha \geq 0, \quad \alpha^\top G(\bar{x}) = 0 \\ \beta \geq 0, \quad \beta^\top g(\bar{x}, \bar{y}) = 0 \\ \nabla_y g(\bar{x}, \bar{y})\gamma \geq 0, \qquad \bar{u}^\top(\nabla_y g(\bar{x}, \bar{y})\gamma) = 0 \end{array}\right\} \Longrightarrow \gamma = 0. \tag{4.31}$$

Then, there exists $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$, with $\|\gamma\| \leq r$ (for some $r > 0$) such that the S-stationarity conditions are satisfied.

Proof. Set $\Omega = \{(x, y, u) | u \geq 0, G(x) \leq 0, g(x, y) \leq 0\}$, and let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (1.6). Then, under assumption 1, there exists $\lambda > 0$ such that $(\bar{x}, \bar{y}, \bar{u})$ is also a local optimal solution of

$$\text{minimize } F(x, y) - \lambda u^\top g(x, y) \text{ subject to } (x, y, u) \in \Omega \cap \mathcal{L}^{-1}(0).$$

Hence, it follows from Proposition 3.1 that if

$$[0 \in \nabla\mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top\gamma + N_\Omega(\bar{x}, \bar{y}, \bar{u}), \gamma \in \mathbb{R}^m] \Rightarrow \gamma = 0, \tag{4.32}$$

then there exists $\gamma \in \mathbb{R}^m$ with $\|\gamma\| \leq r$ (for some $r > 0$) such that

$$0 \in \begin{bmatrix} \nabla F(\bar{x}, \bar{y}) \\ -\lambda g(\bar{x}, \bar{y}) \end{bmatrix} + \nabla\mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top\gamma + N_\Omega(\bar{x}, \bar{y}, \bar{u}). \tag{4.33}$$

Now, observe that $\Omega = \Omega' \times \mathbb{R}^p_+$ with $\Omega' := \{(x, y)|G(x) \le 0, g(x, y) \le 0\}$. It follows from Theorem 2.4 that

$$N_{\Omega'}(\bar{x}, \bar{y}) = \{(\nabla G(\bar{x})^\top \alpha + \nabla_x g(\bar{x}, \bar{y})^\top \beta, \nabla_y g(\bar{x}, \bar{y})^\top \beta)^\top | \text{ (4.21)–(4.22)}\}$$

under assumption 2. Hence, we regain assumption 3 and the desired optimality conditions by applying the last equality to (4.32) and (4.33), while noting that $u^* \in N_{\mathbb{R}^p_+}(\bar{u})$ if and only if $u^* \le 0$ and $\bar{u}^\top u^* = 0$. □

Proceeding as in Remark 4.4, one can easily check that assumption 2 in Theorem 4.7 implies the fulfillment of both assumption 2 and 3 in Theorem 4.8.

To conclude this section, it seems important to mention some sufficient conditions for the partial calmness used in Theorems 4.7 and 4.8. For this purpose, a slightly modified notion of uniform weak sharp minimum was introduced in [5]. We recall that the initial definition first appeared in [30].

Definition 4.9. The family $\{(1.2)|x \in X\}$ is said to have a uniformly weak sharp minimum if there exist $c > 0$ and a neighborhood $\mathcal{N}(x)$ of $S(x)$, $x \in X$ such that

$$f(x, y) - \varphi(x) \ge cd(y, S(x)), \quad \forall y \in K(x) \cap \mathcal{N}(x),\ \forall x \in X.$$

If we set $\mathcal{N}(x) = \mathbb{R}^m$, we obtain the definition in [30]. A uniform weak sharp minimum, as given in the above definition, was shown to exist under the uniform calmness; see [5].

Theorem 4.10. *Let $(\bar{x}, \bar{y}, \bar{u})$ be a local optimal solution of problem (1.6). Then, problem (1.6) is σ-partially calm at $(\bar{x}, \bar{y}, \bar{u})$, provided one of the following assumptions hold:*

1. *The family $\{(1.2)|x \in X\}$ has a uniform weak sharp minimum.*
2. *The set $\mathcal{A} := \{(x, y, u)|G(x) \le 0, \mathcal{L}(x, y, u) = 0, g(x, y) \le 0, u \ge 0\}$ is semismooth and $-(\bar{u}^\top \nabla g(\bar{x}, \bar{y}), g(\bar{x}, \bar{y})) \notin \operatorname{bd} N_{\mathcal{A}}(\bar{x}, \bar{y}, \bar{u})$.*

Proof. Under assumption 1, the result follows from [15]. Under assumption 2, it follows from [37] that the set-valued mapping $\mathcal{M}_3(t) := \{(x, y, u) \in \mathcal{A}| - u^\top g(x, y) + t \le 0\}$ is calm at $(0, \bar{x}, \bar{y}, \bar{u})$. Hence, the result follows from Proposition 3.2. □

For the definition of the semismoothness, the interested reader is referred to [37]. This is automatically satisfied for our set $\mathcal{A}$, if G is convex, and $(x, y) \mapsto \nabla_y f(x, y)$ and g are affine linear.

