Math. Program., Ser. A (2013) 138:447–473
DOI 10.1007/s10107-011-0508-5

FULL LENGTH PAPER

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

S. Dempe · A. B. Zemkoho

Received: 4 March 2010 / Accepted: 10 December 2011 / Published online: 5 January 2012

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 calculus of Mordukhovich.

Keywords Bilevel programming · Optimal value function · Constraint qualifications · Optimality conditions · Demand adjustment problem

Mathematics Subject Classification (2000) 90C30 · 91A65 · 90B06

1 Introduction

In this paper we consider the (optimistic) bilevel programming problem also called the *leader's/lower level* problem, which is a special optimization problem partly constrained by a second (parametric) optimization problem known as the *follower's/lower level* problem. In order to write this problem as a one level mathematical programming

S. Dempe · A. B. Zemkoho (✉)
Institut für Numerische Mathematik und Optimierung,
Technische Universität Bergakademie Freiberg, Akademie Straße 6,
09596 Freiberg, Germany
e-mail: zemkoho@student.tu-freiberg.de

S. Dempe
e-mail: dempe@math.tu-freiberg.de

problem, two major reformulations have been suggested in the literature [5], i.e. the KKT reformulation and the optimal value reformulation. The KKT reformulation usually consists 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 reformulation that would solve the original bilevel programming problem, hence the two problems are not locally equivalent [6]. Although, as noticed in Sect. 3, the equivalence with the (primal) KKT reformulation remains complete, if the normal cone to the feasible set of the lower level problem is not explicitly computed. As for the optimal value reformulation, it is obtained (without any assumption) by replacing the lower level solution set by its description via the optimal value function. This reformulation, introduced in [26], is also completely equivalent to the initial problem. It should however be mentioned that the optimal value function is typically nonsmooth, even when the functions involved are affine linear ones.

Most of the well-known CQs, irrespective of the reformulation, are known to fail for the bilevel optimization problem; see e.g. [11,15,40]. In this paper, we derive necessary optimality conditions for the aforementioned reformulations, under some weaker CQs; and also analyze possible links between them. Considering the partial calmness CQ, an important result is also 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 [38], 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 [14] 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 mappings, we use the sophisticated coderivative tool of Mordukhovich, from the field of variational analysis, to obtain KKT type optimality conditions for the DAP. Hence, greatly improving the works of Chen [3] and Chen and Florian [4], where Fritz-John's type optimality conditions were derived after considering some important simplifications.

In the next section, we first present 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 Sect. 3, we introduce the optimistic bilevel programming problem and the optimal value and KKT reformulations are considered. Their relationships with the initial problem are studied and necessary optimality conditions derived. In Sect. 4, we consider the partial calmness, which is the best known CQ for the optimal value reformulation of 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 Sect. 5, the DAP is introduced and may be for the first time, KKT type necessary optimality conditions are obtained 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 [22] and Rockafellar and Wets [30]. We first consider some initial notations: Let C be a subset of $\mathbb{R}^n$, $\operatorname{co} C$ and $\operatorname{bd} C$ denote the convex hull and the boundary of C, respectively. For a matrix M, $M^\top$ is the transposed matrix of M. For $a \in \mathbb{R}^n$, $a \leq 0$ should be understood componentwise. Finally, $\|\cdot\|$ denotes an arbitrary norm in $\mathbb{R}^n$ and $\langle \cdot, \cdot \rangle$ is used for the inner product of $\mathbb{R}^n$.

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

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

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

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

$$T_C(\overline{x}) := \{x \in \mathbb{R}^n |\, \exists t_k \downarrow 0,\ x_k \to x : \overline{x} + t_k x_k \in C\}$$

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

$$\widehat{N}_C(\overline{x}) := \{x^* \in \mathbb{R}^n |\, \langle x^*, x\rangle \leq 0,\ \forall x \in T_C(\overline{x})\}$$

while the basic normal cone introduced by Mordukhovich is defined as

$$\begin{aligned} N_C(\overline{x}) &:= \{x^* \in \mathbb{R}^n |\, \exists x_k^* \to x^*,\ x_k \to \overline{x}\, (x_k \in C) : x_k^* \in \widehat{N}_C(x_k)\} \\ &:= \{x^* \in \mathbb{R}^n |\, \exists x_k^* \to x^*,\ x_k \to \overline{x} : x_k^* \in \operatorname{cone}(x_k - \Pi(x_k, C))\}, \end{aligned}$$

where "cone" stands for the conic hull of the set in question and Π for the Euclidean projection on C. The second equality (more appropriate for our finite-dimensional framework) corresponds to the initial version of the basic normal cone as introduced in [21]. 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 C. Nevertheless, $\widehat{N}_C(\overline{x})$ and $N_C(\overline{x})$ both coincide with the normal cone of convex analysis when C is convex. Additionally, the Mordukhovich normal cone is contained in the Clarke normal cone for an arbitrary closed set C, and it induces some major tools of Variational Analysis, i.e. the Mordukhovich subdifferential and coderivative.

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

$$\partial\chi(\overline{x}) := \{x^* \in \mathbb{R}^n | (x^*, -1) \in N_{\text{epi}\chi}(\overline{x}, \chi(\overline{x}))\},$$

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

$$\partial\chi(\overline{x}) = \left\{\nabla\chi(\overline{x})\right\}$$

provided χ is continuously differentiable. The following convex hull property

$$\text{co}\,\partial(-\chi)(\overline{x}) = -\text{co}\,\partial\chi(\overline{x}), \tag{2.1}$$

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

$$\partial(\lambda\vartheta + \mu\chi)(\overline{x}) \subseteq \lambda\partial\vartheta(\overline{x}) + \mu\partial\chi(\overline{x}), \tag{2.2}$$

where equality holds if ϑ or χ 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 dom Φ is the set of all $x \in \mathbb{R}^n$ such that $\Phi(x) \neq \emptyset$ and its graph is given as

$$\text{gph}\,\Phi := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m | \, y \in \Phi(x)\}.$$

The coderivative of Φ at $(\overline{x}, \overline{y}) \in \text{gph}\,\Phi$ 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

$$D^*\Phi(\overline{x}, \overline{y})(y^*) := \left\{x^* \in \mathbb{R}^n | \, (x^*, -y^*) \in N_{\text{gph}\Phi}(\overline{x}, \overline{y})\right\}. \tag{2.3}$$

Φ will be said to be 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 $C \subset \mathbb{R}^m$ such that $\Phi(x) \subseteq C, \ \forall x \in U$. The mapping Φ is inner semicontinuous at $(\overline{x}, \overline{y}) \in \text{gph}\,\Phi$ 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 (or Lipschitz-like) property around the point $(\overline{x}, \overline{y}) \in \text{gph}\,\Phi$ 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; see [22].

3 Optimal value versus KKT reformulation

We are mainly concerned in this section with analyzing the optimal value and KKT reformulations of the bilevel programming problem in terms of CQs, stationary points, local and global optimal solutions. A special attention will be given to similarities and possible relationships between the two reformulations. We first start by presenting these reformulations and the links between their optimal solutions and the initial problem.

3.1 Reformulations and optimal solutions

We consider the optimistic bilevel programming problem to

$$\text{minimize } F(x, y) \text{ subject to } x \in X \subseteq \mathbb{R}^n, \; y \in \Psi(x), \tag{3.1}$$

also called the *leader's/upper level* problem, where X is a closed set, $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ a continuously differentiable function and $\Psi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, a set-valued mapping describing the solution set of the parametric optimization problem to

$$\text{minimize } f(x, y) \text{ subject to } y \in K(x), \tag{3.2}$$

$K(x)$ 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 as well. The difference between what is called the optimal value reformulation of problem (3.1) and its so-called KKT reformulation resides in the way the solution set $\Psi(x)$ of the *follower's/lower level* problem (3.2) is expressed in order to have a one level optimization problem.

For the optimal value reformulation, the lower level solution set is chosen as:

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

where φ denotes the optimal value function of problem (3.2), defined as

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

For φ to be well-defined, we assume throughout the paper that the lower level problem (3.2) admits an optimal solution for each $x \in X$. We can then consider the following reformulation of problem (3.1):

$$\text{minimize } F(x, y) \text{ subject to } (x, y) \in \Omega, \ f(x, y) \leq \varphi(x) \tag{3.3}$$

with $\Omega := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m | \, x \in X, \ y \in K(x)\}$. Problem (3.3) is well-known as the optimal value reformulation of the bilevel programming problem; such a technique to rewrite a two level optimization problem is due to Outrata, cf. [26], who also suggested an algorithmic approach for (3.3), see [27]. On the other hand, the theory on CQs and optimality conditions for problem (3.3) may be traced back to the work of Ye and Zhu [40]. For a recent review and new results on the latter issue, the interested reader is referred to [11]. As far as optimal solutions of the optimal value reformulation are concerned, we have the following trivial link with those of the initial problem.

