Open access • Journal Article • DOI:10.1007/S10107-011-0508-5

The bilevel programming problem: reformulations, constraint qualifications and optimality conditions — Source link

Stephan Dempe, Alain B. Zemkoho

Institutions: Freiberg University of Mining and Technology

Published on: 01 Apr 2013 - Mathematical Programming (Springer-Verlag)

Topics: Bilevel optimization, Karush–Kuhn–Tucker conditions and Calmness

Fakultät für Mathematik und Informatik

Preprint 2010-03

Stephan Dempe, Alain B. Zemkoho

The bilevel programming problem: reformulations, constraint qualifications and optimality conditions

ISSN 1433-9307

Stephan Dempe, Alain B. Zemkoho

The bilevel programming problem: reformulations, constraint qualifications and optimality conditions

TU Bergakademie Freiberg
Fakultät für Mathematik und Informatik
Institut für Numerische Mathematik
und Optimierung
Akademiestr. 6 (Mittelbau)
09596 FREIBERG
http://www.mathe.tu-freiberg.de

ISSN 1433 – 9307

Herausgeber: Dekan der Fakultät für Mathematik und Informatik

Herstellung: Medienzentrum der TU Bergakademie Freiberg

The bilevel programming problem: reformulations, constraint qualifications and optimality conditions

Stephan Dempe* and Alain B. Zemkoho†
Institute of Numerical Mathematics and Optimization
Technical University Bergakademie Freiberg
09596 Freiberg, Germany

February 21, 2010

Abstract

We consider the bilevel programming problem and its optimal value and KKT one level reformulations. The two reformulations are studied in a unified manner and compared in terms of optimal solutions, constraint qualifications and optimality conditions. We also show that any bilevel programming problem where the lower level problem is linear with respect to the lower level variable, is partially calm without any restrictive assumption. Finally, we consider the bilevel demand adjustment problem in transportation, and show how KKT type optimality conditions can be obtained under the partial calmness, using the differential theory of Mordukhovich.

Keywords: Bilevel programming, Optimal value function, Constraint qualifications, Optimality conditions, Demand adjustment problem.
2000 AMS subject classifications: 90C30, 91A65, 90B06.

1 Introduction

In this paper we consider the (optimistic) bilevel programming problem also called the leader's problem, which is a special optimization problem partially constrained by a second (parametric) optimization problem known as the follower's problem. In order to write the bilevel programming problem as a one level mathematical programming problem, two major reformulations have been suggested in the literature [6], i.e. the KKT reformulation and the optimal value reformulation. The KKT reformulation usually consist in replacing the follower's problem by its Karush-Kuhn-Tucker (KKT) conditions provided that the latter problem is convex in the lower level variable and an appropriate constraint qualification (CQ) is satisfied. This reformulation introduces new variables, thus complicating the task of constructing a solution point of the KKT

*dempe@math.tu-freiberg.de
†zemkoho@student.tu-freiberg.de

reformulation that would solve the original bilevel programming problem, hence the two problems are not equivalent, at least locally [7]. Also the new variables induce a complementarity slackness constraint which causes the failure of the Mangasarian-Fromowitz constraint qualification (MFCQ).

Considering the convexity of the follower's problem w.r.t. the lower level variable, the latter problem can be rewritten as a generalized equation containing the normal cone to its feasible set. Under this circumstance, the new problem can also be seen as a (primal) KKT reformulation of the bilevel programming problem, even though it is abstract in nature. At this stage, we will show that the bilevel programming problem still maintains a good relationship with its KKT reformulation. Hence the latter problem is also equivalent, for local and global solutions, to the optimal value reformulation. Unfortunately, this nice relationship between the two reformulations is not enjoyed in terms of CQs. In fact, we will show in Section 3 that the KKT and the optimal value reformulations may behave differently under a given CQ.

Further in the paper, we consider a simpler case where the lower level feasible set is defined by inequality constraints and the normal cone to this set is effectively computed. It is well-known that all first category CQs, i.e. those implying the MFCQ also fail for this problem. We show here that the Guignard CQ is weak enough to have some real chances to be satisfied. Also, it is possible to have fruitful relations between the KKT type optimality conditions of the optimal value and the KKT reformulations of the bilevel programming problem. Notably, the KKT optimality conditions of both reformulations may be identical when the follower's problem is linear. Again, in the context of the KKT reformulation of the bilevel problem, the latter and the optimal value reformulation have an analogy in the failure of some well-known CQs. As a matter of fact, the MFCQ fail for the optimal value reformulation because of the optimal value constraint [13], while the failure for the same CQ for the KKT reformulation is due to the complementarity slackness constraint [30]. In order to solve this difficulty, we consider the partial calmness which was introduced by Ye and Zhu [37] and designed for the purpose. In Section 4 of the paper, an important result is proven: the bilevel programming problem where the follower's problem is linear in the lower level variable is partially calm at an arbitrary local optimal solution. This result largely improves a result already established by Ye [35], in the case where no constraint is imposed on the upper level variable.

Finally, in the last section of the paper, we consider an application of bilevel programming in transportation. In fact, the demand adjustment problem (DAP) in road networks has been modeled by Fisk [16] as a bilevel programming problem. But like other bilevel transportation problems, the issue of optimality conditions has not (or very little) been addressed in the literature. A reason for this may be that in addition to the general burdens of bilevel programs, i.e. the nonsmoothness, nonconvexity and the failure of classical CQs, as mentioned above, the feasible set of the traffic assignment problem has a special structure figuring an interplay between the route and link flows, which does not seem easy to handle. Writing the set of feasible link flows as a composition of two set-valued mappings, we use the sophisticated coderivative tool of Mordukhovich, from the field of variational analysis, to design Karush-Kuhn-Tucker type optimality conditions for the DAP. Hence, greatly improving the works of Chen [4] and Chen and Florian [5], where Fritz-John's type optimality conditions were derived after considering some important simplifications.

We first present in the next section some basic notations and background material to be used in the paper. Mainly, relevant properties of the generalized differential theory of Mordukhovich (i.e. the Mordukhovich normal cone, subdifferential and coderivative) are discussed. In Section 3, we introduce the optimistic bilevel programming problem and the optimal value and KKT reformulations are considered. Their behavior for some CQs are analyzed and compared, also their KKT type optimality conditions are derived and possible links between them are studied. In Section 4, we consider the partial calmness, which is an interesting CQ for the bilevel programming problem. Here we show that it is automatically satisfied when the follower's problem is linear in the lower level variable, under very fairly weak considerations. Finally, in section 5, the DAP is introduced and we show that under the natural assumption that routes of the network have limited capacities, the solution set-valued mapping of the traffic assignment problem is inner semicompact. Hence the optimal value function of the traffic assignment problem is automatically Lipschitz continuous and KKT type optimality conditions of the DAP can simply be derived under the partial calmness, which is satisfied at every local optimal solution when the total cost of the route users is linear in the link flows.

2 Background material

In this section we present some basic concepts and notations used in this paper. More details on the material, briefly discussed here can be found in the books of Mordukhovich [24, 25] and Rockafellar and Wets [29]. We first consider some initial notations: Let A be a subset of $\mathbb{R}^n$, co$A$, cl$A$ and bd$A$ denotes the convex hull, the closure and the boundary of A respectively. For a matrix B, $B^\top$ is the transposed matrix of B. For $a \in \mathbb{R}^n$, $a \leq 0$ should be understood componentwise. Finally, $\|.\|$ denotes an arbitrary norm in $\mathbb{R}^n$ and $\langle ., . \rangle$ is used for the inner product of $\mathbb{R}^n$.

A function $\psi : \mathbb{R}^n \to \mathbb{R}^m$ is said to be locally Lipschitz continuous around $\overline{x} \in \mathbb{R}^n$ if there exist $\delta, \kappa > 0$ such that

$$\|\psi(x) - \psi(y)\| \leq \kappa \|x - y\|, \text{ for all } x, y \in \overline{x} + \delta\mathbb{B},$$

where $\mathbb{B}$ is the unit ball of $\mathbb{R}^n$ and κ is called the Lipschitz constant. ψ is locally Lipschitz continuous if it is locally Lipschitz continuous around every point of $\mathbb{R}^n$. The function ψ is said to be Lipschitz continuous if the above inequality holds with $\delta = \infty$. The local Lipschitz continuity of the real-valued function $\psi : \mathbb{R}^n \to \mathbb{R}$ is necessary for its convexity.

Next we assume that A is a closed subset of $\mathbb{R}^n$. The Bouligand tangent cone to A at some point $\overline{x} \in A$ is defined by

$$T_A(\overline{x}) = \left\{ d \in \mathbb{R}^n | \exists t_k \downarrow 0, d_k \to d : \overline{x} + t_k d_k \in A \right\}$$

and the regular normal cone to A at $\overline{x} \in A$ is given as

$$\widehat{N}_A(\overline{x}) = \left\{ d^* \in \mathbb{R}^n | \langle d^*, d \rangle \leq 0 \, \forall d \in T_A(\overline{x}) \right\}$$

while the basic normal cone introduced by Mordukhovich is defined as

$$N_A(\overline{x}) = \left\{ d^* \in \mathbb{R}^n | \exists d_k^* \to d^*, x_k \to \overline{x} (x_k \in A) : d_k^* \in \widehat{N}_A(x_k) \right\}.$$

In contrast to the regular normal cone, which is always convex, the Mordukhovich normal cone is generally nonconvex, thus cannot be polar to any tangential approximation of A. Nevertheless, $\widehat{N}_A(\overline{x})$ and $N_A(\overline{x})$ both coincide with the normal cone of convex analysis when A is convex. Additionally, the Mordukhovich normal cone is contained in the Clarke normal cone for an arbitrary closed set A, and it induces some major tools of Variational Analysis, i.e. the Mordukhovich subdifferential and coderivative.

For a lower semicontinuous function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$, the Mordukhovich subdifferential of ψ at $\overline{x} \in \mathbb{R}^n$ is defined by

$$\partial\psi(\overline{x}) = \big\{x^* \in \mathbb{R}^n | (x^*, -1) \in N_{\mathrm{epi}\psi}(\overline{x}, \psi(\overline{x}))\big\},$$

where epiψ is the epigraph of ψ. The Mordukhovich subdifferential is always nonempty and compact when ψ is locally Lipschitz continuous. Moreover

$$\partial\psi(\overline{x}) = \big\{\nabla\psi(\overline{x})\big\}$$

provided ψ is continuously differentiable. The following convex hull property

$$\mathrm{co}\partial(-\psi)(\overline{x}) = -\mathrm{co}\partial\psi(\overline{x}), \tag{1}$$

also holds true when ψ is locally Lipschitz continuous. If the function $\psi : \mathbb{R}^n \to \mathbb{R}$ is convex, then $\partial\psi(\overline{x})$ coincides with the subdifferential of convex analysis. Also, if we consider two functions ϕ and ψ, locally Lipschitz continuous around $\overline{x}$, and nonnegative real numbers λ and μ, we have the sum rule

$$\partial(\lambda\phi + \mu\psi)(\overline{x}) \subseteq \lambda\partial\phi(\overline{x}) + \mu\partial\psi(\overline{x}), \tag{2}$$

where equality holds if ϕ or ψ is continuously differentiable at $\overline{x}$.

For the rest of this section we consider a set-valued mapping $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, its domain denoted by domΦ is the set of all $x \in \mathbb{R}^n$ such that $\Phi(x) \neq \emptyset$ and its graph is given as

$$\mathrm{gph}\Phi = \big\{(u, v) \in \mathbb{R}^n \times \mathbb{R}^m | v \in \Phi(u)\big\}.$$

The coderivative of Φ at $(\overline{x}, \overline{y}) \in \mathrm{gph}\Phi$ is a positively homogeneous mapping $D^*\Phi(\overline{x}, \overline{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ defined for $y^* \in \mathbb{R}^m$ as

$$D^*\Phi(\overline{x}, \overline{y})(y^*) = \big\{x^* \in \mathbb{R}^n | (x^*, -y^*) \in N_{\mathrm{gph}\Phi}(\overline{x}, \overline{y})\big\}, \tag{3}$$

where the argument $\overline{y}$ is omitted if Φ is single-valued. Precisely, for a locally Lipschitz continuous function $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, we have for any $y^* \in \mathbb{R}^m$

$$D^*\Phi(\overline{x})(y^*) = \partial\langle y^*, \Phi\rangle(\overline{x}), \tag{4}$$

with $\langle y^*, \Phi\rangle(\overline{x}) = \langle y^*, \Phi(\overline{x})\rangle$ and ∂ being the Mordukhovich subdifferential defined above. Again let's mention that if the single-valued mapping Φ is continuously differentiable around $\overline{x}$, then for $y^* \in \mathbb{R}^m$

$$D^*\Phi(\overline{x})(y^*) = \big\{\nabla\Phi(\overline{x})^\top y^*\big\}, \tag{5}$$

where $\nabla\Phi(\overline{x})$ is the Jacobian matrix of Φ.

