Ann Oper Res (2012) 196:223–240
DOI 10.1007/s10479-011-1023-z

Bilevel road pricing: theoretical analysis and optimality conditions

S. Dempe · A.B. Zemkoho

Published online: 16 November 2011

Abstract We consider the bilevel road pricing problem. In contrary to the Karush-Kuhn-Tucker (one level) reformulation, the optimal value reformulation is globally and locally equivalent to the initial problem. Moreover, in the process of deriving optimality conditions, the optimal value reformulation helps to preserve some essential data involved in the traffic assignment problem that may disappear with the Karush-Kuhn-Tucker (KKT) one. Hence, we consider in this work the optimal value reformulation of the bilevel road pricing problem; using some recent developments in nonsmooth analysis, we derive implementable KKT type optimality conditions for the problem containing all the necessary information. The issue of estimating the (fixed) demand required for the road pricing problem is a quite difficult problem which has been also addressed in recent years using bilevel programming. We also show how the ideas used in designing KKT type optimality conditions for the road pricing problem can be applied to derive optimality conditions for the origin-destination (O-D) matrix estimation problem. Many other theoretical aspects of the bilevel road pricing and O-D matrix estimation problems are also studied in this paper.

Keywords Bilevel programming · Road pricing · O-D matrix estimation · Optimal value function · Constraint qualifications · Optimality conditions

1 Introduction

In this paper, we consider the bilevel road pricing problem. In the literature, it has mainly been addressed through the sensitivity analysis and KKT reformulation, as techniques for the one level reformulation. These approaches present some weaknesses including the strong requirements needed for the sensitivity analysis and the unbalanced relationship between the

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

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

KKT reformulation and the initial problem (cf. Dempe and Dutta 2010). Moreover, in the perspective of KKT type conditions for the road pricing problem, the KKT reformulation may cost some essential data for the traffic assignment problem, given the linear structure of the constraint set. Thanks to the optimal value function one level reformulation, these difficulties can directly be dealt with and some new features of the bilevel road pricing problem are highlighted.

The main concern in this paper is to design implementable KKT type optimality conditions containing all the necessary information, in the perspective of new approaches to solve the bilevel road pricing problem, in a rigorous way, and demanding not too strong requirements. It is well-known that the problem of estimating the O-D matrix, necessary in the modeling process of the bilevel road pricing problem presents even more challenges. This problem is also considered in this paper and KKT type optimality conditions are suggested, thereby improving the work of Chen (1994) (also see Chen and Florian 1998), which is discussed in Sect. 4.

In the rest of this section, we first introduce the traffic assignment and bilevel road pricing problems, respectively. Some review of the methodological approaches is also discussed. In Sect. 2 the optimal value function reformulation is introduced and analyzed. The KKT type optimality conditions of the bilevel road pricing and demand adjustment problems are discussed in Sects. 3 and 4, respectively.

We consider a transportation network $\mathscr{G} = (\mathscr{N}, \mathscr{A})$, where $\mathscr{N}$ and $\mathscr{A}$ denote the set of nodes and directed links (arcs), respectively. Let $\mathscr{W} \subset \mathscr{N}^2$ denote the set of origin-destination (O-D) pairs. Each O-D pair $w \in \mathscr{W}$ is connected by a set of routes (paths) $\mathscr{P}_w$, each member of which is a set of sequentially connected links. We denote by $\mathscr{P} = \bigcup_{w\in\mathscr{W}} \mathscr{P}_w$ the set of all routes of the network and by $\alpha = |\mathscr{A}|$, $\omega = |\mathscr{W}|$ and $\pi = |\mathscr{P}|$, the cardinalities of $\mathscr{A}$, $\mathscr{W}$ and $\mathscr{P}$, 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 $p \in \mathscr{P}_w$ and $\Lambda_{wp} = 0$ otherwise, and the matrix $(\Delta = [\Delta_{ap}]) \in \mathbb{R}^{\alpha\times\pi}$ denotes the arc-route incidence matrix with $\Delta_{ap} = 1$ if arc a is in route p and $\Delta_{ap} = 0$ otherwise. The network is assumed to be strongly connected, that is, 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. The column vectors $(c = [c_p]) \in \mathbb{R}^\pi_+$ and $(\tau = [\tau_a]) \in \mathbb{R}^\alpha$ denote the route capacity and arc toll, respectively. 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 of such flows, then