Theorem 3.1 *A point $(\overline{x}, \overline{y})$ is a local (resp. global) optimal solution of* (3.1) *if and only if it is a local (resp. global) optimal solution of* (3.3).

Meanwhile, if we assume the parametric problem (3.2) to be convex, i.e., the function $f(x, .)$ and the set $K(x)$ are convex for all $x \in X$, then the lower level solution set takes the form

$$\Psi(x) = \{y \in \mathbb{R}^m | \, 0 \in \nabla_y f(x, y) + N_{K(x)}(y)\} \tag{3.4}$$

where $N_{K(x)}(y)$ denotes the normal cone (in the sense of convex analysis) to $K(x)$ at y, provided $y \in K(x)$, whereas $N_{K(x)}(y) := \emptyset$, if $y \notin K(x)$. Hence, the bilevel programming problem (3.1) can be reformulated as

$$\text{minimize } F(x, y) \text{ subject to } (x, y) \in X \times \mathbb{R}^m, \ 0 \in \nabla_y f(x, y) + N_{K(x)}(y). \tag{3.5}$$

The link between the two problems can be stated as follows:

Theorem 3.2 *The lower level problem is assumed to be convex. Then, a point $(\overline{x}, \overline{y})$ is a local (resp. global) optimal solution of* (3.1) *if and only if $(\overline{x}, \overline{y})$ is a local (resp. global) optimal solution of* (3.5).

Problem (3.5), which is an optimization problem with generalized equation constraint has been studied for example in [22,25,39], usually under the name of optimization problem with variational inequality constraint. The reason for the latter appellation is that under the convexity of $K(x)$ for $x \in X$, one has the following well-known expression for the normal cone in the sense of convex analysis

$$N_{K(x)}(y) := \{v \in \mathbb{R}^m | \langle v, u - y\rangle \leq 0, \ \forall u \in K(x)\}.$$

If we insert this formula in (3.5), the aforementioned vocabulary is justified. Given that the generalized equation defining the lower level solution set in (3.4) may also be seen as a compact form of the Karush-Kuhn-Tucker (KKT) conditions of the lower level problem, problem (3.5) could consequently be seen as a *primal* KKT reformulation of the bilevel programming problem. Hence, problem (3.5) and its sparse form below, see (3.6), may both be seen as KKT reformulations of the bilevel optimization problem.

Concretely, let us consider a rather simplified bilevel program (3.1) 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 $G : \mathbb{R}^n \to \mathbb{R}^k$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ 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 MFCQ is satisfied at all points $y \in \Psi(x),\ x \in X$, then we have the following equality [30]:

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

Hence the following so-called *KKT reformulation*

$$\begin{aligned} &\text{minimize } F(x, y) \\ &\text{subject to } \begin{cases} G(x) \leq 0,\ \ g(x, y) \leq 0 \\ u \geq 0,\ \ u^\top g(x, y) = 0 \\ \nabla_y f(x, y) + \nabla_y g(x, y)^\top u = 0, \end{cases} \end{aligned} \tag{3.6}$$

given that the above formula for the normal cone induces the Karush-Kuhn-Tucker type optimality conditions for the lower level optimization problem (3.2). Clearly, problem (3.6) is also a mathematical programming problem with complementarity constraints (MPCC), considering the embedded complementarity problem:

$$g(x, y) \leq 0,\ \ u \geq 0,\ \ u^\top g(x, y) = 0. \tag{3.7}$$

We now have the following theorem, resulting from a recent study by Dempe and Dutta [6], on the relationship between the bilevel program and its KKT reformulation after the normal cone present in (3.5) is explicitly computed. Before, we recall that Slater's CQ is said to be satisfied for the lower level problem at $x \in X$, provided there exists $\overline{y}(x)$ such that $g(x, \overline{y}(x)) < 0$. Also, for a given point $(\overline{x}, \overline{y})$ of the graph of Ψ, the set $\Lambda(\overline{x}, \overline{y})$ denotes all the vectors u satisfying: $u \geq 0, u^\top g(\overline{x}, \overline{y}) = 0$ and $\nabla_y f(\overline{x}, \overline{y}) + \nabla_y g(\overline{x}, \overline{y})^\top u = 0$.

Theorem 3.3 *The lower level problem is assumed to be a convex one. Let $(\overline{x}, \overline{y})$ be a global (resp. local) optimal solution of* (3.1) *and assume that Slater's CQ is satisfied at $\overline{x}$. Then, for each $\overline{u} \in \Lambda(\overline{x}, \overline{y})$, the point $(\overline{x}, \overline{y}, \overline{u})$ is a global (resp. local) optimal solution of* (3.6)*. Conversely, let Slater's CQ be satisfied at all $x \in X$ (resp. at $\overline{x}$). Assume that $(\overline{x}, \overline{y}, \overline{u})$ is a global optimal solution (resp. local optimal solution for all $\overline{u} \in \Lambda(\overline{x}, \overline{y})$) of* (3.6)*. Then, $(\overline{x}, \overline{y})$ is a global (resp. local) optimal solution of* (3.1).

Broadly, it is clear from this result that for global solutions, problem (3.6) is equivalent to the initial problem (3.1). But, for local solutions, the requirement that $(\overline{x}, \overline{y}, \overline{u})$ be a local optimal solution of (3.6) for all $\overline{u} \in \Lambda(\overline{x}, \overline{y})$, before one can be sure that $(\overline{x}, \overline{y})$ is a local optimal solution for (3.1), is too strong. Not only it is not realistic at a computational level, but one can easily construct examples of problems where a local optimal solution of problem (3.6) is not a local optimal solution of (3.1), cf. [6]. Hence, it may be fair to say that computing the normal cone in (3.5) destroys the nice link, stated in Theorem 3.2, between the bilevel program (3.1) and its primal KKT reformulation (3.5).

It is important to mention that the CQ (Slater) can not be dropped in any segment of the above theorem [6]. This is the case for the convexity of the lower level problem as well; otherwise a solution of the bilevel program need not even be a stationary point of problem (3.6), see Mirrlees [20]. To circumvent this convexity assumption and some possibly unwanted behavior that may occur using the optimal value reformulation, Ye and Zhu [37] 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 in [37], that the KKT conditions of the lower level problem be satisfied without necessarily requiring the convexity of the lower level problem and hence the following one level optimization problem is considered:

$$
\begin{aligned}
&\text{minimize } F(x, y)\\
&\text{subject to } \begin{cases} f(x, y) - \varphi(x) \leq 0 \\ G(x) \leq 0, \;\; g(x, y) \leq 0 \\ u \geq 0, \;\; u^\top g(x, y) = 0 \\ \nabla_y f(x, y) + \nabla_y g(x, y)^\top u = 0. \end{cases}
\end{aligned}
$$

An equivalence between this problem and the initial one (3.1) was established in [37] around a predefined neighborhood of the considered point. For optimality conditions, techniques known for MPCCs were applied under calmness and partial calmness concepts tailored for the problem. We will not insist on this approach, since we are interested exclusively in considering the KKT and optimal value reformulations separately, and looking at possible links between them. The interested reader is referred to the aforementioned paper for further details.

3.2 CQs and stationary points

The well-known basic CQ, see e.g. [22,30], also called generalized MFCQ, reduces in the case of the optimal value reformulation (3.3), to the following condition:

$$\partial(f - \varphi)(\bar{x}, \bar{y}) \cap (-N_\Omega(\bar{x}, \bar{y})) = \emptyset. \tag{3.8}$$

It was shown in [11, Theorem 3.1] that this condition is violated at any feasible point of (3.3) under the fairly mild assumption that the sum rule

$$\partial((f - \varphi) + \delta_\Omega)(\bar{x}, \bar{y}) \subseteq \partial(f - \varphi)(\bar{x}, \bar{y}) + \partial\delta_\Omega(\bar{x}, \bar{y})$$

be satisfied. By passing from the normal cone N_Ω to its (topological) boundary, one has the weaker CQ:

$$\partial(f - \varphi)(\bar{x}, \bar{y}) \cap (-\text{bd}\, N_\Omega(\bar{x}, \bar{y})) = \emptyset. \tag{3.9}$$

Under this CQ, necessary optimality conditions were derived in [11] provided that the lower level solution set-valued mapping Ψ is inner semicontinuous or inner semicompact. In the next result, we consider the case where the optimal value function φ is convex; and necessary optimality conditions are obtained under CQ (3.9), without the need of the aforementioned requirements on Ψ.