The set-valued mapping Φ is inner semicompact at a point $\overline{x}$, with $\Phi(\overline{x}) \neq \emptyset$, if for every sequence $x_k \to \overline{x}$ with $\Phi(x_k) \neq \emptyset$, there is a sequence of $y_k \in \Phi(x_k)$

that contains a convergent subsequence as $k \to \infty$. It follows that the inner semicompactness holds whenever Φ is uniformly bounded around $\overline{x}$, i.e. there exists a neighborhood U of $\overline{x}$ and a bounded set $A \subset \mathbb{R}^m$ such that $\Phi(x) \subseteq A$, $\forall x \in U$. The mapping Φ is said to be inner semicontinuous at $(\overline{x}, \overline{y}) \in \mathrm{gph}\Phi$ if for every sequence $x_k \to \overline{x}$ there is a sequence of $y_k \in \Phi(x_k)$ that converges to $\overline{y}$ as $k \to \infty$. Clearly, the inner semicontinuity is a property much stronger than the inner semicompactness and is a necessary condition for the following Aubin property to hold. The mapping $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ satisfies the Aubin property around the point $(\overline{x}, \overline{y}) \in \mathrm{gph}\Phi$ if there are neighborhoods U of $\overline{x}$, V of $\overline{y}$ and a constant $L > 0$ such that

$$d(y, \Phi(x_2)) \leq L\|x_1 - x_2\|, \ \forall x_1, x_2 \in U, \ \forall y \in \Phi(x_1) \cap V,$$

where d stands for a distance on $\mathbb{R}^m \times \mathbb{R}^m$. When the graph of Φ is closed, the Aubin property is equivalent to the so-called coderivative (or Mordukhovich) criterion:

$$D^*\Phi(\overline{x}, \overline{y})(0) = \{0\}. \tag{6}$$

3 Optimal value versus KKT reformulation

We are mainly concerned in this section with comparing the optimal value reformulation and KKT reformulation of the bilevel programming problem in terms of constraint qualifications, stationary points, local and global optimal solutions. We first start by presenting these two reformulations and the links between their local and global optimal solutions.

3.1 Reformulations and optimal solutions

We consider the optimistic bilevel programming problem to

$$\text{minimize } F(x, y) \text{ subject to } x \in X \subseteq \mathbb{R}^n, y \in \Psi(x), \tag{7}$$

also called the upper level problem, where the function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and the set-valued mapping Ψ from $\mathbb{R}^n$ to $\mathbb{R}^m$ describes the solution set of the parametric optimization problem to

$$\text{minimize } f(x, y) \text{ subject to } y \in K(x), \tag{8}$$

$K(x)$ being a closed subset of $\mathbb{R}^m$ for all $x \in \mathbb{R}^n$ and the function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ continuously differentiable. The difference between what is called the optimal value reformulation of problem (7) and its so-called KKT reformulation resides in the way the solution set $\Psi(x)$ of the lower level problem (8) is expressed in order to have a one level optimization problem.

For the optimal value reformulation,

$$\Psi(x) = \{y \in K(x) | f(x, y) \leq \varphi(x)\}, \tag{9}$$

where φ is the optimal value function of problem (8), defined as

$$\varphi(x) = \min\{f(x, y) | y \in K(x)\}. \tag{10}$$

Hence the following reformulation of problem (7):

$$\text{minimize } F(x, y) \text{ subject to } (x, y) \in C, \tag{11}$$

where

$$C = \{(x, y) \in \Omega | f(x, y) \leq \varphi(x)\} \tag{12}$$

with $\Omega = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m | x \in X, y \in K(x)\}$.

Theorem 3.1. *A point $(\overline{x}, \overline{y})$ is a local (resp. global) optimal solution of (7) if and only if it is a local (resp. global) optimal solution of (11).*

Proof. Let $(\overline{x}, \overline{y})$ be a local optimal solution of (7), i.e. let $U(\overline{x}, \overline{y})$ be a neighborhood of $(\overline{x}, \overline{y})$ such that

$$F(\overline{x}, \overline{y}) \leq F(x, y), \ \forall (x, y) \in \mathcal{F} \cap U(\overline{x}, \overline{y}), \tag{13}$$

where $\mathcal{F}$ is the feasible set of problem (7). Since $(x, y) \in \mathcal{F}$ implies $x \in X$, $y \in \Psi(x)$ and

$$y \in \Psi(x) \text{ if and only if } y \in K(x), \ f(x, y) \leq \varphi(x), \tag{14}$$

it follows that $(\overline{x}, \overline{y}) \in C$. It also follows from (14) that for every $(x, y) \in C$, $(x, y) \in \mathcal{F}$. Hence $(\overline{x}, \overline{y})$ is also a local optimal solution of problem (11), considering (13). The global case follows in the like manner. The converse can also be proven analogously thanks to (14). □

Meanwhile, if we assume the parametric problem (8) to be convex, i.e., the function $y \to f(x, y)$ and the set $K(x)$ are convex for all $x \in X$, then the lower level solution set takes the form

$$\Psi(x) = \{y \in \mathbb{R}^m | 0 \in \nabla_y f(x, y) + N_{K(x)}(y)\}. \tag{15}$$

Furthermore we consider the following set-valued mapping defined from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^m$

$$N_K(x, y) = \begin{cases} N_{K(x)}(y) & \text{if } y \in K(x) \\ \emptyset & \text{otherwise} \end{cases} \tag{16}$$

suggested by Dempe and Dutta [15]. Then the bilevel programming problem (7) can be reformulated as

$$\text{minimize } F(x, y) \text{ subject to } (x, y) \in C', \tag{17}$$

where

$$C' = \{(x, y) \in \Omega' | H(x, y) \in \text{gph} N_K\}, \tag{18}$$

with $\Omega' = X \times \mathbb{R}^m$ and $H(x, y) = (x, y, -\nabla_y f(x, y))^\top$.

For the next theorem, a point $\overline{x}$ will be said to be a local (resp. global) optimal solution of $\min\limits_{x \in A} f(x)$ on B, if $\overline{x} \in B$ and there exist a neighborhood $U(\overline{x})$ of $\overline{x}$ such that $f(\overline{x}) \leq f(x)$, $\forall x \in A \cap B \cap U(\overline{x})$ (resp. $f(\overline{x}) \leq f(x)$, $\forall x \in A \cap B$). Naturally, if B coincides with the whole space, this needs not be mentioned.

Theorem 3.2. *A point $(\overline{x}, \overline{y})$ is a local (resp. global) optimal solution to problem (7) if and only if $(\overline{x}, \overline{y})$ is a local (resp. global) optimal solution of problem (17) on gphK.*

Proof. Let $(\overline{x}, \overline{y})$ be a local optimal solution to problem (7) on a neighborhood $U(\overline{x}, \overline{y})$ of $(\overline{x}, \overline{y})$. Let's denote by $\mathcal{F}$ the feasible set of (7), then we have

$$F(\overline{x}, \overline{y}) \leq F(x, y), \forall (x, y) \in U(\overline{x}, \overline{y}) \cap \mathcal{F}. \tag{19}$$

Let $(x, y) \in \mathcal{F}$, we have $x \in X, y \in \Psi(x)$ and since the lower level problem is convex in y,

$$y \in \Psi(x) \text{ if and only if } \begin{cases} y \in K(x) \\ -\nabla_y f(x, y) \in N_K(x, y). \end{cases} \tag{20}$$

Thus $(\overline{x}, \overline{y}) \in C' \cap \mathrm{gph}K$. It also follows from (20) that for every $(x, y) \in C' \cap \mathrm{gph}K$, $(x, y) \in \mathcal{F}$. Hence $(\overline{x}, \overline{y})$ is a local optimal solution of problem (17) on $\mathrm{gph}K$, considering (19). The global case follows in the like manner.

Conversely, let $(\overline{x}, \overline{y})$ be a local optimal solution of (17) on $\mathrm{gph}K$ and let $U(\overline{x}, \overline{y})$ be a neighborhood of $(\overline{x}, \overline{y})$ such that

$$F(\overline{x}, \overline{y}) \leq F(x, y), \forall (x, y) \in U(\overline{x}, \overline{y}) \cap C' \cap \mathrm{gph}K. \tag{21}$$

Since $(\overline{x}, \overline{y}) \in \mathrm{gph}K$, it follows from (20) that $(\overline{x}, \overline{y}) \in \mathcal{F}$. Arguing as in the first implication of the proof we conclude that $(\overline{x}, \overline{y})$ is also a local (resp. global) optimal solution of problem (7). □

The term $\mathrm{gph}K$ in Theorem 3.2 can be dropped if we assume that the lower level feasible set is independent of the parameter x, i.e. if we set $K(x) := K$ for all $x \in X$. In this case the feasible set of problem (17) takes the form

$$C' = \{(x, y) \in X \times K | (y, -\nabla_y f(x, y)) \in \mathrm{gph}N_K\},$$

hence the previous result remains true without needing the term $\mathrm{gph}K$.

Following Theorem 3.1 and Theorem 3.2, it can be observed that by transitivity, a point $(\overline{x}, \overline{y})$ will be a local (resp. global) optimal of problem (11) if and only if $(\overline{x}, \overline{y})$ is a local (resp. global) optimal of problem (17) on $\mathrm{gph}K$. But this equivalence will be lost for local solutions, at least if $K(x)$ is defined by a system of inequalities and the normal cone $N_{K(x)}$ is to be computed instead of writing the feasible set of (17) with $\mathrm{gph}N_K$. In fact, the computation of $N_{K(x)}$ will introduce new variables, thus complicating the task of constructing a solution point of the KKT reformulation that would solve the original bilevel problem. We do not pursue this goal here. For more details on this direction, the interested reader is referred to the paper of Dempe and Dutta [7].

3.2 CQs and stationary points

Let's consider a general Lipschitz optimization problem to

$$\text{minimize } \phi(z) \text{ subject to } z \in \mathcal{C}, \tag{22}$$

with $\mathcal{C}$ defined as

$$\mathcal{C} = \{z \in \mathcal{D} | \psi(z) \in \mathcal{E}\}, \tag{23}$$

where $\mathcal{D} \subseteq \mathbb{R}^k$ and $\mathcal{E} \subseteq \mathbb{R}^l$ are closed and $\psi : \mathbb{R}^k \to \mathbb{R}^l$ and $\phi : \mathbb{R}^k \to \mathbb{R}$ are locally Lipschitz continuous functions. Next let's consider the constraint qualification denoted by MMFCQ:

$$\bigcup_{z^* \in N_{\mathcal{E}}(\psi(\overline{z})) \setminus \{0\}} D^*\psi(\overline{z})(z^*) \cap -N_{\mathcal{D}}(\overline{z}) = \emptyset \tag{24}$$

which is nothing but the coderivative criterion (6) for the set-valued mapping

$$\Phi(\lambda) = \{z \in \mathcal{D} | \psi(z) + \lambda \in \mathcal{E}\}$$

at the point $(0, \overline{z}) \in \text{gph}\Phi$ [19]. Equation (24) can also be seen as an extension of the well-known Mangasarian-Fromowitz constraint qualification (MFCQ) to nonsmooth optimization problems, see e.g. Dempe and Zemkoho [13] for an illustration. The following result giving KKT type optimality conditions for problem (22) can now be stated.