Finally, the S-stationarity conditions for the bilevel program (1.1) can be obtained via its classical KKT reformulation as follows.

Corollary 4.11. *Let $(\bar{x}, \bar{y})$ be a local optimal solution of problem (1.1). Assume that the following assertions hold:*

1. *CQ (1.7) holds at $(\bar{x}, \bar{y})$*
2. *Guignard CQ holds at $(\bar{x}, \bar{y}, \bar{u})$, for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$.*

Then, $(\bar{x}, \bar{y})$ is S-stationary in the sense of Definition 1.4.

Similarly to the M-stationarity case, this corollary can be restated with the Guignard CQ of assumption 2 replaced by the CQs in Theorem 4.7 or Theorem 4.8.

5. The primal KKT reformulation

Consider the set-valued mapping Q defined as $Q(x, y) := N_{K(x)}(y)$, for $y \in K(x)$, and $Q(x, y) := \emptyset$, otherwise. To obtain the closedness of $\operatorname{gph} Q$, necessary in order to apply the basic CQ to problem (1.5), we introduce the concept of inner semicontinuity for a set-valued mapping; see [16] for more details.

A set-valued mapping Φ is said to be inner semicontinuous at $(x, y) \in \operatorname{gph} \Phi$ if for every sequence $x^k \to x$ there is a sequence $y^k \in \Phi(x^k)$ such that $y^k \to y$. Φ will be said to be inner semicontinuous if it is inner semicontinuous at every point of its graph. In the following result, we show that $\operatorname{gph} Q$ is closed if the lower level feasible set mapping K is inner semicontinuous.

Proposition 5.1. *Assume that K is inner semicontinuous. Then, $\operatorname{gph} Q$ is closed as a subset of $\operatorname{gph} K \times \mathbb{R}^m$, i.e. if $(x^k, y^k) \to (x, y)((x^k, y^k) \in \operatorname{gph} K)$ and $z^k \to z$ with $z^k \in Q(x^k, y^k)$, then $z \in Q(x, y)$.*

Proof. Let $(x^k, y^k) \to (x, y)((x^k, y^k) \in \operatorname{gph} K)$ and $z^k \to z$ with $z^k \in Q(x^k, y^k)$. Since $K(x^k)$ is assumed to be convex for all k, we have

$$\langle z^k, u^k - y^k\rangle \le 0, \quad \forall u^k \in K(x^k),\ k \in \mathbb{N}. \tag{5.1}$$

Given that K is inner semicontinuous, then for an arbitrary $v \in K(x)$, there exists $v^k \in K(x^k)$ such that $v^k \to v$. It follows from (5.1) that $\langle z^k, v^k - y^k\rangle \le 0$, $\forall k \in \mathbb{N}$. This implies that $\langle z, v - y\rangle \le 0$, $\forall v \in K(x)$, which concludes the proof. □

This result can be seen as an extension to parametric sets, of the result stated in [2, Proposition 3.3], with the difference that our cones are defined in the sense of convex analysis. But, it may well be extended to the case where $N_{K(x)}(y)$ is the more general Mordukhovich normal cone used in [2], when $K(x) := K$ for all x.

Considering the case where the set-valued mapping K is defined as in (1.3), the concept of inner semicontinuity can be brought to usual terms through the following well-known result; see e.g. [38].

Lemma 5.1. *If CQ (1.7) holds at $(\bar{x}, \bar{y})$, then K is inner semicontinuous near $(\bar{x}, \bar{y})$.*

The normal cone to the graph of $N_{\mathbb{R}^p_-}$, which is also useful in this section, can be obtained from the normal cone to Θ (defined in (4.2)) as follows.