Theorem 3.4 *Let $(\overline{x}, \overline{y})$ be a local optimal solution to problem (3.3) and assume that CQ (3.9) is satisfied at $(\overline{x}, \overline{y})$. We also let φ be convex and finite around $\overline{x}$. Furthermore, assume that the functions G and g are convex and there exists $(\widetilde{x}, \widetilde{y})$ such that $G(\widetilde{x}) < 0$ and $g(\widetilde{x}, \widetilde{y}) < 0$. Then, there exists $(\lambda, \alpha, \beta, \gamma) \in \mathbb{R}_+ \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^k$ such that:*

$$\nabla_x F(\overline{x}, \overline{y}) + \nabla_x g(\overline{x}, \overline{y})^\top (\alpha - \lambda\beta) + \nabla G(\overline{x})^\top \gamma = 0 \tag{3.10}$$

$$\nabla_y F(\overline{x}, \overline{y}) + \nabla_y g(\overline{x}, \overline{y})^\top (\alpha - \lambda\beta) = 0 \tag{3.11}$$

$$\nabla_y f(\overline{x}, \overline{y}) + \nabla_y g(\overline{x}, \overline{y})^\top \beta = 0 \tag{3.12}$$

$$\alpha \geq 0, \ \ \alpha^\top g(\overline{x}, \overline{y}) = 0 \tag{3.13}$$

$$\beta \geq 0, \ \ \beta^\top g(\overline{x}, \overline{y}) = 0 \tag{3.14}$$

$$\gamma \geq 0, \ \ \gamma^\top G(\overline{x}) = 0. \tag{3.15}$$

Proof The convexity of the functions G and g imply the convexity of the set Ω. Being convex and finite around $\overline{x}$, the value function φ is locally Lipschitz continuity around $\overline{x}$ [29]. Hence, it follows from [11, Lemma 3.3], [22, Proposition 5.3] and the sum rule (2.2) that there exists $\lambda \geq 0$ such that

$$0 \in \nabla F(\overline{x}, \overline{y}) + \lambda \nabla f(\overline{x}, \overline{y}) + \lambda \partial(-\varphi)(\overline{x}) \times \{0\} + N_\Omega(\overline{x}, \overline{y}). \tag{3.16}$$

We now recall that the convexity and finiteness of φ around $\overline{x}$ also imply $\partial\varphi(\overline{x}) \neq \emptyset$, cf. Rockafellar [29]. Hence, 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 Sect. 1), then

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

Considering the definition of φ and the fact that $\overline{y} \in \Psi(\overline{x})$ (since $(\overline{x}, \overline{y})$ is feasible to (3.3)) 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 exists β such that (3.12), (3.14) and

$$x^* = \nabla_x f(\overline{x}, \overline{y}) + \nabla_x g(\overline{x}, \overline{y})^\top \beta \tag{3.17}$$

hold, given that there exists $(\widetilde{x}, \widetilde{y})$ with $g(\widetilde{x}, \widetilde{y}) < 0$. Moreover, since $\widetilde{x}$ also satisfies $G(\widetilde{x}) < 0$, it follows from Rockafellar and Wets [30, Theorem 6.14] that

$$N_\Omega(\overline{x}, \overline{y}) = \{\nabla g(\overline{x}, \overline{y})^\top \alpha + (\nabla G(\overline{x})^\top \gamma, 0) | \alpha \geq 0, \alpha^\top g(\overline{x}, \overline{y}) = 0, \\ \gamma \geq 0, \gamma^\top G(\overline{x}) = 0\}. \tag{3.18}$$

Combining (3.16)–(3.18), there exists $(\lambda, \alpha, \beta, \gamma)$ satisfying (3.10), (3.12)–(3.15) and

$$\nabla_y F(\overline{x}, \overline{y}) + \lambda \nabla_y f(\overline{x}, \overline{y}) + \nabla_y g(\overline{x}, \overline{y})^\top \alpha = 0 \tag{3.19}$$

given that $\partial(-\varphi)(\overline{x}) \subseteq -\partial\varphi(\overline{x})$, following (2.1) and considering the convexity of φ. By inserting the expression of $\nabla_y f(\overline{x}, \overline{y})$ from (3.12) in Eq. (3.19), we recover Eq. (3.11) and hence, the result. □

The assumption that φ be finite around $\overline{x}$ seems superfluous given that it is required in the previous subsection that the lower level problem admits an optimal solution for each $x \in X$. But obviously, φ may happen not to be finite around $\overline{x}$ if $\overline{x} \in \text{bd}\, X$ (i.e. $G(\overline{x}) = 0$, likely to occur). In order to avoid this ambiguity, one can simply assume that $X := \mathbb{R}^n$. Moreover, for the convexity of φ, in addition to the convexity of g, one can consider the convexity of the lower level cost function f.

An example of bilevel programming problem where CQ (3.9) is satisfied can be found in [11].

Remark 3.1 Optimality conditions similar to those in Theorem 3.4 were also obtained by Dempe, Dutta and Mordukhovich [7,8] in the case where φ is convex and the partial calmness CQ is satisfied. It was required in the latter papers that Ψ be inner semicompact or inner semicontinuous, which are rather strong assumptions. This also justifies the difference between the proof of Theorem 3.4 and those in [7,8], where the aforementioned assumptions are needed for the estimation of the subdifferential of the optimal value function. Our approach is much closer to that of Ye [36], where the above optimality conditions were obtained under different CQs. A more recent result closely related to the one in the above theorem is given in [13], when the lower level feasible set is defined by infinitely many inequality constraints. Here, the partial calmness CQ was also used for the main problem (3.3), while a *closeness qualification condition* was applied on the lower level constraint instead of the Slater CQ used in Theorem 3.4. It is important to mention that CQ (3.9) is stronger than the partial calmness used in [7,8], but the multiplier λ, simply nonnegative in Theorem 3.4, is positive in [7,8] (cf. [11]).

Instead of (3.11), we could have adopted Eq. (3.19) as in [7,8]. But, as it will be clear in what follows, Eq. (3.11) is chosen in order to search a closer relationship between the optimality conditions of the optimal value and KKT reformulations.

Considering the primal KKT reformulation in (3.5), explicit optimality conditions were recently obtained in [25,32]. In what follows, we rather focus our attention on the KKT reformulation in (3.6). Let us first recall that the MFCQ also fails at any feasible point of this problem. Flegel [15] showed that the Guignard CQ, as one of the weakest, may have a greater chance to be satisfied for the more general MPCC. Hence, we apply this CQ to problem (3.6), which is specific to our initial bilevel program (3.1).

For an optimization problem to

$$\text{minimize } f(x) \text{ subject to } g(x) \leq 0,\ h(x) = 0, \tag{3.20}$$

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 C the feasible set of problem (3.20), the Guignard CQ is satisfied at $\overline{x} \in C$, provided that

$$K_C(\overline{x})^* = -\widehat{N}_C(\overline{x}), \tag{3.21}$$

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

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

Theorem 3.5 *Assume that the lower level problem is a convex one and the MFCQ is satisfied at all $y \in \Psi(x)$, $x \in X$. Furthermore, let $(\overline{x}, \overline{y}, \overline{u})$ be a local optimal solution to problem (3.6) and the Guignard CQ be satisfied at $(\overline{x}, \overline{y}, \overline{u})$. Then, there exists $(\lambda, \alpha, \gamma, \nu) \in \mathbb{R}_+ \times \mathbb{R}^p \times \mathbb{R}^k \times \mathbb{R}^m$ such that:*

$$\nabla_x F(\overline{x}, \overline{y}) + \nabla_x g(\overline{x}, \overline{y})^\top (\alpha - \lambda \overline{u}) + \nabla G(\overline{x})^\top \gamma + \nabla_x \mathcal{L}(\overline{x}, \overline{y}, \overline{u})^\top \nu = 0 \tag{3.22}$$

$$\nabla_y F(\overline{x}, \overline{y}) + \nabla_y g(\overline{x}, \overline{y})^\top (\alpha - \lambda \overline{u}) + \nabla_y \mathcal{L}(\overline{x}, \overline{y}, \overline{u})^\top \nu = 0 \tag{3.23}$$

$$\nabla_y f(\overline{x}, \overline{y}) + \nabla_y g(\overline{x}, \overline{y})^\top \overline{u} = 0 \tag{3.24}$$

$$\alpha \geq 0,\ \alpha^\top g(\overline{x}, \overline{y}) = 0 \tag{3.25}$$

$$\overline{u} \geq 0,\ \overline{u}^\top g(\overline{x}, \overline{y}) = 0 \tag{3.26}$$

$$\gamma \geq 0,\ \gamma^\top G(\overline{x}) = 0 \tag{3.27}$$

$$\nabla_y g(\overline{x}, \overline{y})\nu - \lambda g(\overline{x}, \overline{y}) \geq 0,\ \overline{u}^\top (\nabla_y g(\overline{x}, \overline{y})\nu) = 0, \tag{3.28}$$

where $\mathcal{L}(x, y, u) := \nabla_y f(x, y) + \nabla_y g(x, y)^\top u$.