Theorem 3.3. *Let $\overline{z}$ be a local optimal solution to problem* (22) *and assume that $\overline{z}$ satisfies the MMFCQ, then*

$$0 \in \partial\phi(\overline{z}) + \bigcup_{z^* \in N_{\mathcal{E}}(\psi(\overline{z}))} D^*\psi(\overline{z})(z^*) + N_{\mathcal{D}}(\overline{z}). \tag{25}$$

Proof. Since the function ψ is locally Lipschitz continuous and the sets $\mathcal{D}$ and $\mathcal{E}$ are closed, then the feasible set $\mathcal{C}$ of problem (22) is also closed. Hence, it follows from Mordukhovich [25, Proposition 5.3] that being a local optimal solution to (22), $\overline{z}$ satisfies

$$0 \in \partial\phi(\overline{z}) + N_{\mathcal{C}}(\overline{z})$$

given that ϕ is also locally Lipschitz continuous. Now, looking the feasible set of (22) as $\mathcal{C} = \Phi(0)$, it follows from Henrion, Jourani and Outrata [18, Theorem 4.1] that

$$N_{\mathcal{C}}(\overline{z}) \subset \bigcup_{z^* \in N_{\mathcal{E}}(\psi(\overline{z}))} D^*\psi(\overline{z})(z^*) + N_{\mathcal{D}}(\overline{z})$$

under the CQ (24); and the result follows. □

The KKT type optimality condition in (25) can also be obtained under the weaker CQ

$$\bigcup_{z^* \in N_{\mathcal{E}}(\psi(\overline{z}))\setminus\{0\}} D^*\psi(\overline{z})(z^*) \cap -\text{bd}N_{\mathcal{D}}(\overline{z}) = \emptyset \tag{26}$$

where instead of the normal cone $N_{\mathcal{D}}(\overline{z})$ in (24) we consider its boundary $\text{bd}N_{\mathcal{D}}(\overline{z})$, provided that $\mathcal{D}$ is semismooth and regular (in the sense of Clarke) at $\overline{z}$ [19].

Coming back to our bilevel programming problem (7), we first consider the optimal value reformulation (11), then the CQ (24) takes the following form:

$$\partial\mathcal{G}(\overline{x}, \overline{y}) \cap -N_{\Omega}(\overline{x}, \overline{y}) = \emptyset, \tag{27}$$

with $\mathcal{G}(\overline{x}, \overline{y}) = f(\overline{x}, \overline{y}) - \varphi(\overline{x})$. It is easy to show that the CQ (27) is violated at any feasible point of problem (11) under the fairly mild assumption that the sum rule

$$\partial(\mathcal{G} + \delta_{\Omega})(\overline{z}) \subseteq \partial\mathcal{G}(\overline{z}) + \partial\delta_{\Omega}(\overline{z})$$

holds. This last inclusion is automatically satisfied if φ is locally Lipschitz continuous and finite around $\overline{z}$. For more details on condition (27) we refer the interested reader to [13].

We now apply the MMFCQ to the KKT reformulation (17) and we have the following corresponding CQ:

$$\bigcup_{z^* \in N_{\text{gph}N_K}(H(\overline{x}, \overline{y}))\setminus\{0\}} D^*H(\overline{x}, \overline{y})(z^*) \cap -N_{\Omega'}(\overline{x}, \overline{y}) = \emptyset, \tag{28}$$

which can equivalently be written as

$$\left.\begin{array}{r} 0 \in D^*H(\overline{x},\overline{y})(z^*) + N_{\Omega'}(\overline{x},\overline{y}) \\ \text{with } z^* \in N_{\mathrm{gph}N_K}(H(\overline{x},\overline{y})) \end{array}\right\} \Rightarrow z^* = 0. \tag{29}$$

Hence, we have the following corollary of Theorem 3.3 by assuming that the function F is continuously differentiable and f twice continuously differentiable.

Corollary 3.4. *Let $(\overline{x},\overline{y})$ be a local optimal solution to (17) and assume that the following CQ is satisfied:*

$$\left.\begin{array}{r} \big(-u + \nabla^2_{xy}f(\overline{x},\overline{y})^\top w,\ -v + \nabla^2_{yy}f(\overline{x},\overline{y})^\top w\big) \in N_{X\times\mathbb{R}^m}(\overline{x},\overline{y}) \\ \textit{with } (u,v,w) \in N_{gphN_K}(\overline{x},\overline{y},-\nabla_y f(\overline{x},\overline{y})) \end{array}\right\} \Rightarrow \left\{\begin{array}{l} u = 0 \\ v = 0 \\ w = 0. \end{array}\right.$$

Then there exist $(\overline{u},\overline{v},\overline{w}) \in N_{gphN_K}(\overline{x},\overline{y},-\nabla_y f(\overline{x},\overline{y}))$ and $\overline{\gamma} \in N_X(\overline{x})$ such that

(i) $-\nabla_x F(\overline{x},\overline{y}) + \nabla^2_{xy}f(\overline{x},\overline{y})^\top\overline{w} = \overline{u} + \overline{\gamma}$

(ii) $-\nabla_y F(\overline{x},\overline{y}) + \nabla^2_{yy}f(\overline{x},\overline{y})^\top\overline{w} = \overline{v}$.

Proof. We simply have to notice that since the vector-valued function H is continuously differentiable, then it follows from equality (5) that

$$D^*H(\overline{x},\overline{y})(z^*) = \nabla H(\overline{x},\overline{y})^\top z^* = \begin{bmatrix} u - \nabla^2_{xy}f(\overline{x},\overline{y})^\top w \\ v - \nabla^2_{yy}f(\overline{x},\overline{y})^\top w \end{bmatrix}$$

with $z^* = (u,v,w) \in N_{\mathrm{gph}N_K}(\overline{x},\overline{y},-\nabla_y f(\overline{x},\overline{y}))$. The result then follows given that $N_{\Omega'}(\overline{x},\overline{y}) = N_{X\times\mathbb{R}^m}(\overline{x},\overline{y}) = N_X(\overline{x}) \times \{0\}$. □

Considering the case where the lower level feasible set does not depend on the upper level variable x, the variable x makes no sense in the definition of the set-valued mapping in (16). Hence, in this situation the CQ in Corollary 3.4 takes the form

$$\left.\begin{array}{r} \big(\nabla^2_{xy}f(\overline{x},\overline{y})^\top v,\ -u + \nabla^2_{yy}f(\overline{x},\overline{y})^\top v\big) \in N_{X\times\mathbb{R}^m}(\overline{x},\overline{y}) \\ \text{with } (u,v) \in N_{\mathrm{gph}N_K}(\overline{y},-\nabla_y f(\overline{x},\overline{y})) \end{array}\right\} \Rightarrow [u = 0, v = 0], \tag{30}$$

where $K(x) := K$ for all $x \in X$, cf. Dutta and Dempe [15].

The result of Corollary 3.4 was also given in [10] at the difference that we present it here as a consequence of the more general result of Theorem 3.3. It is also worth mentioning that if the upper level feasible set X is convex, then as mentioned above, the optimality conditions in Corollary 3.4 can still be obtained if instead of the normal cone $N_{X\times\mathbb{R}^m}(\overline{x},\overline{y})$ of the CQ, we consider its boundary $\mathrm{bd}N_{X\times\mathbb{R}^m}(\overline{x},\overline{y})$, thus providing a weaker CQ.

We now provide an example of bilevel programming problem, inspired from [10], and presenting a difference of behavior between the optimal value reformulation (11) and the KKT reformulation (17) under the constraint qualification (24) also denoted above as MMFCQ.

Example 1. *Consider the bilevel programming problem (7) with the functions $F, f : \mathbb{R}^2 \to \mathbb{R}$ defined respectively as $F(x,y) = (x-1)^2 + y^2$ and $f(x,y) = (x^2+1)y$. The upper and the lower level feasible sets are respectively defined as $X = \mathbb{R}_+ = [0,\infty)$ and $K(x) = \mathbb{R}_+$.*

Let's first mention that $\Psi(x) = \{0\}$ for all $x \in X$, hence $(1,0)$ is the unique solution of the problem. Next we consider the KKT reformulation (17) and its corresponding MMFCQ as in (30) given that $K(x)$ is independent of x. Since $N_X(1) = \{0\}$ the CQ (30) is satisfied at $(1,0)$ given that $\nabla^2_{xy} f(1,0) = 2$.

Finally, for the optimal value reformulation (11), we have $\varphi(x) = 0$ for all $x \in X$ and $\Omega = \mathbb{R}_+ \times \mathbb{R}_+$. Hence $\nabla \mathcal{G}(1,0) = (0,1)^\top$ and $N_\Omega(1,0) = N_{\mathbb{R}_+}(1) \times N_{\mathbb{R}_+}(0) = \{0\} \times \mathbb{R}_-$. Clearly we have $\nabla \mathcal{G}(1,0) \in \{0\} \times \mathbb{R}_+ = -N_\Omega(1,0)$. Thus illustrating the violation of the MMFCQ at $(1,0)$ for the optimal value reformulation (11).

Our next concern is to compare the stationary points of the KKT reformulation (17) and the optimal value reformulation (11) of problem (7). But given that estimating the normal cone $N_{\text{gph}N_K}$ may be difficult (cf. [15] and references therein), we consider for the rest of this section, a simplified bilevel programming problem (7) where

$$X = \{x \in \mathbb{R}^n | G(x) \leq 0\}, \text{ and } K(x) = \{y \in \mathbb{R}^m | g(x,y) \leq 0\},$$

with the functions $G : \mathbb{R}^n \to \mathbb{R}^p$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ being sufficiently smooth as required in the sequel. If in addition to the convexity of the function $y \to f(x,y)$ and the set $K(x)$, i.e. of $y \to g(x,y)$, an appropriate CQ, like the smooth MFCQ is satisfied at all points $y \in \Psi(x), x \in X$, then we have the following equality [28]:

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

Hence the KKT reformulation takes the form:

$$\begin{array}{l} \text{minimize } F(x,y) \\ \text{subject to } \begin{cases} \nabla_y f(x,y) + \nabla_y g(x,y)^\top \beta = 0 \\ \beta^\top g(x,y) = 0 \\ \beta \geq 0, g(x,y) \leq 0 \\ G(x) \leq 0 \\ (x,y,\beta) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k. \end{cases} \end{array} \tag{31}$$

If we apply the MMFCQ to this problem by setting $\mathcal{D} = \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k$, $\mathcal{E} = \mathbb{R}^{p+2k} \times \{0_{m+1}\}$ and the components of ψ being defined by the constraint functions of problem (31) we obtain a CQ which coincides with the dual form of the ordinary Mangasarian-Fromowitz CQ for smooth problems. Unfortunately, this CQ is violated at an arbitrary feasible point of (31), cf. Scheel and Scholtes [30]. This seems to contradict what has been mentioned in Example 1, where the problem considered clearly falls in the same class of problems in (31). For an attempt to interpret this fact, we properly write the KKT reformulation of the problem of Example 1 in the classical form of (31):

$$\begin{array}{l} \text{minimize } (x-1)^2 + y^2 \\ \text{subject to } x \geq 0,\ y = 0. \end{array} \tag{32}$$

This clearly illustrates that the classical MFCQ is satisfied at $(1,0)$ given that the constraints $x \geq 0$, $y = 0$ are all linear. But the reason for this can be attributed to the disappearance of the variable β, which is supposed to help preserve the complementarity constraint in (31) causing the failure of the MFCQ.

Hence, the problem of Example 1 should be seen not as a legitimization of how good the MFCQ could be as a CQ for the bilevel programming problem, but rather as an artificial problem designed simply with the purpose to show that for a given bilevel optimization problem, the behavior vis-à-vis a CQ, may vary depending on the reformulation considered. Nevertheless, a formal relation between the outcomes of the MMFCQ on the (primal) KKT model (17) and the KKT reformulation in (31) needs to be established, in order to have an overview on the situation. But we do not intend to address this issue here, rather we keep it as topic for future research.