Proposition 5.2. *Let $(\overline{\chi}, \overline{\upsilon}) \in \operatorname{gph} N_{\mathbb{R}^p_-}$, then*

$$N_{\operatorname{gph} N_{\mathbb{R}^p_-}}(\overline{\chi}, \overline{\upsilon}) = \{(-\chi^*, \upsilon^*) \in \mathbb{R}^{2p} | (\chi^*, \upsilon^*) \in N_\Theta(-\overline{\chi}, \overline{\upsilon})\}.$$

Proof. We start by noting that

$$\begin{aligned}\operatorname{gph} N_{\mathbb{R}^p_-} &= \{(\chi, \upsilon) \in \mathbb{R}^{2p} | \chi \leq 0, \upsilon \geq 0, \chi^\top \upsilon = 0\}\\ &= \{(\chi, \upsilon) \in \mathbb{R}^{2p} | (-\chi, \upsilon) \in \Theta\}.\end{aligned}$$

This means that $\operatorname{gph} N_{\mathbb{R}^p_-} = \vartheta^{-1}(\Theta)$, where $\vartheta(\chi, \upsilon) := (-\chi, \upsilon)$ and for $(\overline{\chi}, \overline{\upsilon}) \in \mathbb{R}^{2p}$, one obviously has

$$\nabla \vartheta(\overline{\chi}, \overline{\upsilon}) = \begin{pmatrix} -I_p & O \\ O & I_p \end{pmatrix}$$

with I_p and O denoting the $p \times p$ identity and zero matrix, respectively. Hence, the Jacobian matrix $\nabla \vartheta(\overline{\chi}, \overline{\upsilon})$ is quadratic and nonsingular and it follows from [20, Corollary 2.12] that

$$N_{\operatorname{gph} N_{\mathbb{R}^p_-}}(\overline{\chi}, \overline{\upsilon}) = \nabla\vartheta(\overline{\chi}, \overline{\upsilon})^\top N_\Theta(\vartheta(\overline{\chi}, \overline{\upsilon})),$$

given that $\operatorname{gph} N_{\mathbb{R}^p_-}$ and Θ are closed sets. The result then follows. □

Exploiting the polyhedrality of $\operatorname{gph} N_{\mathbb{R}^p_-}$ and Θ, the equality in the above result can also be proven, at least in the case where $\overline{\chi} = 0$, by using a combination of [39, Proposition 1] and [22, Theorem 5].

In the next theorem, we present a slightly modified version of Theorem 6.1 in [1]. We also include a sketch of the proof for further references in the rest of the paper, and also to simplify the comparison between the primal and classical KKT reformulation approaches, in deriving necessary optimality conditions for the bilevel program (1.1).

Theorem 5.2. *Let $(\bar{x}, \bar{y})$ be a local optimal solution of problem (1.5) (i.e. of (1.1)). Assume that the following assertions hold:*

1. *CQ (1.7) holds at $(\bar{x}, \bar{y})$*
2. *The following set-valued mapping is calm at $(0, \bar{x}, \bar{y}, \bar{u})$, for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$*

$$M(\vartheta) := \left\{ (x, y, u) \middle| \begin{bmatrix} g(x, y) \\ u \end{bmatrix} + \vartheta \in \operatorname{gph} N_{\mathbb{R}^p_-} \right\}$$

3. *The following set-valued mapping is calm at $(0, 0, \bar{x}, \bar{y}, \bar{u})$, for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$*

$$P(z, \vartheta) := \left\{ (x, y, u) \middle| \begin{bmatrix} G(x) \\ \mathcal{L}(x, y, u) \end{bmatrix} + z \in \mathbb{R}^k_- \times \{0_m\} \right\} \cap M(\vartheta).$$

Then, there exist $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ and $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ such that the M-stationarity conditions (1.8)–(1.12) are satisfied, with $\|(\alpha, \gamma)\| \leq r$ (for some $r > 0$).

Proof. We organize the proof in three steps in order to simplify further reference.

Step 1. Consider the following values for ψ and Λ, respectively:

$$\psi(x, y) := [G(x), x, y, -\nabla_y f(x, y)], \quad \text{and} \quad \Lambda := \mathbb{R}^k_- \times \operatorname{gph} Q.$$

Then it follows from Proposition 3.1 that there exists $(\alpha, \gamma) \in \mathbb{R}^{k+m}$, with $\|(\alpha, \gamma)\| \leq r$ (for some $r > 0$) such that

$$0 \in \nabla F(\bar{x}, \bar{y}) + \begin{bmatrix} \nabla G(\bar{x})^\top \alpha + \nabla^2_{xy} f(\bar{x}, \bar{y})^\top \gamma \\ \nabla^2_{yy} f(\bar{x}, \bar{y})^\top \gamma \end{bmatrix} + D^* Q((\bar{x}, \bar{y}) | -\nabla_y f(\bar{x}, \bar{y}))(\gamma) \tag{5.2}$$

provided Λ is closed and the set-valued mapping Ψ in (3.3) (with $\Omega := \mathbb{R}^n \times \mathbb{R}^m$, ψ and Λ given as in the beginning of this proof) is calm at $(0, \bar{x}, \bar{y})$. Obviously, the closedness of Λ is ensured by assumption 1, by a combination of Proposition 5.1 and Lemma 5.1. As for the calmness of Ψ, it is obtained by assumption 3. In fact the proof of the latter claim can be adapted from the proof of Theorem 4.3 in [1].