Proof From the discussion in the previous subsection, the convexity of the lower level problem and the satisfaction of the MFCQ at all $y \in \Psi(x)$, $x \in X$, are simple formalities necessary to make sure that problem (3.6) is closely related to the initial bilevel program (3.1), as summarized in Theorem 3.3. Now, let us set $\psi(x, y, u) := (-u^\top g(x, y),\ g(x, y),\ G(x), -u)^\top$, then $(\overline{x}, \overline{y}, \overline{u})$ is a local optimal solution of the problem

$$\text{minimize } F(x, y) \text{ subject to } \psi(x, y, u) \leq 0,\ \mathcal{L}(x, y, u) = 0, \tag{3.29}$$

since $u^\top g(x, y) = 0$ if and only if $-u^\top g(x, y) \leq 0$, given that $u \geq 0$ and $g(x, y) \leq 0$. Further let us denote by C the feasible set of the previous problem. It follows from Mordukhovich [22, Proposition 5.1] that

$$-\nabla \mathcal{F}(\overline{x}, \overline{y}, \overline{u}) \in \widehat{N}_C(\overline{x}, \overline{y}, \overline{u}),$$

with $\mathcal{F}(x, y, u) := F(x, y)$. Applying the Guignard CQ, $\nabla\mathcal{F}(\overline{x}, \overline{y}, \overline{u}) \in K_C(\overline{x}, \overline{y}, \overline{u})^*$; i.e.

$$-\nabla\mathcal{F}(\overline{x}, \overline{y}, \overline{u})^\top d \leq 0, \ \forall d \in K_C(\overline{x}, \overline{y}, \overline{u}).$$

The rest of the proof then follows by the well-known Farkas' Lemma of the alternative. □

It is worth mentioning that the equivalent transformation, in the above proof, of equation $u^\top g(x, y) = 0$ into an inequality helped to obtain the nonnegativity of the multiplier λ. Hence, contributing to establish an outlook closer to that of the optimality conditions obtained for the optimal value reformulation. As far as condition (3.28) is concerned, one can easily check that it is obtained by combining the derivative of the Lagrangian function of problem (3.6) w.r.t. to u, and the complementarity system resulting from the constraint $-u \leq 0$. This leads to:

$$\beta = \nabla_y g(\overline{x}, \overline{y})\nu - \lambda g(\overline{x}, \overline{y}) \text{ and } \beta \geq 0, \ \overline{u}^\top \beta = 0.$$

Hence, implying (3.28). We now provide an example of bilevel programming problem where the Guignard CQ is satisfied for the KKT reformulation.

Example 3.1 We consider the problem to

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

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

$$\text{minimize } x + y \text{ subject to } x, y \geq 0, \ xy = 0$$

and following Flegel [15], the Guignard CQ holds at the unique optimal solution point $(0, 0)$.

For some more details on the application of the Guignard CQ to MPCCs, we refer the interested reader to the PhD thesis of Flegel [15].

Ye [36] mentioned that under a suitable CQ, the optimality conditions (3.22)–(3.27) could be obtained. Theorem 3.5 suggests that the Guignard CQ is one of such CQs. But, surprisingly, (3.28) is not among the conditions in [36]. It seems to have been omitted purposely. In fact, if we drop condition (3.28), then, as observed in [36], the optimality conditions of Theorem 3.4 and those of Theorem 3.5 coincide in two cases: (a) when $\nu = 0$ and (b) if the lower level problem is affine linear in (x, y). However, by observing the optimality conditions in the results above, the following implication is obvious.

Assume that there exists $\overline{v} = (\lambda, \alpha, \gamma, \nu)$ such that $(\overline{x}, \overline{y}, \overline{u}, \overline{v})$ satisfies conditions (3.22)–(3.28) with $\nu = 0$ or $\nabla_{x,y}\mathcal{L}(\overline{x}, \overline{y}, \overline{u}) = 0$. Then $(\overline{x}, \overline{y}, \overline{v}')$ (with $\overline{v}' = (\lambda, \alpha, \overline{u}, \gamma)$) satisfies (3.10)–(3.15). Clearly, this implication is satisfied if the lower level cost and constraint functions f and g are affine linear in (x, y). In general though, there is no relationship between the optimality conditions of the optimal value and KKT reformulations of the bilevel programming problem; cf. e.g. Henrion and Surowiec [16], for an illustrating example. This difference can be briefly explained as follows.

At first, consider the Lagrange function $\mathcal{L}(x, y, u) := \nabla_y f(x, y) + \nabla_y g(x, y)^\top u$, present in the constraints of the KKT reformulation (3.6). The process of deriving optimality conditions generates second order derivatives for f and g, something which is not necessary when considering the generalized derivative of the value function φ in the optimal value reformulation (3.3). On the other hand, the complementarity system (3.7) embedded in problem (3.6) leads to various types of optimality conditions known in the MPEC literature, such as the S, M and C types, just to name a few [15,31]. It may be important to mention that the optimality conditions in Theorem 3.5 are closely related to the S type conditions. Our interest in the latter category of conditions has been solely motivated by our aim to seek the closest link possible between the optimality conditions of both reformulations. It is worth mentioning that there are many other types of optimality conditions for the optimal value reformulation as well. For instance, necessary optimality conditions for (3.3) have been obtained in [7,11] highlighting a convex combination of an upper bound of the subdifferential of φ. In [11], a new type of optimality conditions was also provided in the case where the lower level problem is strongly stable in the sense of Kojima. Clearly, the variety of optimality conditions for the KKT reformulation is due to the complementarity system (3.7) while for the optimal value reformulation the difference between the types of optimality conditions is justified by the existence of various estimates for the subdifferential of φ.

Finally, on the CQs, as already mentioned above, the MFCQ fails for the optimal value reformulation, see [11,40]. It also fails for the KKT reformulation when considered as a usual nonlinear optimization problem. But for the latter reformulation, the MFCQ could regain its effectiveness provided a suitable transformation of the feasible set is made, see e.g. [10,31]. For the partial calmness (equivalent to a partial exact penalization) introduced by Ye and Zhu [40] for the optimal value reformulation, it has been shown to be very effective for some classes of problems [7,11,38,40]. Moreover, the result of the next section (see Theorem 4.2) reinforces this fact since it shows that problem (3.3) is partially calm if the follower's problem is linear in the lower level variable. For the KKT reformulation (3.6), the concept of partial exact penalization is also applicable, see [10,41]. But in this case, the constraint of interest is the complementarity slackness condition $u^\top g(x, y) = 0$, which is responsible for the failure of the MFCQ. So, like the optimal value constraint $f(x, y) - \varphi(x) \leq 0$ for the optimal value reformulation, the constraint $u^\top g(x, y) = 0$ can be moved to the upper level objective function in problem (3.6) provided the uniform weak sharp minimum condition known for the optimal value reformulation is satisfied [41]. For problem (3.6), this approach would lead to the same optimality conditions as in Theorem 3.5 [10]. For the weaker form of the MFCQ (3.9) applied in Theorem 3.4, the same concept, i.e. a CQ with boundary, remains applicable but with a different form. The interested reader is referred to [11] and references therein, for such results. But it should be mentioned that such a CQ may not lead to the same type of optimality conditions obtained in Theorem 3.5. As far as other CQs are concerned (including the

Guingard CQ), their behavior on the optimal value and KKT reformulations still have to be compared. This could be a topic for future research.