In order to derive KKT type optimality conditions for problem (31), we consider the Guignard CQ defined as follows and which is weaker than the MFCQ. For an optimization problem to

$$\begin{array}{l}\text{minimize } f(x)\\ \text{subject to } g(x) \leq 0,\ h(x) = 0,\end{array} \tag{33}$$

where the functions $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^p$ and $h : \mathbb{R}^n \to \mathbb{R}^q$ are continuously differentiable, if we denote by A the feasible set of problem (33), the Guignard CQ is satisfied at $\overline{x} \in A$ if

$$K_A(\overline{x})^* = -\widehat{N}_A(\overline{x}), \tag{34}$$

where the set in the left hand side of the equality denotes the dual cone of the linearized tangent cone to A at $\overline{x} \in A$:

$$K_A(\overline{x}) = \{d \in \mathbb{R}^n |\quad \begin{array}{l}\nabla g_i(\overline{x})^\top d \leq 0, \forall i : \ g_i(\overline{x}) = 0\\ \nabla h_i(\overline{x})^\top d = 0, \forall i : \ i = 1, \ldots, q\}.\end{array}$$

Theorem 3.5. *Let $(\overline{x}, \overline{y}, \beta)$ be a local optimal solution to problem (31) and assume that the Guignard CQ is satisfied at $(\overline{x}, \overline{y}, \beta)$. Then we have $\beta \geq 0$ and there exist $\lambda \geq 0$, $\alpha \geq 0$, $\gamma \geq 0$ and ν such that:*

$$\begin{aligned}
\nabla_x F(\overline{x}, \overline{y}) + \nabla_x g(\overline{x}, \overline{y})^\top(\alpha - \lambda\beta) + \nabla G(\overline{x})^\top \gamma &\\
+ \left[\nabla^2_{xy} f(\overline{x}, \overline{y}) + \nabla^2_{xy} g(\overline{x}, \overline{y})^\top \beta\right]^\top \nu &= 0 \quad (35)\\
\nabla_y F(\overline{x}, \overline{y}) + \nabla_y g(\overline{x}, \overline{y})^\top(\alpha - \lambda\beta) + \left[\nabla^2_{yy} f(\overline{x}, \overline{y}) + \nabla^2_{yy} g(\overline{x}, \overline{y})^\top \beta\right]^\top \nu &= 0 \quad (36)\\
\nabla_y f(\overline{x}, \overline{y}) + \nabla_y g(\overline{x}, \overline{y})^\top \beta &= 0 \quad (37)\\
\alpha^\top g(\overline{x}, \overline{y}) &= 0 \quad (38)\\
\beta^\top g(\overline{x}, \overline{y}) &= 0 \quad (39)\\
\gamma^\top G(\overline{x}) &= 0. \quad (40)
\end{aligned}$$

Proof. Let us set

$$\phi(x, y, \beta) = (-\beta^\top g(x, y),\ g(x, y),\ G(x), -\beta)^\top$$

and

$$\psi(x, y, \beta) = \nabla_y f(x, y) + \nabla_y g(x, y)^\top \beta,$$

then $(\overline{x}, \overline{y}, \beta)$ is a local optimal solution of the problem

$$\begin{array}{l}\text{minimize } F(x, y)\\ \text{subject to } \phi(x, y, \beta) \leq 0,\ \psi(x, y, \beta) = 0,\end{array} \tag{41}$$

since $\beta^\top g(x,y) = 0$ if and only if $-\beta^\top g(x,y) \le 0$, given that $\beta \ge 0$ and $g(x,y) \le 0$. Further let's denote by A the feasible set of the previous problem. It follows from Mordukhovich [25, Proposition 5.1] that

$$-\nabla F(\overline{x}, \overline{y}, \beta) \in \widehat{N}_A(\overline{x}, \overline{y}, \beta).$$

Applying the Guignard CQ, we have

$$\nabla F(\overline{x}, \overline{y}, \beta) \in K_A(\overline{x}, \overline{y}, \beta)^*.$$

That is

$$-\nabla F(\overline{x}, \overline{y}, \beta)^\top d \le 0, \ \forall d \in K_A(\overline{x}, \overline{y}, \beta).$$

Hence, it follows from Farkas' Theorem of the alternative [20] that there exist $\mu = (\lambda, \alpha, \gamma)^\top \ge 0$ and ν such that

$$\nabla F(\overline{x}, \overline{y}) + \nabla \phi(\overline{x}, \overline{y}, \beta)^\top \mu + \nabla \psi(\overline{x}, \overline{y}, \beta)^\top \nu = 0 \quad (42)$$

$$\nabla_y f(\overline{x}, \overline{y}) + \nabla_y g(\overline{x}, \overline{y})^\top \beta = 0 \quad (43)$$

$$\alpha^\top g(\overline{x}, \overline{y}) = 0 \quad (44)$$

$$\beta^\top g(\overline{x}, \overline{y}) = 0 \quad (45)$$

$$\gamma^\top G(\overline{x}) = 0, \quad (46)$$

by including the feasibility of $(\overline{x}, \overline{y}, \beta)$ to problem (31) and considering only the derivative w.r.t. x and y in equality (42). It should also be clear that equality (44) is due to the definition of $K_A(\overline{x}, \overline{y}, \beta)$ and by setting $\alpha_i = 0$, for all i such that $g_i(\overline{x}, \overline{y}) < 0$. It is also worth mentioning that considering $-\beta^\top g(x,y) \le 0$ in (41) simply helps to obtain the nonnegativity of the multiplier λ. The result then follows by decomposing (42) into the partial derivative w.r.t. x and y respectively. □

We now give an example of bilevel programming problem showing that the Guignard CQ has some real chances to hold.

Example 2. *We consider the problem to*

$$\begin{array}{l} \textit{minimize } x + y \\ \textit{subject to } \left\{ \begin{array}{l} x \ge 0 \\ \quad \textit{minimize } xy \\ \quad \textit{subject to } y \ge 0. \end{array} \right. \end{array}$$

In this case, $f(x,y) = xy$, $X = \{x \ge 0\} \subset \mathbb{R}$ *and* $K(x) = \{y \ge 0\} \subset \mathbb{R}$. *Hence,* $-\nabla_y f(x,y) \in N_{K(x)}(y)$ *is equivalent to* $x \ge 0$ *and* $xy = 0$, *given that* $N_{K(x)}(y) = \{-\beta | \beta \ge 0, \beta y = 0\}$. *The above problem then takes the form*

$$\begin{array}{l} \textit{minimize } x + y \\ \textit{subject to } x, y \ge 0, xy = 0 \end{array} \quad (47)$$

and following Flegel [17], the Guignard CQ holds at the unique optimal solution point $(0,0)$.

An important point to make here is that even though the variable β does not appear in the KKT reformulation (47), the MFCQ still fails here, which is not the case for (32).

For some more details on the application of the Guignard CQ to MPECs (mathematical programming problems with equilibrium constraints), we refer the interested reader to Flegel [17]. We also remaind that problem (31) is a special MPEC.

As far as the stationary points of the optimal value reformulation (11) are concerned, we have already shown above that if we plug the data of (11) in the MMFCQ, the latter is violated at an arbitrary feasible point provided simply that φ is locally Lipschitz continuous and finite around the corresponding point. By passing from the normal cone N_Ω to its boundary $\mathrm{bd}N_\Omega$, i.e. instead of the CQ (27), the weaker CQ:

$$\partial\mathcal{G}(\overline{x},\overline{y}) \cap -\mathrm{bd}N_\Omega(\overline{x},\overline{y}) = \emptyset \tag{48}$$

may have some chances to hold for problem (11). Hence the following result.

Theorem 3.6. *Let $(\overline{x},\overline{y})$ be a local optimal solution to problem (11) and assume that the CQ (48) is satisfied at $(\overline{x},\overline{y})$. We also let φ be convex and finite on $\mathbb{R}^n$. We further assume that the functions G and g are convex and there exist $(\widetilde{x},\widetilde{y})$ such that $G(\widetilde{x}) < 0$ and $g(\widetilde{x},\widetilde{y}) < 0$. Then there exist $\lambda \geq 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\gamma \geq 0$ such that:*

$$\nabla_x F(\overline{x},\overline{y}) + \nabla_x g(\overline{x},\overline{y})^\top(\alpha - \lambda\beta) + \nabla G(\overline{x})^\top\gamma = 0 \tag{49}$$

$$\nabla_y F(\overline{x},\overline{y}) + \nabla_y g(\overline{x},\overline{y})^\top(\alpha - \lambda\beta) = 0 \tag{50}$$

$$\nabla_y f(\overline{x},\overline{y}) + \nabla_y g(\overline{x},\overline{y})^\top\beta = 0 \tag{51}$$

$$\alpha^\top g(\overline{x},\overline{y}) = 0 \tag{52}$$

$$\beta^\top g(\overline{x},\overline{y}) = 0 \tag{53}$$

$$\gamma^\top G(\overline{x}) = 0. \tag{54}$$

Proof. The convexity of the functions G and g imply the convexity of the set $\Omega = \{(x,y)|G(x) \leq 0, g(x,y) \leq 0\}$. Next, being finite and convex on $\mathbb{R}^n$, the value function φ is locally Lipschitz continuity around $\overline{x}$. Hence, it follows from Dempe and Zemkoho [13, Lemma 3.3], Mordukhovich [25, Proposition 5.3] and the sum rule (2) that there exist $\lambda \geq 0$ such that

$$0 \in \nabla F(\overline{x},\overline{y}) + \lambda\nabla f(\overline{x},\overline{y}) + \lambda\partial(-\varphi)(\overline{x}) \times \{0\} + N_\Omega(\overline{x},\overline{y}). \tag{55}$$

Now, let $x^* \in \partial\varphi(\overline{x})$. Since $\partial\varphi(\overline{x})$ coincides with the subdifferential of φ in the sense of convex analysis (as mentioned in Section 1), then

$$\varphi(x) - \varphi(\overline{x}) \geq \langle x^*, x - \overline{x}\rangle, \forall x \in \mathbb{R}^n.$$

Considering the definition of φ including the fact that $\overline{y} \in \Psi(\overline{x})$ (since $(\overline{x},\overline{y})$ is feasible to (11)) we have

$$f(x,y) - \langle x^*, x\rangle \geq f(\overline{x},\overline{y}) - \langle x^*, \overline{x}\rangle, \forall(x,y) : g(x,y) \leq 0,$$

which means that $(\overline{x},\overline{y})$ is an optimal solution of the problem to

$$\text{minimize } f(x,y) - \langle x^*, x\rangle \text{ subject to } g(x,y) \leq 0.$$

Hence, from the classical Lagrange multiplier rule, there exist $\beta \geq 0$ such that

$$x^* = \nabla_x f(\overline{x},\overline{y}) + \nabla_x g(\overline{x},\overline{y})^\top\beta \tag{56}$$

$$\nabla_y f(\overline{x},\overline{y}) + \nabla_y g(\overline{x},\overline{y})^\top\beta = 0 \tag{57}$$

$$\beta^\top g(\overline{x},\overline{y}) = 0 \tag{58}$$

given that there exist $(\widetilde{x}, \widetilde{y})$ such that $g(\widetilde{x}, \widetilde{y}) < 0$.

On the other hand, given the existence of $(\widetilde{x}, \widetilde{y})$ such that $G(\widetilde{x}) < 0$ and $g(\widetilde{x}, \widetilde{y}) < 0$, it follows from Rockafellar [28, Theorem 4.3] that

$$N_\Omega(\overline{x}, \overline{y}) \subseteq \Big\{\nabla g(\overline{x}, \overline{y})^\top \alpha + (\nabla G(\overline{x})^\top \gamma, 0) | \alpha \geq 0, \alpha^\top g(\overline{x}, \overline{y}) = 0 \\ \gamma \geq 0, \gamma^\top G(\overline{x}) = 0\Big\}. \tag{59}$$

Combining (55)-(59), there exist $\lambda \geq 0$, $\alpha \geq 0$, $\beta \geq 0$ and $\gamma \geq 0$ satisfying (51)-(54) and

$$\nabla_x F(\overline{x}, \overline{y}) + \nabla_x g(\overline{x}, \overline{y})^\top (\alpha - \lambda\beta) + \nabla G(\overline{x})^\top \gamma = 0 \tag{60}$$

$$\nabla_y F(\overline{x}, \overline{y}) + \lambda \nabla_y f(\overline{x}, \overline{y}) + \nabla_y g(\overline{x}, \overline{y})^\top \alpha = 0 \tag{61}$$

given that $\partial(-\varphi)(\overline{x}) \subseteq -\partial\varphi(\overline{x})$, following (1) and considering the convexity of φ. By inserting the expression of $\nabla_y f(\overline{x}, \overline{y})$ from (57) in equation (61), we have the result. □

For φ to be convex and finite on $\mathbb{R}^n$, one can assume that f is convex and $X = \mathbb{R}^n = \text{dom}\Psi$.

Remark 3.7. *Something interesting about the result in Theorem 3.6 is that on the initial optimal value reformulation (11), the Slater CQ is violated at an arbitrary feasible point because the optimal value constraint $f(x, y) \leq \varphi(x)$ is in fact an equality. But thanks to the CQ (48), this difficulty is discarded, thus paving the way for the Slater CQ to be brought in as a second CQ to help estimate the normal cone to Ω. As we will realize in the next sections, the partial calmness acts almost identically.*

In a natural way, we define the system of equations (49)-(54) of Theorem 3.6 as the KKT type optimality conditions of the optimal value reformulation (11), while the equations in (35)-(40) are considered as the KKT type optimality conditions of the KKT reformulation (31). Hence, the following consequences are obvious.