Step 2. Under assumption 1 and 2, an upper bound for the coderivative of Q at $(\bar{x}, \bar{y}, -\nabla_y f(\bar{x}, \bar{y}))$ is derived from Theorem 3.1 in [1]:

$$D^*Q((\bar{x}, \bar{y})| - \nabla_y f(\bar{x}, \bar{y}))(\gamma) \subseteq \bigcup_{\bar{u}\in\Lambda(\bar{x},\bar{y})} \left\{(\nabla(\nabla_y g(\bar{x}, \bar{y})^\top \bar{u}))^\top \gamma + \nabla g(\bar{x}, \bar{y})^\top D^* N_{\mathbb{R}^p_-}(g(\bar{x}, \bar{y}), \bar{u})(\nabla_y g(\bar{x}, \bar{y})\gamma)\right\}.$$

Step 3. To conclude the proof, note that $\beta \in D^* N_{\mathbb{R}^p_-}(g(\bar{x}, \bar{y}), \bar{u})(\nabla_y g(\bar{x}, \bar{y})\gamma)$ if and only if $(\beta, -\nabla_y g(\bar{x}, \bar{y})\gamma) \in N_{\text{gph}\, N_{\mathbb{R}^p_-}}(g(\bar{x}, \bar{y}), \bar{u})$. Hence, the result follows by considering the equality in Proposition 5.2. □

In our case, constraint $G(x) \leq 0$ is included in ψ whereas it is part of Ω in [1]. The reason for this is to get a close link between CQ (4.3) and assumptions 2 and 3. Also, $\Lambda(\bar{x}, \bar{y})$ is not a singleton as in [1]. To obtain this, at the place of assumption 1, it is required in [1] that $\nabla_y g(\bar{x}, \bar{y})$ have full rank. Finally, in contrary to Theorem 6.1 of [1], the multipliers α and γ are bounded in Theorem 5.2 by a known number, something which can be useful when constructing an algorithm for the bilevel program.

There are two equivalent ways to interpret the coderivative term in the right hand side of the inclusion in step 2 of the proof of Theorem 5.2. The first one used in [40] consists in writing it directly in terms of $g(\bar{x}, \bar{y}), \bar{u}$ and $\nabla_y g(\bar{x}, \bar{y})\gamma$. It should however be mentioned that in [40], g does not depend on the parameter x. The second one that we have used here consists in first computing the normal cone to the graph of $N_{\mathbb{R}^p_-}$ at $(g(\bar{x}, \bar{y}), \bar{u})$. Then, translating inclusion $(\beta, -\nabla_y g(\bar{x}, \bar{y})\gamma) \in N_{\text{gph}\, N_{\mathbb{R}^p_-}}(g(\bar{x}, \bar{y}), \bar{u})$, directly leads to the M-stationarity conditions in the sense of Definition 1.3.

At first view, it is not apparent that the optimality conditions in [40] are in fact equivalent to the M-stationarity conditions in Definition 1.3. Moreover, a condition was later suggested in [41], in order to obtain S-type optimality conditions for an OPEC from the M-ones. By the way, let us mention that in the case of our problem, the S-stationarity conditions defined in [41] correspond to those of Definition 1.4. Hence, a CQ similar to the one suggested in [9] and called Partial MPEC LICQ can also lead from M- to S-type optimality conditions for OPEC. For problem (1.1), one can easily check that the Partial MPEC LICQ takes the form:

$$\left.\begin{array}{l}\nabla G(\bar{x})^\top \alpha + \nabla_x g(\bar{x}, \bar{y})^\top \beta + \nabla_x \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0\\ \nabla_y g(\bar{x}, \bar{y})^\top \beta + \nabla_y \mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma = 0\\ \beta_\nu = 0, \qquad \nabla_y g_\eta(\bar{x}, \bar{y})\gamma = 0\end{array}\right\} \Longrightarrow \beta_\mu = 0, \qquad \nabla_y g_\mu(\bar{x}, \bar{y})\gamma = 0.$$

For a more clear comparison between the approach in the previous section and the current one, one should note that assumption 2 and 3 of Theorem 5.2 are satisfied, provided that CQ (4.3) holds at $(\bar{x}, \bar{y}, \bar{u})$, for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$. This follows similarly as in Remark 4.4. Hence, the following corollary of the last theorem.