4 Partial calmness

From now on we focus our attention on the optimal value reformulation of the bilevel programming problem (3.1). That is, the problem to

$$\begin{aligned} &\text{minimize } F(x, y) \\ &\text{subject to } \begin{cases} f(x, y) \leq \varphi(x) \\ x \in X, y \in K(x). \end{cases} \end{aligned} \tag{4.1}$$

Let $(\bar{x}, \bar{y})$ be a feasible point of (4.1). Problem (4.1) is partially calm at $(\bar{x}, \bar{y})$ if there is a number $\mu > 0$ and a neighborhood U of $(\bar{x}, \bar{y}, 0)$ such that

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

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

$$\begin{aligned} &\text{minimize } F(x, y) \\ &\text{subject to } \begin{cases} f(x, y) - \varphi(x) + u = 0 \\ x \in X, y \in K(x). \end{cases} \end{aligned}$$

Our interest to the partial calmness is led by its capacity to help to 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 [40] *Let $(\bar{x}, \bar{y})$ be a local optimal solution of problem (4.1). This problem is partially calm at $(\bar{x}, \bar{y})$ if and only if there exists $\lambda > 0$ such that $(\bar{x}, \bar{y})$ is a local optimal solution of the partially penalized problem to*

$$\begin{aligned} &\textit{minimize } F(x, y) + \lambda(f(x, y) - \varphi(x)) \\ &\textit{subject to } x \in X, y \in K(x). \end{aligned}$$

From this result, it is clear that the optimality conditions derived in Theorem 3.4 could also be obtained if we replace CQ (3.9) by the partial calmness. But, as mentioned in Remark 3.1, the multiplier λ of Theorem 3.4 will be positive under the partial calmness, even though the latter CQ, as shown in [11], is weaker than CQ (3.9).

In their seminal paper [40] 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

$$\text{minimize } F(x, y) \text{ subject to } x \in \mathbb{R}^n, y \in \Psi(x), \tag{4.2}$$

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

$$\text{minimize } a(x)^\top y + b(x) \text{ subject to } C(x)y \leq d(x), \tag{4.3}$$

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 (4.2):

$$\begin{aligned} &\text{minimize } F(x, y) \\ &\text{subject to } \begin{cases} a(x)^\top y + b(x) \leq \varphi(x) \\ C(x)y \leq d(x). \end{cases} \end{aligned} \tag{4.4}$$

Theorem 4.2 *Let $(\overline{x}, \overline{y})$ be an optimal solution of problem* (4.4)*, the function F be Lipschitz continuous and dom $\Psi = \mathbb{R}^n$. Then, problem* (4.4) *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

$$\begin{aligned} &a(x')^\top y' + b(x') - \varphi(x') + u = 0 \\ &C(x')y' \leq d(x'). \end{aligned} \tag{4.5}$$

Since dom $\Psi = \mathbb{R}^n$, let $y(x')$ be a solution to the lower level problem (4.3) for the parameter x', i.e. the couple $(x', y(x'))$ is feasible to problem (4.4). 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

$$\begin{aligned} \|y' - y_o(x')\| &= \min_y \{\|y' - y\| : y \in \Psi(x')\} \\ &= \min_{\varepsilon, y} \{\varepsilon : \|y' - y\| \leq \varepsilon, y \in \Psi(x')\} \\ &= \min_{\varepsilon, y} \{\varepsilon : -\varepsilon e \leq y' - y \leq \varepsilon e, y \in \Psi(x')\}. \end{aligned}$$

The last equality describes the linear program

$$\begin{aligned} &\min_{\varepsilon, y} \varepsilon \\ &\begin{cases} -\varepsilon e + y \leq y' \\ -\varepsilon e - y \leq -y' \\ a(x')^\top y \leq \varphi(x') - b(x') \\ C(x')y \leq d(x'), \end{cases} \end{aligned}$$

having as dual the problem

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

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

$$\max_{\xi_1,\xi_2,\xi_3,\xi_4} \xi_3\big(\varphi(x') - a(x')^\top y' - b(x')\big) + \big(d(x') - C(x')y'\big)^\top \xi_4$$
$$-e^\top \xi_1 - e^\top \xi_2 = 1,\ \xi_1, \xi_2, \xi_3, \xi_4 \le 0.$$

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

$$-e^\top \xi_1 - e^\top \xi_2 = 1,\ \xi_1, \xi_2, \xi_3, \xi_4 \le 0 \tag{4.6}$$

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

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

given that (x', y') satisfies (4.5). Also notice that $u \le 0$ considering the definition of φ. Since the number of vertices satisfying (4.6) is finite, let $\xi_3^B \in \mathbb{R}$ be the smallest $(2m+1)$th component of such vertices, then

$$\|y' - y_o(x')\| \le |\xi_3^B||u|. \tag{4.7}$$

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 (4.4) and $(x', y_o(x'))$ being feasible, we have

$$\begin{aligned} F(x', y') - F(\overline{x}, \overline{y}) &\ge F(x', y') - F(x', y_o(x')) \\ &\ge -K_F\|y' - y_o(x')\| \\ &\ge -\mu|u|, \end{aligned}$$

with $\mu = K_F|\xi_3^B|$ and taking into account inequality (4.7). □

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

We now consider the bilevel programming problem (4.2) in the case where $X \ne \mathbb{R}^n$. 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 [40].

Definition 4.1 The family of parametric optimization problems $\{(3.2)|x \in X\}$ is said to have a uniformly weak sharp minimum if there exist $\mu > 0$ such that

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

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 (3.1) is partially calm at every local optimal solution, provided that the family of parametric optimization problems $\{(3.2)|x \in X\}$ has a uniformly weak sharp minimum. Hence, this concept has been adopted as a standard criterion for the partial calmness and many sufficient conditions for the uniform weak sharp minimum have been given in the literature [11,38,40,41]. Very recently though, Henrion and Surowiec [16] showed that the calmness of the set-valued mapping

$$\mathcal{M}(v) := \{(x, y) \in \Omega |\, f(x, y) - \varphi(x) \leq v\}$$

could be inserted between the uniform weak sharp minimum and the partial calmness. Hence, providing a weaker sufficient condition for the latter to be satisfied. It should, nevertheless, be mentioned that it was already clear from [11], see proof of Theorem 4.10, that the calmness of $\mathcal{M}$ is a sufficient condition for the partial calmness.

Ye [38] considered the bilevel programming problem (3.1) where the lower level problem is defined as

$$\min_y \{f(x, y)|\, y \in \mathbb{R}^m,\ g(x, y) \leq 0\}, \tag{4.8}$$

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

Theorem 4.3 *Assume that* dom $\Psi = X$. *Let*

$$\overline{g}(x, y)^\top := (g_1(x, y), \ldots, g_p(x, y), f(x, y) - \varphi(x))$$

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

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

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 (4.8) is nothing but the lower level problem considered in (4.3). Hence if $X = \mathbb{R}^n$, then the result in Theorem 4.2 remains true for the bilevel programming problem (4.2) where the lower level problem is defined in (4.8). It thus seems clear that in this case the assumption of Theorem 4.3, which may be quite difficult to check (see Mangasarian and Shiau [18]), is not useful. Nevertheless, when $X \neq \mathbb{R}^n$, the previous result takes all its importance.

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 $\mathcal{W} \subseteq \mathcal{N}^2$ 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 us denote by Q the set-valued mapping from $\mathbb{R}^{\omega}$ to $\mathbb{R}^{\pi}$ describing the set of such flows, then

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

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 $\mathbb{R}^{\alpha}$

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

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 [33] 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

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

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

$$\begin{aligned} &\text{minimize } f(v) := \sum_{a \in \mathcal{A}} \int_0^{v_a} t_a(s) ds \\ &\text{subject to } v \in V(d), \end{aligned} \tag{5.1}$$

provided that each link cost function $t_a : \mathbb{R} \to \mathbb{R}$ is continuous and positive. Problem (5.1) is known as the traffic assignment problem. In general, the total cost of travel represented by f may have a more general expression. The interested reader is referred to Patriksson [28] for a detailed discussion on this issue. In this paper, we simply consider f in problem (5.1) to be a function of the demand d and the link flow v.

In transportation planing, the demand mentioned in the traffic assignment problem above is a strategic data in a broader sense given that a comprehensive decision making process highly depends on how accurate it is estimated. Fisk [14] has suggested a bilevel formulation of the problem of estimating the origin-destination (O-D) matrix (or the O-D demand adjustment problem (DAP)) from an outdated matrix and using some observed traffic counts:

$$\begin{aligned} &\text{minimize } F(d, v) \\ &\text{subject to } d \in D, v \in \Psi(d) \end{aligned} \tag{5.2}$$

where D is a closed set and $\Psi(d)$ the solution set of the traffic assignment problem (5.1) (where $f(v)$ is replaced by $f(d, v)$) 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. For more details on the model and some solution approaches, we refer the interested reader to Abrahamsson [1], Chen [3], Chen and Florian [4], Fisk [14], Migdalas [19], Yang [34] and Yang et al. [35].