- Let the point $(\overline{x}, \overline{y}, \alpha, \beta, \gamma, \lambda, \nu)$ satisfy (35)-(40) with $\nu = 0$, then the point $(\overline{x}, \overline{y}, \alpha, \beta, \gamma, \lambda)$ satisfies (49)-(54).
- If the point $(\overline{x}, \overline{y}, \alpha, \beta, \gamma, \lambda)$ satisfies (49)-(54), then $(\overline{x}, \overline{y}, \alpha, \beta, \gamma, \lambda, 0)$ satisfies (35)-(40).
- If the lower level functions f and g are linear in (x, y), then the point $(\overline{x}, \overline{y}, \alpha, \beta, \gamma, \lambda)$ satisfies (35)-(40) if and only if it satisfies (49)-(54), given that in this case the second order derivative terms will disappear from (35)-(40).

Part of the above observations have also been made by Ye [36] in a different context.

Before we conclude this section, it may be of a particular interest to mention that when the lower level problem is not convex, it would not be legitimate to attempt solving the bilevel programming problem by using the KKT reformulation (31). As a matter of fact, Mirrlees [23] showed that in such a situation, one may not be able to identify an optimal solution of the initial problem. Moreover, the optimal solution of the bilevel programming problem may not even satisfy the optimality conditions derived using the KKT reformulation. In order to solve this difficulty and some possibly unwanted behavior that may occur using the optimal value reformulation, Ye and Zhu [39] recently suggested a combination of the KKT and the optimal value reformulations in order to obtain optimality conditions for the bilevel programming problem. Concretely, it is assumed that the KKT conditions of the lower level problem are satisfied without considering the convexity assumption above and hence the following one level optimization problem is considered:

$$\begin{array}{ll}\text{minimize } F(x,y) & \\ \text{subject to} & \left\{\begin{array}{l} f(x,y)-\varphi(x)\leq 0\\ \nabla_y f(x,y)+\nabla_y g(x,y)^\top\beta=0\\ \beta^\top g(x,y)=0\\ \beta\geq 0, g(x,y)\leq 0\\ G(x)\leq 0\\ (x,y,\beta)\in\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^k.\end{array}\right.\end{array}$$

The techniques used to design KKT type optimality conditions for the MPECs are then applied to the above problem.

To close this section, we recall that in general, the smooth MFCQ is violated at any feasible point of the KKT reformulation (31) of the bilevel programming problem (7). Thus, also implying the failure of the smooth Slater and the linear independence CQ. The same observation can be made for the optimal value reformulation (11), considering the nonsmooth versions of the MFCQ, the Slater and the linear independence CQ [13, 37]. An interesting question is how would the other CQs behave for both reformulations. Already Chen [4] showed that the smooth version of the Arrow-Hurwicz-Uzawa fails for a class of bilevel problems, considering the KKT reformulation. So it could be interesting to see how it behaves for the optimal value reformulation. It is also worth mentioning that the failure of the MMFCQ for the KKT model (31) is due to the complementarity constraint $\beta^\top g(x,y)=0$, while its failure for the optimal value model (11) is caused by the optimal value constraint $f(x,y)\leq\varphi(x)$. A possible way to avoid this common problem to the two reformulations consist, for each reformulation, to move the corresponding constraint to the leader's objective. In the next section, we consider the partial calmness, a CQ that will be shown to be very effective in doing exactly that.

4 Partial calmness

From now on we focus our attention on the optimal value reformulation of the bilevel programming problem (7). That is the problem to

$$\begin{array}{ll}\text{minimize } F(x,y) & \\ \text{subject to} & \left\{\begin{array}{l} f(x,y)\leq\varphi(x)\\ x\in X, y\in K(x).\end{array}\right.\end{array}\tag{62}$$

Let $(\overline{x},\overline{y})$ be a feasible point of (62). Problem (62) is partially calm at $(\overline{x},\overline{y})$ if there is a number $\mu>0$ and a neighborhood U of $(\overline{x},\overline{y},0)$ such that

$$F(x,y)-F(\overline{x},\overline{y})+\mu|u|\geq 0,\tag{63}$$

for all $(x, y, u) \in U$ feasible to the partially perturbed problem to

$$\begin{array}{l} \text{minimize } F(x,y) \\ \text{subject to } \left\{ \begin{array}{l} f(x,y) - \varphi(x) + u = 0 \\ x \in X, y \in K(x). \end{array} \right. \end{array} \tag{64}$$

As we already mentioned in the conclusion of the previous section, our interest to the partial calmness is led by its capacity to help move the optimal value constraint function $(x, y) \to f(x, y) - \varphi(x)$ from the feasible set to the upper level objective function, thus providing an order one exact penalty function. This paves the way to more tractable constraints in the perspective of KKT type optimality conditions for the bilevel programming problem.

Theorem 4.1. *[37] Let $(\overline{x}, \overline{y})$ be a local optimal solution of problem (62). This problem is partially calm at $(\overline{x}, \overline{y})$ if and only if there exists $\mu > 0$ such that $(\overline{x}, \overline{y})$ is a local optimal solution of the partially penalized problem to*

$$\begin{array}{l} \textit{minimize } F(x,y) + \mu(f(x,y) - \varphi(x)) \\ \textit{subject to } x \in X, y \in K(x). \end{array} \tag{65}$$

From this result, it is clear that the optimality conditions derived in Theorem 3.6 could also be obtained if we replace the CQ (48) by the partial calmness. But a difference in the multipliers needs to be pointed. In fact, under the partial calmness, the multiplier λ of Theorem 3.6 is positive, even though the partial calmness is a weaker CQ than the CQ in (48) as shown by Dempe and Zemkoho [13].

In their seminal paper [37] where Ye and Zhu introduced the partial calmness, it was proven that a bilevel programming problem with a lower level problem linear in (x, y), is partially calm. We show in the next theorem that this proof can be adapted to the case where the follower's problem is linear only in the lower level variable y. Clearly we consider the optimistic bilevel programming problem to

$$\text{minimize } F(x,y) \text{ subject to } x \in \mathbb{R}^n, y \in \Psi(x), \tag{66}$$

where the set-valued mapping Ψ from $\mathbb{R}^n$ to $\mathbb{R}^m$ describes the solution set of the parametric optimization problem to

$$\text{minimize } a(x)^\top y + b(x) \text{ subject to } C(x)y \leq d(x), \tag{67}$$

with $a : \mathbb{R}^n \to \mathbb{R}^m$, $b : \mathbb{R}^n \to \mathbb{R}$, $C : \mathbb{R}^n \to \mathbb{R}^{p\times m}$ and $d : \mathbb{R}^n \to \mathbb{R}^p$. Next we consider the following optimal value reformulation of problem (66):

$$\begin{array}{l} \text{minimize } F(x,y) \\ \text{subject to } \left\{ \begin{array}{l} a(x)^\top y + b(x) \leq \varphi(x) \\ C(x)y \leq d(x). \end{array} \right. \end{array} \tag{68}$$

Theorem 4.2. *Let $(\overline{x}, \overline{y})$ be an optimal solution to (68), we assume that F is Lipschitz continuous and dom$\Psi = \mathbb{R}^n$. Then (68) is partially calm at $(\overline{x}, \overline{y})$.*

Proof. Consider a neighborhood U of $(\overline{x}, \overline{y}, 0)$ and let $(x', y', u) \in U$ satisfying

$$\begin{array}{l} a(x')^\top y' + b(x') - \varphi(x') + u = 0 \\ C(x')y' \leq d(x'). \end{array} \tag{69}$$

Since $\mathrm{dom}\Psi = \mathbb{R}^n$, let $y(x')$ be a solution to the lower level problem (67) for the parameter x', i.e. the couple $(x', y(x'))$ is feasible to problem (68). Further let $y_o(x')$ be the projection of y' on $\Psi(x')$ then, denoting by $e = (1, \ldots, 1)^\top$ a m-dimensional vector, we have

$$\begin{array}{rcl} \|y' - y_o(x')\| & = & \min\limits_{y}\{\|y' - y\| : y \in \Psi(x')\} \\ & = & \min\limits_{\varepsilon, y}\{\varepsilon : \|y' - y\| \leq \varepsilon, y \in \Psi(x')\} \\ & = & \min\limits_{\varepsilon, y}\{\varepsilon : -\varepsilon e \leq y' - y \leq \varepsilon e, y \in \Psi(x')\}. \end{array}$$

The last equality describes the linear program

$$\begin{array}{l} \min\limits_{\varepsilon, y} \varepsilon \\ \begin{cases} -\varepsilon e + y \leq y' \\ -\varepsilon e - y \leq -y' \\ a(x')^\top y \leq \varphi(x') - b(x') \\ C(x')y \leq d(x'), \end{cases} \end{array}$$

having as dual the problem

$$\begin{array}{l} \max\limits_{\xi_1, \xi_2, \xi_3, \xi_4} \left\langle (\xi_1, \xi_2, \xi_3, \xi_4)^\top, (y', -y', \varphi(x') - b(x'), d(x'))^\top \right\rangle \\ \begin{cases} -e^\top \xi_1 - e^\top \xi_2 = 1 \\ \xi_1 - \xi_2 + \xi_3 a(x') + C(x')^\top \xi_4 = 0 \\ (\xi_1, \xi_2, \xi_3, \xi_4) \in \mathbb{R}^m_- \times \mathbb{R}^m_- \times \mathbb{R}_- \times \mathbb{R}^p_-. \end{cases} \end{array}$$

By inserting the constraint $\xi_1 - \xi_2 + \xi_3 a(x') + C(x')^\top \xi_4 = 0$ in the objective function of the dual problem, we have the equivalent problem

$$\begin{array}{l} \max\limits_{\xi_1, \xi_2, \xi_3, \xi_4} \xi_3\big(\varphi(x') - a(x')^\top y' - b(x')\big) + \big(d(x') - C(x')y'\big)^\top \xi_4 \\ -e^\top \xi_1 - e^\top \xi_2 = 1,\ \xi_1, \xi_2, \xi_3, \xi_4 \leq 0. \end{array}$$

Thus there is at least one vertex $(\xi_1^o, \xi_2^o, \xi_3^o, \xi_4^o)$ of the system

$$-e^\top \xi_1 - e^\top \xi_2 = 1,\ \xi_1, \xi_2, \xi_3, \xi_4 \leq 0 \tag{70}$$

such that

$$\|y' - y_o(x')\| = \xi_3^o\big(\varphi(x') - a(x')^\top y' - b(x')\big) + \big(d(x') - C(x')y'\big)^\top \xi_4^o,$$

which implies

$$\|y' - y_o(x')\| \leq \xi_3^o u$$

given that (x', y') satisfies (69). Also notice that $u \leq 0$ considering the definition of φ. Since the number of vertices satisfying (70) is finite, let $\xi_3^B \in \mathbb{R}$ be the smallest $(2m+1)^{th}$ component of such vertices, then

$$\|y' - y_o(x')\| \leq |\xi_3^B||u|. \tag{71}$$

On the other hand we recall that F is Lipschitz continuous. Denote by K_F its Lipschitz constant, given that $(\overline{x}, \overline{y})$ is an optimal solution to problem (68) and $(x', y_o(x'))$ being feasible, we have

$$\begin{array}{rcl} F(x', y') - F(\overline{x}, \overline{y}) & \geq & F(x', y') - F(x', y_o(x')) \\ & \geq & -K_F\|y' - y_o(x')\| \\ & \geq & -\mu|u|, \end{array} \tag{72}$$

where $\mu = K_F|\xi_3^B|$ and considering inequality (71). □

In [12], a similar proof was already given for the simpler case where $f(x,y) = x^\top y$ and $K(x) = \{y|Ay = b, y \geq 0\}$. It is easy to see that this case can simply be imbedded in Theorem 4.2 by rearranging the lower level constraints as: $Ay \leq b$, $-Ay \leq -b$ and $-y \leq 0$.

We now consider the bilevel programming problem (66) in the case where $X \neq \mathbb{R}^n$. It is no more certain that this problem would be partially calm without an additional assumption. For this reason we consider the following definition of the notion of uniformly weak sharp minimum introduced in [37].