Corollary 5.3. *Let $(\bar{x}, \bar{y})$ be a local optimal solution of problem (1.1). Assume that the following assertions hold:*

1. *CQ (1.7) holds at $(\bar{x}, \bar{y})$*
2. *CQ (4.3) holds at $(\bar{x}, \bar{y}, \bar{u})$, for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$.*

Then, there exist $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ and $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ such that the M-stationarity conditions (1.8)–(1.12) are satisfied, with $\|(\alpha, \gamma)\| \leq r$ (for some $r > 0$).

If we neglect the bounds on the multipliers, the only difference between Corollaries 4.5 and 5.3 is that for the former, the M-stationarity conditions (1.8)–(1.12) have to be satisfied *for all* $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, whereas for the latter, they have to hold *for some* $\bar{u} \in \Lambda(\bar{x}, \bar{y})$. It cannot be otherwise in the case of Corollary 5.3, if one considers the inclusion in Step 2 of the proof of Theorem 5.2. This means that, for the bilevel optimization problem (1.1), if we adopt the definition of the M-stationarity of a local optimal solution as in Definition 1.3, the primal KKT reformulation leads us to weaker conditions. Hence, one could say that the gain we have with the primal KKT reformulation in terms of local optimal solution (cf. Theorem 1.1), is lost when considering M-stationarity. From the view point of CQs, obviously, the same effort (i.e. in terms of CQs) leads to M-type optimality conditions for the primal KKT reformulation, which are weaker than those obtained via the classical KKT reformulation.

Moreover, the classical KKT reformulation provides a much bigger flexibility in designing surrogates for the other optimality conditions known for MPCCs; see for example the S-type stationarity conditions obtained in Section 4.2. However, it may be very difficult to derive S-type optimality conditions for an OPEC. The reason for this is that the Fréchet normal cone does not have as good calculus rules as that of Mordukhovich.

For the last part of this section, we now introduce a concept of partial *calmness/exact penalization* for the primal KKT reformulation of the bilevel programming problem, that will allow us to substantially weaken CQ (4.3), while still being able to obtain the optimality conditions in Theorem 5.2. To bring the concept of partial calmness in Definition 3.6 to the primal KKT reformulation (1.5), one possibility is to observe that

$$\psi(x, y) \in \Lambda \Longleftrightarrow \rho(x, y) := d_\Lambda o \psi(x, y) = 0,$$

where $\Lambda := \text{gph}\, Q$, $\psi(x, y) := (x, y, -\nabla_y f(x, y))$ and d_Λ denotes the distance function. Hence, the primal KKT reformulation can equivalently be written as

$$\text{minimize } F(x, y) \text{ subject to } (x, y) \in X \times \mathbb{R}^m, \qquad \rho(x, y) = 0. \tag{5.3}$$

Such a transformation for an OPEC has already been suggested in [37], but with no further details. We start by showing in the next result that problem (5.3) leads to the same optimality conditions as in Theorem 5.2. Second, we show that the latter reformulation of an OPEC induces new but not so fruitful ideas for CQs. $\bar{x} \in X$ will be said to be *upper level regular* if there exists no nonzero vector $\alpha \geq 0$: $\alpha^\top G(\bar{x}) = 0$ and $\nabla G(\bar{x})^\top \alpha = 0$.

Theorem 5.4. *Let $(\bar{x}, \bar{y})$ be a local optimal solution of (5.3) (i.e. of problem (1.1)). Assume that the following assertions hold:*

1. *$\bar{x}$ is upper level regular*
2. *assumption 1 and 2 in Theorem 5.2*
3. *ρ-partial calmness at $(\bar{x}, \bar{y})$.*

Then, there exist $(\alpha, \beta, \gamma) \in \mathbb{R}^{k+p+m}$ and $\bar{u} \in \Lambda(\bar{x}, \bar{y})$ such that the M-stationarity conditions (1.8)–(1.12) are satisfied, with $\|(\alpha, \gamma)\| \leq r$ (for some $r > 0$).

Proof. Under assumption 3, it follows from Theorem 3.7, that there exists $r_1 > 0$ such that $(\bar{x}, \bar{y})$ is a local optimal solution of

$$\text{minimize } F(x, y) + r_1\rho(x, y) \text{ subject to } G(x) \leq 0.$$

Hence, from Theorems 2.1 and 4.2, there exists α, with $\|\alpha\| \leq r_2$, for some $r_2 > 0$ such that (1.10) and the following inclusion hold

$$0 \in \nabla F(\bar{x}, \bar{y}) + (\nabla G(\bar{x})^\top \alpha, 0)^\top + r_1 \partial\rho(\bar{x}, \bar{y}). \tag{5.4}$$

Applying Theorem 2.2 to ρ, it follows that

$$\partial\rho(\bar{x}, \bar{y}) \subseteq \bigcup [\partial\langle u^*, \psi\rangle(\bar{x}, \bar{y}), \ , u^* \in \mathbb{B} \cap N_\Lambda(\psi(\bar{x}, \bar{y}))], \tag{5.5}$$

given that Λ is (locally) closed and hence, $\partial d_\Lambda(a) = \mathbb{B} \cap N_\Lambda(a)$ for any $a \in \Lambda$; cf. [18, Example 8.53]. Thus, the combination of (5.4) and (5.5) implies that there exists $\gamma \in \mathbb{R}^m$, with $\|\gamma\| \leq r_1$, such that inclusion (5.2) holds.