One interesting thing about the traffic assignment problem (5.1) 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. (5.4). 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 $V(d) := J \circ Q(d)$), 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 DAP.

We have from Theorem 3.1 that problem (5.2) can be reformulated as

$$\begin{aligned} &\text{minimize } F(d, v) \\ &\text{subject to } f(d, v) \le \varphi(d), (d, v) \in \Omega, \end{aligned} \tag{5.3}$$

where $\Omega := (D \times \mathbb{R}^{\alpha}) \cap \text{gph}V$, with gph$V$ representing the graph of the link flow set-valued mapping V. Here, φ denotes the optimal value function of the traffic assignment problem parameterized in the demand d. Chen [3] and Chen and Florian [4] suggested Fritz John's type optimality conditions for problem (5.3) 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. As already mentioned above $V(d) := J \circ Q(d)$ and we can easily observe that the graph of Q is given by

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

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

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

of unused routes of the network and the set

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

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 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

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

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

$$N_{\text{gph}Q}(\overline{d}, \overline{q}) = \left\{ \left(-\sum_{w \in \mathcal{W}} \lambda_w e^w,\ \sum_{r \in \mathcal{P}^c(\overline{q})} \lambda_r^c e^r - \sum_{r \in \mathcal{P}^o(\overline{q})} \lambda_r^o e^r + \sum_{w \in \mathcal{W}} \lambda_w \Lambda_w^{\top} \right) : \right.$$
$$\left. (\lambda_w)_{w \in \mathcal{W}} \in \mathbb{R}^{\omega},\ (\lambda_r^o)_{r \in \mathcal{P}^o(\overline{q})}, (\lambda_r^c)_{r \in \mathcal{P}^c(\overline{q})} \geq 0 \right\},$$

following Rockafellar and Wets [30, Theorem 6.46], where $[e^w]_{w \in \mathcal{W}}$ is the collection of columns of the identity matrix of $\mathbb{R}^{\omega \times \omega}$.

We now provide some variational properties of the feasible link flows set-valued mapping V, which will be very useful in deriving KKT type optimality conditions for the DAP. For this purpose, we consider the following set of route flows corresponding to a given feasible demand-link flow couple $(\overline{d}, \overline{v})$

$$\mathcal{H}(\overline{d},\overline{v}) := \{q \in \mathbb{R}^{\pi} \,|\, \Delta q = \overline{v},\, (\overline{d}, q) \in \text{gph} Q\}. \tag{5.4}$$

Lemma 5.1 *For any* $(\overline{d},\overline{v}) \in gph\, V$*, we have*

$$N_{gph\,V}(\overline{d},\overline{v}) \subseteq \bigcup_{\overline{q}\in\mathcal{H}(\overline{d},\overline{v})} \{(d^*, v^*) \in \mathbb{R}^{\omega} \times \mathbb{R}^{\alpha} \,|\, (d^*, \Delta^{\top} v^*) \in N_{gph\,Q}(\overline{d},\overline{q})\}.$$

Moreover, V *is Lipschitz-like around* $(\overline{d},\overline{v})$*, provided the following CQ holds at* $(\overline{d},\overline{v})$*:*

$$\left.\begin{array}{r} \sum_{r\in\mathcal{P}^c(\overline{q})} \lambda_r^c e^r - \sum_{r\in\mathcal{P}^o(\overline{q})} \lambda_r^o e^r + \sum_{w\in\mathcal{W}} \lambda_w \Lambda_w^{\top} = 0 \\ (\lambda_r^o)_{r\in\mathcal{P}^0(\overline{q})}, (\lambda_r^c)_{r\in\mathcal{P}^c(\overline{q})} \geq 0 \\ \overline{q} \in \mathcal{H}(\overline{d},\overline{v}) \end{array}\right\} \Longrightarrow [\lambda_w = 0,\ w \in \mathcal{W}]. \tag{5.5}$$

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 we have $d^* \in D^*V(\overline{d},\overline{q})(-v^*)$. Since $Q(d) \subseteq |c|\mathbb{B}$, 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 $(d, v) \rightrightarrows Q(d) \cap J^{-1}(v) = \mathcal{H}(d, v)$ is uniformly bounded around $(\overline{d},\overline{v})$. Hence, including the continuous differentiability of the function J, it follows from Rockafellar and Wets [30, p. 454] that there exists $\overline{q} \in \mathcal{H}(\overline{d},\overline{v})$ such that $d^* \in D^*Q(\overline{d},\overline{q})(\Delta^{\top}(-v^*))$, that is $(d^*, \Delta^{\top} v^*) \in N_{\text{gph}\,Q}(\overline{d},\overline{q})$.

For the Lipschitz-like property of V, it follows from the upper estimate of $N_{\text{gph}\,V}$ in the theorem that

$$D^*V(\overline{d},\overline{v})(v^*) \subseteq \bigcup_{\overline{q}\in\mathcal{H}(\overline{d},\overline{v})} \{d^* \in \mathbb{R}^{\omega} \,|\, (d^*, -\Delta^{\top} v^*) \in N_{\text{gph}\,Q}(\overline{d},\overline{q})\}.$$

Furthermore, if one considers the expression of the normal cone to the graph of Q, one has

$$D^*V(\overline{d},\overline{v})(0) \subseteq \bigcup_{\overline{q}\in\mathcal{H}(\overline{d},\overline{v})} \left\{ -\sum_{w\in\mathcal{W}} \lambda_w e^w \,\middle|\, \sum_{r\in\mathcal{P}^c(\overline{q})} \lambda_r^c e^r - \sum_{r\in\mathcal{P}^o(\overline{q})} \lambda_r^o e^r + \sum_{w\in\mathcal{W}} \lambda_w \Lambda_w^{\top} = 0 \right. \\ \left. (\lambda_w)_{w\in\mathcal{W}} \in \mathbb{R}^{\omega},\ (\lambda_r^o)_{r\in\mathcal{P}^0(\overline{q})}, (\lambda_r^c)_{r\in\mathcal{P}^c(\overline{q})} \geq 0 \right\}.$$

We recall that V is Lipschitz-like around $(\overline{d},\overline{v})$ provided the following coderivative criterion of Mordukhovich, see [22], holds

$$D^*V(\overline{d},\overline{v})(0) = \{0\}.$$

Considering the latter inclusion, this is automatically satisfied under CQ (5.5). □

Example 5.1 We consider the problem of Example 1 in the paper of Lu [17], with a network of four links ($\alpha = 4$), two O-D pairs ($\omega = 2$) and four routes ($\pi = 4$). Furthermore, let the O-D-route incidence matrix (Λ) and the link-route incidence matrix (Δ)

be defined as in Lu's Example, then it follows from [17] that for $\overline{d} = [20, 20]^\top$ and $\overline{v} = [10, 10, 10, 10]^\top$, one has $\overline{v} \in V(\overline{d})$ and $\mathcal{H}(\overline{d}, \overline{v}) = \{\overline{q} = [10, 10, 10, 10]^\top\}$. Clearly, all the routes are used. If we set the route capacity vector to be $[10, 30, 30, 10]^\top$, meaning that routes 1 and 4 are used at full capacity, while routes 2 and 3 are underused, the left hand side of implication (5.5) reduces to

$$[\lambda_1^c + \lambda_1,\ \lambda_1,\ \lambda_2,\ \lambda_4^c + \lambda_2] = [0, 0, 0, 0],\ \text{ with } \lambda_1^c, \lambda_4^c \geq 0.$$

This obviously implies that $\lambda_1 = \lambda_2 = 0$. Thus, CQ (5.5) is satisfied at $(\overline{d}, \overline{v})$.

Remark 5.1 For a given demand-link flow couple $(\overline{d}, \overline{v})$, if we assume that all routes are used but not at full capacity, i.e.

$$\mathcal{P}^c(\overline{q}) = \emptyset \text{ and } \mathcal{P}^c(\overline{q}) = \emptyset,\ \forall \overline{q} \in \mathcal{H}(\overline{d}, \overline{v}),$$

then CQ (5.5) reduces to the fulfillment of the implication:

$$\sum_{w \in \mathcal{W}} \lambda_w \Lambda_w^\top = 0 \Longrightarrow [\lambda_w = 0,\ w \in \mathcal{W}].$$

This amounts to saying that the O-D-route incidence matrix has full rank. This assumption is satisfied for Λ in the network of the above example.