Definition 4.3. *The family of parametric optimization problems $\{(8)|x \in X\}$ is said to have a uniformly weak sharp minimum if there exist $\mu > 0$ such that*

$$f(x,y) - \varphi(x) \geq \mu d(y, \Psi(x)), \forall y \in K(x), \forall x \in X.$$

It can easily be shown that if the upper level objective function F is Lipschitz continuous in y uniformly in x, then the bilevel programming problem (7) is partially calm at every local optimal solution, provided that the family of parametric optimization problems $\{(8)|x \in X\}$ has a uniformly weak sharp minimum. Ye [35] considered the bilevel programming problem (7) where the lower level problem is defined as

$$\text{minimize } f(x,y) \text{ subject to } y \in \mathbb{R}^m,\ g(x,y) \leq 0, \tag{73}$$

with the functions f and g both linear in y and the following was proven:

Theorem 4.4. *Assume that $dom\Psi = X$. Let*

$$\overline{g}(x,y)^\top = (g_1(x,y), \dots, g_k(x,y), f(x,y) - \varphi(x))$$

and assume that there exists $\mu > 0$ such that

$$\begin{aligned} c(x) := \sup_{w,y',I} \Big\{ & w_{k+1} : y' \in \Psi(x), w_i > 0, g_i(x,y') = 0, \forall i \in I \\ & \|\textstyle\sum_{i\in I} w_i \nabla_y \overline{g}_i(x,y')\|_1 = 1, \\ & \textit{vectors } \{\nabla_y \overline{g}_i(x,y') : i \in I\} \textit{ are linearly independent}, \\ & \{k+1\} \subseteq I \subseteq \{1,\dots,k+1\} \Big\} \\ \leq \mu, \ & \forall x \in X \textit{ such that there exists } I \textit{ as in the previous line.} \end{aligned}$$

Then, there exist $\alpha > 0$ such that:

$$f(x,y) - \varphi(x) \geq \mu^{-1} \alpha d(y, \Psi(x)), \forall y \in K(x), \forall x \in X.$$

We can easily observe that the follower's problem (73) is nothing but the lower level problem considered in (67). Hence if $X = \mathbb{R}^n$, then the result in Theorem 4.2 remains true for the bilevel programming problem (66) where the lower level problem is defined in (73). It thus seems clear that in this case the assumption of Theorem 4.4 , which may be quite difficult to check (see Mangasarian and Shiau [21]), is not useful. Nevertheless, when $X \neq \mathbb{R}^n$, the previous result takes all its importance. For more details on sufficient conditions for the uniformly weak sharp minimum to be satisfied we refer the interested reader to Dempe and Zemkoho [13], Ye [35], Ye and Zhu [37, 38], and Ye, Zhu and Zhu [40].

5 Application to the DAP

We consider a transportation network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ and $\mathcal{A}$ denote the set of nodes and directed links (arcs), respectively. Let $\mathcal{W} \subset \mathcal{N}^2$ denote the set of origin-destination (O-D) pairs. Each O-D pair $w \in \mathcal{W}$ is connected by a set of routes (paths) $\mathcal{P}_w$, each member of which is a set of sequentially connected links. We denote by $\mathcal{P} = \bigcup_{w \in \mathcal{W}} \mathcal{P}_w$ the set of all routes of the network and by $\alpha = |\mathcal{A}|$, $\omega = |\mathcal{W}|$ and $\pi = |\mathcal{P}|$, the cardinalities of $\mathcal{A}$, $\mathcal{W}$ and $\mathcal{P}$, respectively. Let the matrix $(\Lambda = [\Lambda_{wp}]) \in \mathbb{R}^{\omega \times \pi}$ denote the O-D-route incidence matrix in which $\Lambda_{wp} = 1$ if route $p \in \mathcal{P}_w$ and $\Lambda_{wp} = 0$ otherwise, and the matrix $(\Delta = [\Delta_{ap}]) \in \mathbb{R}^{\alpha \times \pi}$ denotes the arc-route incidence matrix; here $\Delta_{ap} = 1$ if arc a is in route p and $\Delta_{ap} = 0$ otherwise. The network is assumed to be strongly connected, i.e. at least one route joins each O-D pair.

We also consider the column vectors $(d = [d_w]) \in \mathbb{R}^{\omega}$, $(q = [q_p]) \in \mathbb{R}^{\pi}_{+}$ and $(v = [v_a]) \in \mathbb{R}^{\alpha}$ to denote the travel demand, the route flow and arc flow, respectively. Finally we denote by $(c = [c_p]) \in \mathbb{R}^{\pi}_{+}$ the column vector denoting the route capacity. For a given demand d, a route flow q is feasible if it does not exceed the capacity and satisfies the O-D demand constraint $\Lambda q = d$. Let's denote by Q the set-valued mapping from $\mathbb{R}^{\omega}$ to $\mathbb{R}^{\pi}$ describing the set of such flows, then

$$Q(d) = \{q \in \mathbb{R}^{\pi}_{+} | q \leq c, \Lambda q = d\}. \tag{74}$$

For a given demand d, a link flow is feasible if there is a corresponding feasible route flow q such that the flow conservation constraint $\Delta q = v$, is satisfied. Hence, the following set-valued mapping from $\mathbb{R}^{\omega}$ to $\mathbb{R}^{\alpha}$

$$V(d) = \{v \in \mathbb{R}^{\alpha} | \exists q \in Q(d), \Delta q = v\}, \tag{75}$$

denotes the set of feasible link flows. We let the separable function t from $\mathbb{R}^{\alpha}$ to $\mathbb{R}^{\alpha}$ denote the route cost, i.e. for each $a \in \mathcal{A}$, the component $t_a(v_a)$ of the vector $t(v)$ gives the traffic cost on the arc a, under the flow v_a. We assume the route cost to be additive, thus the components of $\bar{c}(v) = \Delta^{\top} t(v)$ give the cost on each route $p \in \mathcal{P}$. Finally, we introduce the vector $\vartheta(v) = [\vartheta_w(v)] \in \mathbb{R}^{\omega}$ of minimum cost between each O-D pair $w \in \mathcal{W}$, i.e. $\vartheta_w(v) = \min_{p \in \mathcal{P}_w} \bar{c}_p(v)$.

The user equilibrium principle of Wardrop [31] states that for every O-D pair $w \in \mathcal{W}$, the travel cost of the routes utilized are equal and minimal for each individual user, i.e. for any route $p \in \mathcal{P}$ and O-D pair $w \in \mathcal{W}$, we have

$$\begin{cases} \bar{c}_p(v) = \vartheta_w(v) & \text{if } q_p > 0 \\ \bar{c}_p(v) \geq \vartheta_w(v) & \text{if } q_p = 0 \end{cases} \tag{76}$$

for any given demand d. It follows from Beckmann, McGuire and Winsten [2] that for any given demand d, obtaining Wardrop's user equilibrium is equivalent to solving the parametric optimization problem

$$\begin{array}{l} \text{minimize } f(d, v) = \sum_{a \in \mathcal{A}} \int_0^{v_a} t_a(s) ds \\ \text{subject to } v \in V(d), \end{array} \tag{77}$$

provided that each link cost function $t_a : \mathbb{R} \to \mathbb{R}$ is continuous and positive. The latter assumptions will be maintained for the rest of this section such that for each demand vector d, the Wardrop's user equilibrium arc flow will be defined as the solution of the optimization problem (77) also called the traffic assignment problem.

In transportation planing, the demand mentioned in the model of the traffic assignment model above is a strategic data in a broader sense given that a comprehensive decision making, in many other aspects, highly depends on how accurate it is estimated. Fisk [16] has suggested a bilevel formulation of the problem of estimating the origin-destination (O-D) matrix or O-D demand adjustment problem (DAP) from an outdated matrix and using some observed traffic counts:

$$\begin{array}{l} \text{minimize } F(d,v) \\ \text{subject to } d \in D, v \in \Psi(d) \end{array} \tag{78}$$

where D is a closed set and $\Psi(d)$ is the solution set of the traffic assignment problem (77) for a given O-D matrix d organized as a vector. Here the leader's cost function F may be the combination of the error measurements between the target matrix and the observed traffic flows, respectively. For more details on the model and some solution approaches, we refer the interested reader to Abrahamsson [1], Chen [4], Chen and Florian [5], Fisk [16], Migdalas [22], Yang [32] and Yang et al. [33].

One interesting thing about the traffic assignment problem (77) or the DAP model as a whole is the flow conservation constraint $\Delta q = v$ materializing the relation between the link flow v and the route flow q. A difficulty in handling this problem seems to be caused by this constraint since it is not really clear w.r.t. which variable between v and q the lower level problem should be considered. But what is perceivable is that one may find many combinations of q that would give v. Thus, the uniqueness of v would not imply that of q. We materialize this fact by explicitly defining the set of feasible route flows for a given couple of demand and link flow, cf. (86). Then, exploiting the mentioned flow conservation constraint, which also induces a special structure for the feasible set of the traffic assignment problem, defined for each demand vector d as the composition of the route flow set-valued mappings Q and the function $J : \mathbb{R}^\pi \to \mathbb{R}^\alpha$, with $J(q) = \Delta q$ such that $V(d) = J(Q(d))$, we design a new estimation of the subdifferential of the value function φ of the traffic assignment problem leading, perhaps for the first time, to KKT type optimality conditions for the demand adjustment problem.

Problem (78) can be reformulated as

$$\begin{array}{l} \text{minimize } F(d,v) \\ \text{subject to } \left\{ \begin{array}{l} f(v) \le \varphi(d) \\ d \in D, v \in V(d), \end{array} \right. \end{array} \tag{79}$$

φ being the optimal value function of the problem (77) parameterized by the demand d. Chen [4] and Chen and Florian [5] suggested Fritz John's type optimality conditions for problem (79) after a number of simplifications, including that of considering the flow conservation constraint $\Delta q = v$ as exogenous, thus dropping it in some sense. We do not make such simplifications here; hence problem (79) can be rewritten as

$$\begin{array}{l} \text{minimize } F(d,v) \\ \text{subject to } f(v) \le \varphi(d), (d,v) \in \Omega, \end{array} \tag{80}$$

where $\Omega = (D \times \mathbb{R}^\alpha) \cap \mathrm{gph} V$, with $\mathrm{gph} V$ representing the graph of the link flow set-valued mapping V. As already mentioned above $V(d) = J(Q(d))$ and we can easily observe that the graph of Q is given by

$$\text{gph}Q = \{(d, q) \in \mathbb{R}^\omega \times \mathbb{R}^\pi | -d + \Lambda q = 0,\, 0 \leq q \leq c\}. \tag{81}$$

For $(\overline{d}, \overline{q}) \in \text{gph}Q$, we consider the set

$$\mathcal{P}^o(\overline{q}) = \{r \in \mathcal{P} | \overline{q}_r = 0\} \tag{82}$$

of unused routes of the network and the set

$$\mathcal{P}^c(\overline{q}) = \{r \in \mathcal{P} | \overline{q}_r = c_r\} \tag{83}$$

of routes used at their full capacity. Then $\mathcal{P}$ can be partitioned into $\mathcal{P}^o(\overline{q}), \mathcal{P}^c(\overline{q})$ and $\mathcal{P}^u(\overline{q})$, where $\mathcal{P}^u(\overline{q})$ is the set of routes used but which are not at full capacity. Thus, $\mathcal{P} = \mathcal{P}^o(\overline{q}) \cup \mathcal{P}^c(\overline{q}) \cup \mathcal{P}^u(\overline{q})$.

To make the further explanations more clear, we make the following technical assumption: we assume that $\mathcal{P}$ is an ordered set such that for a route $r \in \mathcal{P}$, we associate an index $|r| \in \mathbb{N}$ and we define the $\pi-$dimensional vector e^r as

$$e^r = (0, \ldots, 0, 1, 0, \ldots, 0)^\top, \tag{84}$$

where 1 is at position $|r|$, in order to symbolize the utilization of the corresponding route by a network user. Next we let $[\Lambda_w]_{w\in\mathcal{W}}$ be the collection of rows of the OD-route incidence matrix Λ. Then, for $(\overline{d}, \overline{q}) \in \text{gph}Q$ the normal cone to $\text{gph}Q$ at $(\overline{d}, \overline{q})$ is given as

$$\begin{aligned} N_{\text{gph}Q}(\overline{d}, \overline{q}) = \Big\{\big(-\sum_{w\in\mathcal{W}} \lambda_w e^w,\ \sum_{r\in\mathcal{P}^c(\overline{q})} \lambda_r^c e^r - \sum_{r\in\mathcal{P}^o(\overline{q})} \lambda_r^o e^r + \sum_{w\in\mathcal{W}} \lambda_w \Lambda_w^\top\big) : \\ (\lambda_w)_{w\in W} \in \mathbb{R}^\omega,\ (\lambda_r^o)_{r\in\mathcal{P}^o(\overline{q})}, (\lambda_r^c)_{r\in\mathcal{P}^c(\overline{q})} \geq 0\Big\}, \end{aligned} \tag{85}$$

following Rockafellar [28, Theorem 4.3], where $[e^w]_{w\in\mathcal{W}}$ is the collection of columns of the identity matrix of $\mathbb{R}^{\omega\times\omega}$.