The rest of the proof then follows as that of Theorem 5.2. In this case, r can be chosen as $r = \min\{r_1, r_2\}$. □

Remark 5.5. The need of the partial calmness in order to have ρ (composition of ψ and the distance function on Λ) as an exact penalization term can be avoided by considering a classical result of Clarke [31, Proposition 2.4.3], which amounts to saying that the distance function is automatically an exact penalty term provided the objective function (F in our case) is Lipschitz continuous. To proceed, one should observe that $\psi(x, y) \in \Lambda$ is also equivalent to $d_{\psi^{-1}(\Lambda)}(x, y) = 0$. Hence, $(\bar{x}, \bar{y})$ is a local optimal solution of problem (1.5) implies that, there exists $r > 0$ such that $(\bar{x}, \bar{y})$ is a local optimal solution of the problem to

$$\text{minimize } F(x, y) + rd_{\psi^{-1}(\Lambda)}(x, y) \text{ subject to } (x, y) \in X \times \mathbb{R}^m,$$

without any CQ. In exchange though, computing the basic subdifferential of the distance function $d_{\psi^{-1}(\Lambda)}$ would then require an assumption closely related to the ρ-partial calmness, i.e. the calmness of a certain set-valued mapping.

Consider the set-valued mapping $\Psi(\vartheta) := \{(x, y) \in X \times \mathbb{R}^m | \psi(x, y) + \vartheta \in \Lambda\}$, where $\psi(x, y) := (x, y, -\nabla_y f(x, y))$ and $\Lambda := \text{gph}\, Q$. The next result helps to show that the ρ-partial calmness is closely related to the CQ in assumption 3 of Theorem 5.2.

Theorem 5.6. *Let $(\bar{x}, \bar{y}) \in \Psi(0)$. Then, Ψ is calm at $(0, \bar{x}, \bar{y})$ if and only if the following set-valued mapping is calm at $(0, \bar{x}, \bar{y})$*

$$\widetilde{\Psi}(t) := \{(x, y) \in X \times \mathbb{R}^m | \rho(x, y) \leq t\}. \tag{5.6}$$

This result established in [22], for the case where $X \times \mathbb{R}^m$ corresponds to a normed space, remains valid in our setting. Additionally, one can easily check that the calmness of $\widetilde{\Psi}$ is equivalent to the calmness of a set-valued mapping obtained by replacing $\rho(x, y) \leq t$ in (5.6) by $\rho(x, y) + t \leq 0$. Hence, by the combination of Proposition 3.2 and Theorem 5.6, it is clear that the calmness of

$$\overline{P}(z, \vartheta) := \{(x, y, u) \in X \times \mathbb{R}^m \times \mathbb{R}^p | \mathcal{L}(x, y, u) + z = 0\} \cap M(\vartheta)$$

at $(0, 0, \bar{x}, \bar{y}, \bar{u})$, for all $\bar{u} \in \Lambda(\bar{x}, \bar{y})$, is a sufficient condition for the ρ-partial calmness to hold. Here, $M$is defined as in assumption 2 of Theorem 5.2. Hence, a dual condition, similar to (and weaker than) CQ (4.3), sufficient for the ρ-partial calmness, can be stated.

Moreover, applying Proposition 3.2, the ρ-partial calmness also inspires a different kind of CQ in the dual form. In fact, it follows from Proposition 3.2, that a sufficient condition for the ρ-partial calmness to hold is that:

$$\partial\rho(\bar{x}, \bar{y}) \cap -N_{X \times \mathbb{R}^m}(\bar{x}, \bar{y}) = \emptyset. \tag{5.7}$$

Proposition 5.3. *If equality holds in (5.5), then CQ (5.7) fails at* $(\bar{x}, \bar{y})$.