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 us denote by

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

the set of Lagrange multipliers for the traffic assignment problem. Then, the local Lipschitz continuity and the subdifferential of the optimal value function of the traffic assignment problem can be obtained as follows.

Theorem 5.1 *Let $\overline{d} \in D$ such that for all $\overline{v} \in \Psi(\overline{d})$, CQ (5.5) holds at $(\overline{d}, \overline{v})$. Then, the optimal value function φ of the traffic assignment problem is locally Lipschitz continuous around $\overline{d}$ and*

$$\partial\varphi(\overline{d}) \subseteq \bigcup_{\overline{v} \in \Psi(\overline{d})} \bigcup_{\overline{q} \in \mathcal{H}(\overline{d}, \overline{v})} \bigcup_{(\lambda^\omega, \lambda^c, \lambda^0) \in \Lambda(\overline{d}, \overline{q})} \left\{ -\sum_{w \in \mathcal{W}} \lambda_w e^w + \nabla_d f(\overline{d}, \overline{v}) \right\}. \tag{5.6}$$

Alternatively, if we assume that Ψ is inner semicontinuous at $(\overline{d}, \overline{v})$, and CQ (5.5) holds at $(\overline{d}, \overline{v})$, then φ is also locally Lipschitz continuous around $\overline{d}$ and

$$\partial\varphi(\bar{d}) \subseteq \bigcup_{\bar{q}\in\mathcal{H}(\bar{d},\bar{v})} \bigcup_{(\lambda^{\omega},\lambda^{c},\lambda^{0})\in\Lambda(\bar{d},\bar{q})} \left\{-\sum_{w\in\mathcal{W}}\lambda_w e^w + \nabla_d f(\bar{d},\bar{v})\right\}. \tag{5.7}$$

Proof We first start by noting that since J is a continuous function, we 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. Sect. 2. In addition to the Lipschitz-like property of V around $(\bar{d},\bar{v})$, for all $\bar{v}\in\Psi(\bar{d})$ (cf. Lemma 5.1), the value function φ is locally Lipschitz continuous following Mordukhovich and Nam [23, Theorem 5.2].

On the other hand, with the continuous differentiability of f, it follows from [24, Theorem 7] that

$$\partial\varphi(\bar{d}) \subseteq \bigcup_{\bar{v}\in\Psi(\bar{d})} \{\nabla_d f(\bar{d},\bar{v}) + D^*V(\bar{d},\bar{v})(\nabla_v f(\bar{d},\bar{v}))\} \tag{5.8}$$

considering the inner semicompactness of Ψ at $\bar{d}$. Moreover, it follows from the definition of the coderivative that $d^* \in D^*V(\bar{d},\bar{v})(\nabla_v f(\bar{d},\bar{v}))$ if and only if $(d^*, -\nabla_v f(\bar{d},\bar{v})) \in N_{\mathrm{gph}V}(\bar{d},\bar{v})$. Hence, from Lemma 5.1, there exists $\bar{q}\in\mathcal{H}(\bar{d},\bar{v})$ such that $(d^*, -\Delta^\top\nabla_v f(\bar{d},\bar{v})) \in N_{\mathrm{gph}Q}(\bar{d},\bar{q})$, which implies the existence of $(\lambda^\omega,\lambda^c,\lambda^o)$, with $\lambda^\omega = (\lambda_w)$, $\lambda^c = (\lambda_r^c) \geq 0$, $\lambda^o = (\lambda_r^o) \geq 0$ such that

$$d^* = -\sum_{w\in\mathcal{W}}\lambda_w e^w \tag{5.9}$$

$$\Delta^\top\nabla_v f(\bar{d},\bar{v}) = -\sum_{r\in\mathcal{P}^c(\bar{q})}\lambda_r^c e^r + \sum_{r\in\mathcal{P}^o(\bar{q})}\lambda_r^o e^r - \sum_{w\in\mathcal{W}}\lambda_w\Lambda_w^\top. \tag{5.10}$$

Hence, the inclusion in (5.6) follows by the combination of (5.8) and (5.9)–(5.10).

In the case where Ψ is inner semicontinuous, the local Lipschitz continuity and the inclusion (5.7) follow similarly by applying [23, Theorem 5.2] and [24, Theorem 7], respectively. □

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

Theorem 5.2 *Let $(\bar{d},\bar{v})$ be a local optimal solution to problem (5.3) and assume that CQ (5.5) holds at $(\bar{d},\bar{v})$, for all $\bar{v}\in\Psi(\bar{d})$. Furthermore, let problem (5.3) be partially calm at $(\bar{d},\bar{v})$. Then, there exist $\bar{q}\in\mathbb{R}^\pi$, $\mu > 0$, $(\lambda^\omega,\lambda^c,\lambda^o)$, and $v_s\in\Psi(\bar{d})$, $q_s\in\mathbb{R}^\pi$, $(\lambda_s^\omega,\lambda_s^c,\lambda_s^o)$ and $\eta_s\geq 0$, $s=1,\ldots,\omega+1$, with $\sum_{s=1}^{\omega+1}\eta_s = 1$ such that*

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

$$\Delta^\top\left(\nabla_v F(\bar{d},\bar{v}) + \mu\nabla_v f(\bar{d},\bar{v})\right) - \sum_{r\in\mathcal{P}^o(\bar{q})}\lambda_r^o e^r + \sum_{r\in\mathcal{P}^c(\bar{q})}\lambda_r^c e^r + \sum_{w\in\mathcal{W}}\lambda_w\Lambda_w^\top = 0$$

$$\Delta^\top \nabla_v f(\overline{d}, v_s) + \sum_{r\in\mathcal{P}^c(q_s)} \lambda^c_{sr} e^r - \sum_{r\in\mathcal{P}^o(q_s)} \lambda^o_{sr} e^r + \sum_{w\in\mathcal{W}} \lambda_{sw} \Lambda_w^\top = 0$$
$$0 \leq \overline{q} \leq c,\ \Lambda\overline{q} = \overline{d},\ \Delta\overline{q} = \overline{v}$$
$$0 \leq q_s \leq c,\ \Lambda q_s = \overline{d},\ \Delta q_s = v_s$$
$$\lambda^\omega = (\lambda_w),\ \lambda^c = (\lambda^c_r) \geq 0,\ \lambda^o = (\lambda^o_r) \geq 0$$
$$\lambda^\omega_s = (\lambda_{sw}),\ \lambda^c_s = (\lambda^c_{sr}) \geq 0,\ \lambda^o_s = (\lambda^o_{sr}) \geq 0.$$

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 gph$V$ is closed and φ is locally Lipschitz continuous around $\overline{d}$, it follows from Mordukhovich [22, Proposition 5.3] that

$$0 \in \partial(F + \mu(f - \varphi))(\overline{d}, \overline{v}) + N_{\text{gph } V}(\overline{d}, \overline{v}). \tag{5.11}$$

Considering the sum rule (2.2) and the convex hull property (2.1) 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 $\lambda^\omega = (\lambda_w),\ \lambda^c = (\lambda^c_r) \geq 0,\ \lambda^o = (\lambda^o_r) \geq 0$ such that

$$-\sum_{w\in\mathcal{W}} \lambda_w e^w + \nabla_d F(\overline{d}, \overline{v}) + \mu\nabla_d f(\overline{d}, \overline{v}) \in \mu\text{co}\partial\varphi(\overline{d}) \tag{5.12}$$
$$\Delta^\top v^* = \sum_{w\in\mathcal{W}} \lambda_w \Lambda_w^\top + \sum_{r\in\mathcal{P}^c(\overline{q})} \lambda^c_r e^r - \sum_{r\in\mathcal{P}^o(\overline{q})} \lambda^o_r e^r \tag{5.13}$$
$$v^* = -\nabla_v F(\overline{d}, \overline{v}) - \mu\nabla_v f(\overline{d}, \overline{v}) \tag{5.14}$$
$$0 \leq \overline{q} \leq c,\ \Lambda\overline{q} = \overline{d},\ \Delta\overline{q} = \overline{v}. \tag{5.15}$$

On the other hand, if we take $d^* \in \text{co}\varphi(\overline{d})$, then we have from Caratheodory's Theorem (see e.g. Rockafellar [29]) 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 (5.12)–(5.15) and the inclusion (5.6). □

In the case where the solution set-valued mapping Ψ is inner semicontinuous, 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 5.3 *Let $(\overline{d}, \overline{v})$ be an optimal solution to problem (5.3), Ψ be inner semicontinuous at $(\overline{d}, \overline{v})$ and CQ (5.5) be satisfied at $(\overline{d}, \overline{v})$. Furthermore, let problem (5.3) 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 $q_s \in \mathbb{R}^\pi$, $(\lambda^\omega_s, \lambda^c_s, \lambda^o_s)$ and $\eta_s \geq 0, s = 1, \ldots, \omega + 1$, with $\sum_{s=1}^{\omega+1} \eta_s = 1$ such that*