We now establish an estimation of the normal cone to the graph of the feasible link flows set-valued mapping, which will be very useful in deriving KKT type optimality conditions for the DAP.

Lemma 5.1. *We let $(\overline{d}, \overline{v}) \in gphV$; if $(d^*, v^*) \in N_{gphV}(\overline{d}, \overline{v})$, then there exists $\overline{q}$ with $\overline{v} = \Delta\overline{q}$ and $(\overline{d}, \overline{q}) \in gphQ$ such that $(d^*, \Delta^\top v^*) \in N_{gphQ}(\overline{d}, \overline{q})$.*

Proof. Let $(\overline{d}, \overline{v}) \in \text{gph}V$, by definition $(d^*, v^*) \in N_{\text{gph}V}(\overline{d}, \overline{v})$ if and only if $d^* \in D^*V(\overline{d}, \overline{q})(-v^*)$. Since $Q(d) \subset |c|\mathbb{B}$, for all $d \in \mathbb{R}^\omega$, where $\mathbb{B}$ is the unit ball of $\mathbb{R}^\pi$ and $|c| = \max\{c_i | i = 1, \ldots, \pi\}$ (c is the route capacity vector), then the set-valued mapping $(d, v) \rightrightarrows Q(d) \cap J^{-1}(v)$ is uniformly bounded around $(\overline{d}, \overline{v})$. Hence, including the continuous differentiability of the function J, it follows from Rockafellar and Wets [29, page 454] that there exists $\overline{q}$ with $\overline{v} = \Delta\overline{q}$ and $(\overline{d}, \overline{q}) \in \text{gph}Q$ such that $d^* \in D^*Q(\overline{d}, \overline{q})(\Delta^\top(-v^*))$, that is $(d^*, \Delta^\top v^*) \in N_{\text{gph}Q}(\overline{d}, \overline{q})$. □

In order to focus on the main ideas we consider the simplified situation where $D = \mathbb{R}^\omega$ and the upper and lower level cost functions F and f are all continuously differentiable. Nevertheless, all the results here can easily be extended to more general cases. Let's denote by

$$\Lambda(\overline{d},\overline{q}) = \Big\{(\lambda^\omega,\lambda^c,\lambda^o)|\lambda^\omega = (\lambda_w),\ \lambda^c = (\lambda^c_r) \geq 0,\ \lambda^o = (\lambda^o_r) \geq 0,$$
$$- \sum_{r\in\mathcal{P}^c(\overline{q})} \lambda^c_r e^r + \sum_{r\in\mathcal{P}^o(\overline{q})} \lambda^o_r e^r - \sum_{w\in\mathcal{W}} \lambda_w \Lambda_w^\top = \Delta^\top \nabla_v f(\overline{d},\overline{v})\Big\}$$

the set of Lagrange multipliers for the traffic assignment problem and by

$$\mathcal{H}(\overline{d},\overline{v}) = \{q \in \mathbb{R}^\pi | \Delta q = \overline{v}, (\overline{d}, q) \in \text{gph}Q\} \tag{86}$$

the set of route flows corresponding to the feasible demand-link flow couple $(\overline{d},\overline{v})$. Then, the subdifferential of the optimal value function of the traffic assignment problem can be estimated as follows.

Theorem 5.2. *For every $(\overline{d},\overline{v}) \in gph\Psi$, the optimal value function φ of the traffic assignment problem is locally Lipschitz continuous around $\overline{d}$ and*

$$\partial\varphi(\overline{d}) \subseteq \bigcup_{\overline{v}\in\Psi(\overline{d})} \bigcup_{\overline{q}\in\mathcal{H}(\overline{d},\overline{v})} \bigcup_{(\lambda^\omega,\lambda^c,\lambda^0)\in\Lambda(\overline{d},\overline{q})} \Big\{ - \sum_{w\in\mathcal{W}} \lambda_w e^w + \nabla_d f(\overline{d},\overline{v})\Big\}. \tag{87}$$

Moreover, if we assume that Ψ is inner semicontinuous at $(\overline{d},\overline{v})$, then

$$\partial\varphi(\overline{d}) \subseteq \bigcup_{\overline{q}\in\mathcal{H}(\overline{d},\overline{v})} \bigcup_{(\lambda^\omega,\lambda^c,\lambda^0)\in\Lambda(\overline{d},\overline{q})} \Big\{ - \sum_{w\in\mathcal{W}} \lambda_w e^w + \nabla_d f(\overline{d},\overline{v})\Big\}. \tag{88}$$

Proof. At first the set-valued mapping $V = JoQ$ satisfies Aubin property as the composition of the set-valued mapping Q satisfying Aubin property (the system $0 \leq q \leq c$, $\Lambda q = d$ is linear in q) and the locally Lipschitz continuity of J (see [29, Corollary 10.38]) given that Q is uniformly bounded, i.e. we have $Q(d) \subseteq |c|\mathbb{B}, \forall d \in \mathbb{R}^\omega$, with $\mathbb{B}$ being the unit ball of $\mathbb{R}^\pi$ and $|c| = \max\{c_i | i = 1,\ldots,\pi\}$. Since J is a continuous function, we then have $\Psi(d) \subseteq V(d) = J(Q(d)) \subseteq J(|c|\mathbb{B}), \forall d \in \mathbb{R}^\omega$, with $J(|c|\mathbb{B})$ bounded. This means that the set-valued mapping Ψ is uniformly bounded, hence inner semicompact, cf. Section 2. Combining the inner semicompactness of Ψ and the Aubin property for V, the value function φ is locally Lipschitz continuous following Mordukhovich and Nam [26, Theorem 5.2].

On the other hand, with the continuous differentiability of f, it follows from [27, Theorem 7] that

$$\partial\varphi(\overline{d}) \subseteq \bigcup_{\overline{v}\in\Psi(\overline{d})} \Big\{\nabla_d f(\overline{d},\overline{v}) + D^*V(\overline{d},\overline{v})(\nabla_v f(\overline{d},\overline{v}))\Big\} \tag{89}$$

and

$$\partial\varphi(\overline{d}) \subseteq \Big\{\nabla_d f(\overline{d},\overline{v}) + D^*V(\overline{d},\overline{v})(\nabla_v f(\overline{d},\overline{v}))\Big\}, \tag{90}$$

under the inner semicompactness of Ψ at $\overline{d}$ (fulfilled) and the inner semicontinuity of Ψ at $(\overline{d},\overline{v})$, respectively. Moreover, it follows from the definition of the coderivative that $d^* \in D^*V(\overline{d},\overline{v})(\nabla_v f(\overline{d},\overline{v}))$ if and only if $(d^*, -\nabla_v f(\overline{d},\overline{v})) \in N_{\text{gph}V}(\overline{d},\overline{v})$. Hence, from Lemma 5.1, there exists $\overline{q} \in \mathcal{H}(\overline{d},\overline{v})$ such that

$(d^*, -\Delta^\top \nabla_v f(\overline{d}, \overline{v})) \in N_{\text{gph}Q}(\overline{d}, \overline{q})$, which implies the existence of $(\lambda^\omega, \lambda^c, \lambda^o)$, with $\lambda^\omega = (\lambda_w)$, $\lambda^c = (\lambda^c_r) \geq 0$, $\lambda^o = (\lambda^o_r) \geq 0$ such that

$$d^* = -\sum_{w \in \mathcal{W}} \lambda_w e^w \tag{91}$$

and

$$\Delta^\top \nabla_v f(\overline{d}, \overline{v}) = -\sum_{r \in \mathcal{P}^c(\overline{q})} \lambda^c_r e^r + \sum_{r \in \mathcal{P}^o(\overline{q})} \lambda^o_r e^r - \sum_{w \in \mathcal{W}} \lambda_w \Lambda^\top_w. \tag{92}$$

Hence, the inclusion in (87) follows by the combination of (89) and (91)-(92). Inclusion (88) follows analogously by exploiting (90). □

We are now able to give KKT optimality conditions for the DAP under the partial calmness defined in section 4.

Theorem 5.3. *Let $(\overline{d}, \overline{v})$ be a local optimal solution to problem (80), assumed to be partially calm at $(\overline{d}, \overline{v})$. Then there exist $\overline{q} \in \mathbb{R}^\pi$, $\mu > 0$, $(\lambda^\omega, \lambda^c, \lambda^o)$, and $v_s \in \Psi(\overline{d})$, $q_s \in \mathbb{R}^\pi$, $(\lambda^\omega_s, \lambda^c_s, \lambda^o_s)$ and $\eta_s \geq 0$, $s = 1, \ldots, \omega+1$, with $\sum_{s=1}^{\omega+1} \eta_s = 1$ such that*

$$\nabla_d F(\overline{d}, \overline{v}) + \mu \nabla_d f(\overline{d}, \overline{v}) - \sum_{w \in \mathcal{W}} \lambda_w e^w + \mu \sum_{s=1}^{\omega+1} \eta_s \Big(\sum_{w \in W} \lambda_{sw} e^w - \nabla_d f(\overline{d}, v_s) \Big) = 0$$

$$\Delta^\top \Big(\nabla_v F(\overline{d}, \overline{v}) + \mu \nabla_v f(\overline{d}, \overline{v}) \Big) = \sum_{r \in \mathcal{P}^o(\overline{q})} \lambda^o_r e^r - \sum_{r \in \mathcal{P}^c(\overline{q})} \lambda^c_r e^r - \sum_{w \in \mathcal{W}} \lambda_w \Lambda^\top_w$$

$$\Delta^\top \nabla_v f(\overline{d}, v_s) = -\sum_{r \in \mathcal{P}^c(q_s)} \lambda^c_{sr} e^r + \sum_{r \in \mathcal{P}^o(q_s)} \lambda^o_{sr} e^r - \sum_{w \in \mathcal{W}} \lambda_{sw} \Lambda^\top_w$$

$$0 \leq \overline{q} \leq c,\ \Lambda \overline{q} = \overline{d},\ \Delta \overline{q} = \overline{v}$$

$$0 \leq q_s \leq c,\ \Lambda q_s = \overline{d},\ \Delta q_s = v_s$$

$$\lambda^\omega = (\lambda_w),\ \lambda^c = (\lambda^c_r) \geq 0,\ \lambda^o = (\lambda^o_r) \geq 0$$

$$\lambda^\omega_s = (\lambda_{sw}),\ \lambda^c_s = (\lambda^c_{sr}) \geq 0,\ \lambda^o_s = (\lambda^o_{sr}) \geq 0.$$

Proof. Under the partial calmness, it follows from Theorem 4.1 that there exists $\mu > 0$ such that $(\overline{d}, \overline{v})$ solves

$$\begin{aligned} &\text{minimize } F(d, v) + \mu(f(d, v) - \varphi(d)) \\ &\text{subject to } (d, v) \in \text{gph}V. \end{aligned}$$

Since $\text{gph}V$ is closed and φ is locally Lipschitz continuous around $\overline{d}$, it follows from Mordukhovich [25, Proposition 5.3] that

$$0 \in \partial(F + \mu(f - \varphi))(\overline{d}, \overline{v}) + N_{\text{gph}V}(\overline{d}, \overline{v}). \tag{93}$$

Considering the sum rule (2) and the convex hull property (1) we have

$$0 \in \nabla F(\overline{d}, \overline{v}) + \mu \nabla f(\overline{d}, \overline{v}) - \mu \text{co}\varphi(\overline{d}) \times \{0\} + N_{\text{gph}V}(\overline{d}, \overline{v}).$$

Hence it follows from Lemma 5.1 that there exist $v^* \in \mathbb{R}^\alpha$, $\overline{q} \in \mathbb{R}^\pi$, $\mu > 0$, and

$(\lambda^\omega, \lambda^c, \lambda^o)$, with $\lambda^\omega = (\lambda_w)$, $\lambda^c = (\lambda_r^c) \geq 0$, $\lambda^o = (\lambda_r^o) \geq 0$ such that

$$-\sum_{w\in\mathcal{W}} \lambda_w e^w + \nabla_d F(\overline{d}, \overline{v}) + \mu \nabla_d f(\overline{d}, \overline{v}) \in \mu \mathrm{co}\partial\varphi(\overline{d}) \tag{94}$$

$$\Delta^\top v^* = \sum_{w\in\mathcal{W}} \lambda_w \Lambda_w^\top + \sum_{r\in\mathcal{P}^c(\overline{q})} \lambda_r^c e^r - \sum_{r\in\mathcal{P}^o(\overline{q})} \lambda_r^o e^r \tag{95}$$

$$v^* = -\nabla_v F(\overline{d}, \overline{v}) - \mu \nabla_v f(\overline{d}, \overline{v}) \tag{96}$$

$$0 \leq \overline{q} \leq c,\ \Lambda\overline{q} = \overline{d},\ \Delta\overline{q} = \overline{v}. \tag{97}$$

On the other hand, if we take $d^* \in \mathrm{co}\varphi(\overline{d})$, then it follows from Caratheodory's Theorem (see e.g. Mangasarian [20]) that there exist $\eta_s \in \mathbb{R}$ and $d_s^* \in \partial\varphi(\overline{d})$, with $s = 1, \ldots, \omega + 1$ such that $d^* = \sum_{s=1}^{\omega+1} \eta_s d_s^*$, $\sum_{s=1}^{\omega+1} \eta_s = 1$, $\eta_s \geq 0$. Hence, the result follows from (94)-(97) and the inclusion (87). □

Additionally, if we consider the inner semicontinuity of the solution set-valued mapping Ψ, the following result can be stated and we omit the proof since it is analogous to the previous one, given that only the estimation of the subdifferential of the optimal value function differs.