Proof. If equality holds in (5.5), then $0 \in \partial\rho(\bar{x}, \bar{y})$ since as a normal cone, $N_\Lambda(\psi(\bar{x}, \bar{y}))$ always contains the origin point. For the latter reason, we also have $0 \in N_{X\times\mathbb{R}^m}(\bar{x}, \bar{y})$. Hence, the result. □

This behavior of CQ (5.7) is close to that of a similar CQ considered in [5] for the optimal value reformulation (1.14), where instead of $\rho(x, y)$ one has $f(x, y) - \varphi(x)$. In the case of problem (1.14), the corresponding CQ was shown to automatically fail, provided the value function is locally Lipschitz continuous. Even though one can easily construct examples where equality holds in (5.5), the generalization of this fact would generally require the set Λ to be normally regular at $\psi(\bar{x}, \bar{y})$, which may not be easy to have given that Λ is the graph of a normal cone mapping. Moreover, if we assume X to be convex, passing to the boundary of $N_{X\times\mathbb{R}^m}$ generates a CQ, analogous to the one also considered in [5] for (1.14), that may have more chances to be satisfied:

$$\partial\rho(\bar{x}, \bar{y}) \cap -\mathrm{bd}\, N_{X\times\mathbb{R}^m}(\bar{x}, \bar{y}) = \emptyset. \tag{5.8}$$

Considering the abstract nature of CQ (5.8) (which is weaker than CQ (5.7)), an immediate attempt to write it in terms of problem data, i.e. in a verifiable form, produces the following condition

$$\begin{aligned}&\nabla g(\bar{x}, \bar{y})^\top \beta + \nabla_{x,y}\mathcal{L}(\bar{x}, \bar{y}, \bar{u})^\top \gamma \notin \mathrm{bd}\, N_{X\times\mathbb{R}^m}(\bar{x}, \bar{y}) \quad \forall \bar{u} \in \Lambda(\bar{x}, \bar{y}),\ \forall \gamma \in \mathbb{R}^m,\\ &\quad \forall \beta \in D^*N_{\mathbb{R}^p}(g(\bar{x}, \bar{y}), \bar{u})(\nabla_y g(\bar{x}, \bar{y})\gamma),\end{aligned} \tag{5.9}$$

provided inclusion (5.5) holds, while noting that for $A, B, C \subseteq \mathbb{R}^l$, $A \subseteq B$ and $B \cap C = \emptyset$ imply $A \cap C = \emptyset$. It should however be mentioned that CQ (5.9) also fails in many cases, in particular, if $0 \in \mathrm{bd}\, N_{X\times\mathbb{R}^m}(\bar{x}, \bar{y})$.

Acknowledgment

The second author acknowledges the support by the Deutscher Akademischer Austausch Dienst (DAAD).