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

A generalization of Theorems 5.2 and 5.3 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

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

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 (5.11). For the inner semicontinuity of Ψ, this is automatically satisfied 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 [3,34,35]. It is obtained by assuming that for every link $a \in \mathcal{A}$, the link cost t_a is strictly increasing in v_a. Given that gph V is a convex set, CQ (5.5) can be dropped in the above results, provided that the function f is convex in (d, v), as it is the case in (5.1). In fact, in this case, φ would be convex and hence, the approach to optimality conditions in Theorem 3.4 is more appropriate. The interested reader is referred to [9] for a detailed discussion on this case. Finally, as far as the partial calmness of the DAP is concerned, it follows from Theorem 4.2 that this will, for example, be satisfied if the traffic cost of the road users expressed by f is linear in the link flow. That is, if there is no congestion in the network, this function can take the form $f(d, v) := \sum_{a\in\mathcal{A}} \alpha_a v_a$, where α_a denotes the fix cost of travel for each traveler on link a which can represent a combination of the time spent and the fuel consumed on this link.

Acknowledgments The authors would like to thank the two referees who's comments and suggestions have contributed to important improvements in the first version of the paper. The second author also acknowledges the support by the Deutscher Akademischer Austausch Dienst (DAAD).

References

1. Abrahamsson, T.: Estimation of origin-destination matrices using traffic counts: a literature survey. Interim Report IR-98-021/May, Int. Inst. Appl. Syst. Anal., Laxenburg, Austria. http://www.iiasa.ac.at/Admin/PUB/Documents/IR-98-021.pdf (1998)
2. Beckmann, M., Mcguire, C.B., Winsten, C.B.: Studies in the Economics of Transportation. Yale University Press, New Haven (1956)

3. Chen, Y.: Bilevel programming problems: analysis, algorithms and applications. PhD Thesis, Centre de Recherche sur les Transports, Université de Montréal, CRT-984 (1994)
4. Chen, Y., Florian, M.: Congested O-D trip demand adjustement problem: bilevel programming formulation and optimality conditions. In: Migdalas, A. et al. (eds.) Multilevel Optimization: Algorithms and Applications, pp. 1–22. Kluwer Academic Publishers, Dordrecht (1998)
5. Dempe, S.: Foundations of Bilevel Programming. Kluwer Academic Publishers, Dordrecht (2002)
6. Dempe, S., Dutta, J.: Is bilevel programming a special case of mathematical programming with equilibrium constraints? Math. Program. (2010). doi:10.1007/s10107-010-0342-1
7. Dempe, S., Dutta, J., Mordukhovich, B.S.: New necessary optimality conditions in optimistic bilevel programming. Optimization **56**, 577–604 (2007)
8. Dempe, S., Dutta, J., Mordukhovich, B.S. : Variational analysis in bilevel programming. In: Neogy, S.K., Bapat, R.B., Das, A.K., Parthasarathy, T. (eds.) Statistical Science and Interdisciplinary Research 1, Mathematical Programming and Game Theory for Decision Making, pp. 257–277. World Scientific, Hackensack (2008)
9. Dempe, S., Zemkoho, A.B.: Bilevel road pricing: theoretical analysis and optimality conditions. Ann. Oper. Res. (2011). doi:10.1007/s10479-011-1023-z
10. Dempe, S., Zemkoho, A.B.: On the Karush-Kuhn-Tucker reformulation of the bilevel optimization problem. Nonlinear Anal. **75**, 1202–1218 (2012)
11. Dempe, S., Zemkoho, A.B.: The generalized Mangasarian-Fromowitz constraint qualification and optimality conditions for bilevel programs. J. Optim. Theory Appl. **148**, 433–441 (2011)
12. Dempe, S., Zemkoho, A.B.: A bilevel approach for traffic management in capacitated networks. Preprint 2008-05, Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg (2008)
13. Dinh, N., Mordukhovich, B.S., Nghia, T.T.A.: Subdifferentials of value functions and optimality conditions for DC and bilevel infinite and semi-infinite programs. Math. Program. Ser. B **123**, 101–138 (2010)
14. Fisk, C.S.: On combining maximum maximum entropy trip matrix with useroptimal assigment. Transp. Res. **22B**, 69–73 (1988)
15. Flegel, M.: Constraint qualification and stationarity concepts for mathematical programs with equilibrium constraints. PhD Thesis, Institute of Applied Mathematics and Statistics, University of Würzburg (2005)
16. Henrion, R., Surowiec, T.M.: On calmness conditions in convex bilevel programming. Appl. Anal. **90**, 951–970 (2011)
17. Lu, S.: Sensitivity of static traffic user equilibria with perturbations in arc cost function and travel demand. Transp. Sci. **42**, 105–123 (2008)
18. Mangasarian, O.L., Shiau, T.-H.: Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems. SIAM J. Control Optim. **25**, 583–595 (1987)
19. Migdalas, A.: Bilevel programming in traffic planning: models, methods and challenge. J. Global Optim. **7**, 381–405 (1995)
20. Mirrlees, J.A.: The theory of moral hazard and unobservable behaviour. Rev. Econ. Stud. **66**, 3–21 (1999)
21. Mordukhovich, B.S.: Maximum principle in the problem of time optimal response with nonsmooth constraints. J. Appl. Math. Mech. **40**, 960–969 (1976)
22. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation. I: Basic Theory. II: Applications. Springer, Berlin (2006)
23. Mordukhovich, B.S., Nam, N.M.: Variational stability and marginal functions via generalized differentiation. Math. Oper. Res. **30**, 800–816 (2005)
24. Mordukhovich, B.S., Nam, N.M., Yen, N.D.: Subgradients of marginal functions in parametric mathematical programming. Math. Program. **116**, 369–396 (2007)
25. Mordukhovich, B.S., Outrata, J.V.: Coderivative analysis of quasi-variational inequalities with applications to stability and optimization. SIAM J. Optim. **18**, 389–412 (2007)
26. Outrata, J.V.: A note on the usage of nondifferentiable exact penalties in some special optimization problems. Kybernetika **24**, 251–258 (1988)
27. Outrata, J.V.: On the numerical solution of a class of Stackelberg problems. Z. Oper. Res. **34**, 255–277 (1990)
28. Patriksson, M.: The Traffic Assignment Problem—Models and Methods. VSP BV, Utrecht, The Netherlands (1994)
29. Rockafellar, R.T.: Convex Analysis. Princeton University Press, New Jersey (1970)
30. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)

31. Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. Math. Oper. Res. **25**, 1–22 (2000)
32. Surowiec, T.M.: Explicit stationary conditions and solution characterization for equilibrium problems with equilibrium constraints. PhD Thesis, Mathematisch-Naturwissentschaften Fakultät II, Humbolt Universität zu Berlin (2010)
33. Wardrop, J.G.: Some theoretical aspects of road traffic research. Proc. Inst. Civil Engeneers **Part II**, 325–378 (1952)
34. Yang, H.: Heuristic algorithms for the bilevel origin-destination matrix estimation problem. Transp. Res. **29B**, 231–242 (1995)
35. Yang, H., Sasaki, T., Iida, Y., Asakura, Y.: Estimation of origin-destination matrices from traffic counts on congested networks. Transp. Res. **26B**, 417–434 (1992)
36. Ye, J.J.: Nondifferentiable multiplier rules for optimization and bilevel optimization problems. SIAM J. Optim. **15**, 252–274 (2004)
37. Ye, J.J., Zhu, D.L.: New necessary optimality conditions for bilevel programs by combining MPEC and the value function approach. SIAM J. Optim. **20**, 1885–1905 (2010)
38. Ye, J.J.: New uniform parametric error bounds. J. Optim. Theory Appl. **98**, 197–219 (1998)
39. Ye, J.J., Ye, X.Y.: Necessary optimality conditions for optimization problems with variational inequality constraints. Math. Oper. Res. **22**, 977–997 (1997)
40. Ye, J.J., Zhu, D.L.: Optimality conditions for bilevel programming problems. Optimization **33**, 9–27 (1995)
41. Ye, J.J., Zhu, D.L., Zhu, Q.J.: Exact penalization and necessary optimality conditions for generalized bilevel programming problems. SIAM J. Optim. **7**, 481–507 (1997)