Theorem 5.4. *Let $(\overline{d}, \overline{v})$ be an optimal solution to problem (80), assumed to be partially calm at $(\overline{d}, \overline{v})$. We also assume that Ψ is inner semicontinuous $(\overline{d}, \overline{v})$. Then there exist $\overline{q} \in \mathbb{R}^\pi$, $\mu > 0$, $(\lambda^\omega, \lambda^c, \lambda^o)$, and $q_s \in \mathbb{R}^\pi$, $(\lambda_s^\omega, \lambda_s^c, \lambda_s^o)$ and $\eta_s \geq 0$, $s = 1, \ldots, \omega + 1$, with $\sum_{s=1}^{\omega+1} \eta_s = 1$ such that*

$$\begin{aligned}
\nabla_d F(\overline{d}, \overline{v}) - \sum_{w\in\mathcal{W}} \lambda_w e^w + \mu \sum_{s=1}^{\omega+1} \eta_s \sum_{w\in W} \lambda_{sw} e^w &= 0 \\
\Delta^\top \left(\nabla_v F(\overline{d}, \overline{v}) + \mu \nabla_v f(\overline{d}, \overline{v})\right) &= \sum_{r\in\mathcal{P}^o(\overline{q})} \lambda_r^o e^r - \sum_{r\in\mathcal{P}^c(\overline{q})} \lambda_r^c e^r - \sum_{w\in\mathcal{W}} \lambda_w \Lambda_w^\top \\
\Delta^\top \nabla_v f(\overline{d}, \overline{v}) &= -\sum_{r\in\mathcal{P}^c(q_s)} \lambda_{sr}^c e^r + \sum_{r\in\mathcal{P}^o(q_s)} \lambda_{sr}^o e^r - \sum_{w\in\mathcal{W}} \lambda_{sw} \Lambda_w^\top \\
0 \leq \overline{q} \leq c,\ \Lambda\overline{q} &= \overline{d},\ \Delta\overline{q} = \overline{v} \\
0 \leq q_s \leq c,\ \Lambda q_s &= \overline{d},\ \Delta q_s = \overline{v} \\
\lambda^\omega = (\lambda_w),\ \lambda^c = (\lambda_r^c) &\geq 0,\ \lambda^o = (\lambda_r^o) \geq 0 \\
\lambda_s^\omega = (\lambda_{sw}),\ \lambda_s^c = (\lambda_{sr}^c) &\geq 0,\ \lambda_s^o = (\lambda_{sr}^o) \geq 0.
\end{aligned}$$

A generalization of Theorem 5.3 and Theorem 5.4 to the case where the set D is different from $\mathbb{R}^\omega$ is possible if we additionally require $(\overline{d}, \overline{v})$ to satisfy

$$N_{\mathrm{gph}V}(\overline{d}, \overline{v}) \cap N_{D\times\mathbb{R}^\alpha}(\overline{d}, \overline{v}) = \{0\}$$

in order to add the normal cone $N_{D\times\mathbb{R}^\alpha}(\overline{d}, \overline{v})$ to the right hand side of the condition in (93). For the inner semicontinuity of Ψ, it is automatically achieved if we assume that the solution of the traffic assignment problem is locally unique and determines a continuous function. This is usually the case for most of the bilevel transportation problems considered in the literature [4, 32, 33]. It is obtained by assuming that for every link $a \in \mathcal{A}$, the link cost t_a is strictly increasing in v_a. Finally, as far as the partial calmness of the DAP is concerned, it follows from Theorem 4.2 that this will be satisfied if the traffic cost of the road users expressed by f is linear in the link flow.

Acknowledgements: The work of the second author was partly supported by the Deutscher Akademischer Austausch Dienst (DAAD).

References

[1] T. Abrahamsson, Estimation of origin-destination matrices using traffic counts A literature survey, Interim Report, International Institute for Applied Systems Analysis, Laxenburg, Austria, IR-98-021/May, 1998.

[2] M. Beckmann, C.B. Mcguire and C.B. Winsten, Studies in the economics of transportation, Yale University Press, New Haven, CT, 1956.

[3] D. Bertsimas and J.N. Tsitsiklis, Introduction to Linear optimization, Athena Scientific, Belmont, Massachusetts, 1997.

[4] Y. Chen, Bilevel programming problems: Analysis, algorithms and applications, PhD Thesis, Centre de Recherche sur les Transports (C.R.T.), Université de Montréal, CRT-984, 1994.

[5] Y. Chen and M. Florian, Congested O-D trip demand adjustment problem: Bilevel programming formulation and optimality condtions, A. Migdalas et al. (eds), Multilevel optimization: Algorithms and aplications, 1-22, Kluwer Academic Publishers, Netherlands, 1998.

[6] S. Dempe, Foundations of bilevel programming, Kluwer Academic Publishers, Dordrecht, 2002.

[7] S. Dempe and J. Dutta, Is bilevel programming a special case of mathematical programming with equilibrium constraints?, To appear in Mathematical Programming.

[8] S. Dempe, J. Dutta and S. Lohse, Optimality conditions for bilevel programming problems, Optimization, Vol. 55, Nos. 5-6, 505-524, 2006.

[9] S. Dempe, J. Dutta and B. S. Mordukhovich, New necessary optimal conditions in optimistic bilevel programming, Optimization, Vol. 56, 2007.

[10] S. Dempe, J. Dutta and B.S. Mordukhovich, Variational analysis in bilevel programming, Proceedings of the International Symposium on Mathematical Programming for Decision Making: Indian Statistical Institute, Delhi, December 10-11, 2007, Word Scientific, Singapore 2008.

[11] S. Dempe, V.V. Kalashnikov and N. Kalashnikova, Optimality conditions for bilevel programming problems, S. Dempe and V. Kalashnikov (eds.), Opimization with Multivalued Mappings, 3-28, Springer Science+Business Media, LLC, 2006.

[12] S. Dempe and A.B. Zemkoho, A bilevel approach for traffic management in capacitated networks, Preprint 2008-05, Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg, 2008.

[13] S. Dempe and A.B. Zemkoho, The generalized Mangasarian-Fromowitz constraint qualification and optimality conditions for bilevel programs, Submitted.

[14] S. Dempe and A.B. Zemkoho, Bilevel road pricing: Some theoretical analysis and optimality conditions, In preparation.

[15] J. Dutta and S. Dempe, Bilevel programming with convex lower level problems, S. Dempe and V. Kalashnikov (eds.), Optimization with Multivalued Mappings, 51-71, Springer Science+Business Media, LLC, 2006.

[16] C.S. Fisk, On combining maximum entropy trip matrix estimation with useroptimal assignment, Transportation Research 22B, 1988.

[17] M. Flegel, Constraint qualification and stationarity concepts for mathematical programs with equilibrium constraints, PhD Thesis, Institute of Applied Mathematics and Statistics, University of Würzburg, 2005.

[18] R. Henrion, A. Jourani and J. Outrata, On the Calmness of a Class of Multifunctions, SIAM Journal of Optimization, Vol. 13, No. 2, 603-618, 2002.

[19] R. Henrion and J.V. Outrata, A subdifferential condition for calmness of Multifunctions, Journal of Mathematical Analysis and Applications, Vol. 258, 110-130, 2001.

[20] O.L. Mangasarian, Nonlinear programming, SIAM, 1994.

[21] O.L. Mangasarian and T.-H. Shiau, Lipschitz continuity of solutions of linear inequalities, programs and complementary problems, SIAM Journal of Control and Optimization, Vol. 25, No. 3, 1987.

[22] A. Migdalas, Bilevel programming in traffic planning: Models, methods and challenge, Journal of Global Optimization, Vol. 7, 381-405, 1995.

[23] J.A. Mirrlees, The theory of Moral hazard and unobservable behaviour: Part I, Review of Economic Studies, Vol. 66, 3-21, 1999.

[24] B.S. Mordukhovich, Variational analysis and generalized differentiation I: Basic Theory, Grundlehren der mathematischen Wissenschaften, Springer, 2006.

[25] B.S. Mordukhovich, Variational analysis and generalized differentiation II: Applications, Grundlehren der mathematischen Wissenschaften, Springer, 2006.

[26] B.S. Mordukhovich and N.M. Nam, Variational stability and marginal functions via generalized differentiation, Mathematics of Operations Research, Vol. 30, No. 4, 800-816, 2005.

[27] B.S. Mordukhovich, N.M. Nam, N.D. Yen, Subgradients of marginal functions in parametric mathematical programming, Mathematical Programming, Ser. B, 2006.

[28] R.T. Rockafellar, Lagrange multipliers and optimality, SIAM Review, Vol. 35, No. 2, 1993.

[29] R.T. Rockafellar and R.J.-B. Wets, Variational analysis, Grundlehren der mathematischen Wissenschaften, Springer, 1998.

[30] H. Scheel and Scholtes, Mathematical programs with equilibrium constraints: stationarity, optimality, and sensitivity Mathematics of Operations Research, Vol. 25, 1-22, 2000.

[31] J.G. Wardrop, Some theoretical aspects of road traffic research, Proceedings of the Institute of Civil Engeneers, Part II, 325-378, 1952.

[32] H. Yang, Heuristic algorithms for the bilevel origin-destination matrix estimation problem, Trasportation Ressearch B, Vol. 29B, No. 4, 231-242, 1995.

[33] H. Yang, T. Sasaki, Y. Iida and Y. Asakura, Estimation of origin-destination matrices from link traffic counts on congested networks, Transportation Research B, Vol. 26B, No. 6, 417-434, 1992.

[34] J.J. Ye, Constraint qualifications and KKT conditions for bilevel programming problems, Mathematics of Operational Research, Vol. 31, No. 4, 2006.

[35] J.J. Ye, New uniform parametric error bounds, Journal of Optimization Theory and Applications, Vol. 98, No. 1, 197-219, 1998.

[36] J.J. Ye, Nondifferentiable multiplier rules for optimization and bilevel optimization problems, SIAM Journal of Optimization, Vol. 15, No. 1, 252-274, 2004.

[37] J.J. Ye and D.L. Zhu, Optimality conditions for bilevel programming problems, Optimization, Vol. 33, No. 1, 9-27, 1995.

[38] J.J. Ye and D.L. Zhu, A note on optimality conditions for bilevel programming problems, Optimization, Vol. 39, No. 4, 361-366, 1997.

[39] J.J. Ye and D.L. Zhu, New necessary optimality conditions for bilevel programs by combined MPEC and the value function approach, http://www.math.uvic.ca/faculty/janeye/papers/YeZhuNov1.pdf (to appear on SIAM Journal on Optimization).

[40] J.J. Ye, D.L. Zhu and Q.J. Zhu, Exact penalization and necessary optimality conditions for generalized bilevel programming problems, SIAM Journal of Optimization, Vol. 7, No. 2, 481-507, 1997.