References

[1] B.S. Mordukhovich, J.V. Outrata, Coderivative analysis of quasi-variational inequalities with applications to stability and optimization, SIAM J. Optim. 18 (2) (2007) 389–412.
[2] R.T. Rockafellar, Lagrange multipliers and optimality, SIAM Rev. 35 (2) (1993) 183–238.
[3] S. Dempe, J. Dutta, Is bilevel programming a special case of mathematical programming with equilibrium constraints? Math. Program., in press (doi:10.1007/s10107-010-0342-1).
[4] M.L. Flegel, Constraint qualification and stationarity concepts for mathematical programs with equilibrium constraints, Ph.D. Thesis, Institute of Applied Mathematics and Statistics, University of Würzburg, Würzburg, 2005.
[5] S. Dempe, A.B. Zemkoho, The generalized Mangasarian-Fromowitz constraint qualification and optimality conditions for bilevel programs, J. Optim. Theory Appl. 148 (2011) 46–68.
[6] J.V. Outrata, On the numerical solution of a class of Stackelberg problems, Z. Oper. Res. 34 (4) (1990) 255–277.
[7] R. Henrion, T.M. Surowiec, On calmness conditions in convex bilevel programming, Preprint 1508, Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin, 2010.
[8] H. Scheel, S. Scholtes, Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity, Math. Oper. Res. 25 (1) (2000) 1–22.
[9] J.J. Ye, Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints, J. Math. Anal. Appl. 307 (1) (2005) 350–369.
[10] S. Dempe, A necessary and a sufficient optimality condition for bilevel programming problems, Optimization 25 (4) (1992) 341–354.
[11] S. Dempe, Foundations of Bilevel Programming, Kluwer Academic Publishers, Dordrecht, 2002.
[12] S. Dempe, J. Dutta, S. Lohse, Optimality conditions for bilevel programming problems, Optimization 55 (5–6) (2006) 505–524.
[13] S. Dempe, S. Lohse, Inverse linear programming, in: A. Seeger (Ed.), Recent Advances in Optimization, in: Proceedings of the Twelfth French–German–Spanish Conference on Optimization, Springer, Avignon, France, Berlin, 2004, 2006, pp. 19-28.
[14] Y. Chen, M. Florian, The nonlinear bilevel programming problem: formulations, regularity and optimality conditions, Optimization 32 (3) (1995) 193–209.
[15] J.J. Ye, D.L. Zhu, Q.J. Zhu, Exact penalization and necessary optimality conditions for generalized bilevel programming problems, SIAM J. Optim. 7 (2) (1997) 481–507.
[16] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation. I: Basic Theory. II: Applications, Springer, Berlin, 2006.
[17] R. Henrion, A. Jourani, J.V. Outrata, On the calmness of a class of multifunctions, SIAM J. Optim. 13 (2) (2002) 603–618.
[18] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Springer, Berlin, 1998.
[19] B.S. Mordukhovich, Generalized differential calculus for nonsmooth and set-valued mappings, J. Math. Anal. Appl. 183 (1) (1994) 250–288.
[20] B.S. Mordukhovich, Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis, Trans. Amer. Math. Soc. 343 (2) (1994) 609–657.
[21] J.-P. Aubin, Lipschitz behavior of solutions to convex minimization problems, Math. Oper. Res. 9 (1984) 87–111.
[22] R. Henrion, J.V. Outrata, Calmness of constraint systems with applications, Math. Program. 104 (2–3) (2005) 437–464.
[23] B.S. Mordukhovich, Nonsmooth analysis with nonconvex generalized differentials and conjugate mappings, Dokl. Akad. Nauk BSSR 28 (1984) 976–979.
[24] B.S. Mordukhovich, Lipschitzian stability of constraint systems and generalized equations, Nonlinear Anal. 22 (2) (1994) 173–206.
[25] A. Jourani, Constraint qualifications and Lagrange multipliers in nondifferentiable programming problems, J. Optim. Theory Appl. 81 (3) (1994) 533–548.
[26] S.-P. Han, O.L. Mangasarian, Exact penalty functions in nonlinear programming, Math. Program. 17 (1979) 251–269.
[27] J.J. Ye, X.Y. Ye, Necessary optimality conditions for optimization problems with variational inequality constraints, Math. Oper. Res. 22 (4) (1997) 977–997.
[28] R. Zhang, Problems of hierarchical optimization in finite dimensions, SIAM J. Optim. 4 (3) (1994) 521–536.
[29] J. Dutta, Generalized derivatives and nonsmooth optimization, a finite dimensional tour (with comments and rejoinder), Top 13 (2) (2005) 185–314.
[30] J.J. Ye, D.L. Zhu, Optimality conditions for bilevel programming problems, Optimization 33 (1) (1995) 9–27.

[31] F.H. Clarke, Optimization and Nonsmooth Analysis, in: SIAM Classics in Applied Mathematics, vol. 5, Wiley, New York, 1984, Reprint. Philadelphia, 1994.
[32] M.L. Flegel, C. Kanzow, A direct proof for *M*-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints, in: S. Dempe, V.V. Kalashnikov (Eds.), Optimization with Multivalued Mappings, Springer, New York, 2006, pp. 111–122.
[33] J.J. Ye, Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints, SIAM J. Optim. 10 (4) (2000) 943–962.
[34] T. Hoheisel, C. Kanzow, J.V. Outrata, Exact penalty results for mathematical programs with vanishing constraints, Nonlinear Anal. 72 (2010) 2514–2526.
[35] S. Dempe, A.B. Zemkoho, The bilevel programming problem: reformulations, constraint qualifications and optimality conditions, Preprint 2010-03, Fakultät für Mathematik und Informatik, Technische Universität Bergakademie Freiberg, Freiberg, 2010.
[36] M.L. Flegel, C. Kanzow, On the Guignard constraint qualification for mathematical programs with equilibrium constraints, Optimization 54 (6) (2005) 517–534.
[37] R. Henrion, J.V. Outrata, A subdifferential condition for calmness of multifunctions, J. Math. Anal. Appl. 258 (1) (2001) 110–130.
[38] S. Dempe, J. Dutta, B.S. Mordukhovich, New necessary optimality conditions in optimistic bilevel programming, Optimization 56 (5–6) (2007) 577–604.
[39] S.M. Robinson, Some continuity properties of polyhedral multifunctions, Math. Program. Stud. 14 (1981) 206–214.
[40] T.M. Surowiec, Explicit stationary conditions and solution characterization for equilibrium problems with equilibrium constraints, Ph.D. Thesis, Mathematisch-Naturwissentschaften Fakultät II, Humbolt Universität zu Berlin, 2010.
[41] R. Henrion, J.V. Outrata, T.M. Surowiec, A note on the relation between strong and *M*-stationarity for a class of mathematical programs with equilibrium constraints, Kybernetika 46 (3) (2010) 423–434.