$$Q = \{q \in \mathbb{R}^\pi_+ | q \leq c, \Lambda q = d\}. \tag{1.1}$$

A link flow v is feasible if there exists a feasible route flow q such that the flow conservation constraint $\Delta q = v$, is satisfied. Hence,

$$V = \{v \in \mathbb{R}^\alpha | \exists q \in Q, \Delta q = v\} \tag{1.2}$$

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

Wardrop's user equilibrium principle (Wardrop 1952) states that for every O-D pair $w \in \mathscr{W}$, the travel cost of the routes utilized are equal and minimal for each individual user, that is for each $w \in \mathscr{W}$ and $p \in \mathscr{P}_w$, we have

$$\begin{cases} \overline{c}_p(v,\tau) = \vartheta_w(v,\tau) & \text{if } q_p > 0 \\ \overline{c}_p(v,\tau) \geq \vartheta_w(v,\tau) & \text{if } q_p = 0 \end{cases} \tag{1.3}$$

for any fixed toll pattern $\tau \in \Gamma$. It follows from Beckmann et al. (1956) that for every fixed toll pattern $\tau \in \Gamma$, the Wardrop's user equilibrium problem (1.3) is equivalent to the parametric optimization problem

$$\begin{aligned} &\text{minimize} \quad f(v,\tau) = \sum_{a \in \mathscr{A}} \int_0^{v_a} t_a(s,\tau)ds \\ &\text{subject to} \quad v \in V, \end{aligned} \tag{1.4}$$

provided that for each link $a \in \mathscr{A}$, the link cost takes the form $t_a(v_a, \tau)$; that is, it does not depend on the flow on the other links. In other words, the link costs are separable with respect to the link flows. In addition, they should also be continuous and positive. These assumptions will be maintained for the rest of the paper such that for each toll pattern τ, Wardrop's user equilibrium arc flow will be defined as the solution of the optimization problem (1.4) also called the traffic assignment problem.

We now consider a road authority who intends to improve the circulation on the network $\mathscr{G}$. He/she chooses road pricing as a method to modify the behavior of the road users, by setting tolls on some links of the network to discourage the use of the tolled links in favor of some perhaps abandoned of less used ones. For some simplification in the presentation of the model, we first assume that all the links are tolled (we can have $\tau_a = 0$ for some link $a \in \mathscr{A}$). Later in Sect. 3, we will discuss a possible way to introduce some fairness in the model. The bilevel formulation of this problem enables the road authority to decide while considering the reactions of the road users. If we assume that for each toll pattern τ, the road users choose their origin-destination pairs in a way that favors the road authority, then the problem to be solved by the authority is the so-called optimistic bilevel problem to

$$\begin{aligned} &\text{minimize} \quad F(v,\tau) \\ &\text{subject to} \quad \tau \in \Gamma, v \in S(\tau), \end{aligned} \tag{1.5}$$

where $F(v,\tau)$ is the disutility function of the road authority who is also called the *leader*, Γ is a closed set representing the set of feasible tolls, and for any given toll $\tau \in \Gamma$, $S(\tau)$ denotes the set of optimal link flows for the collection of all the road users also called the *follower*. In other words, $S(\tau)$ is the solution set of the traffic assignment problem (1.4) under the toll pattern τ. As was mentioned by Dempe and Zemkoho (2008), the model in (1.5) can be altered to tackle other hierarchical problems like the reduction of road accidents in some developing countries. The modification consist of separating the road users in two categories: the heavy goods vehicles and the rest of the users. Hence, only the heavy goods vehicles may be charged a toll equivalent to the level of risk to which they expose the other road users, by using the corresponding link. In the same way, the problem of reducing the level of pollution caused by heavy goods vehicles on some links of a network can be addressed. Many other economical or traffic improvements goals can be achieved by road pricing. A major deciding factor is the leader or road authority's objective function. Various expressions of $F(v,\tau)$ have been considered in the literature, including that of minimizing the total travel time experienced by all vehicles

$$F(v,\tau) = \sum_{a \in \mathscr{A}} v_a t_a(v_a, \tau_a). \tag{1.6}$$

The total revenue arising from toll charges can also be maximized, hence the authority's cost function takes the form

$$F(v,\tau)=\sum_{a\in\mathscr{A}} v_a\tau_a. \tag{1.7}$$

A combination of objectives (1.6) and (1.7) is also possible through a weighted sum or the maximization of the ratio of the total revenue to the total cost, that is the function:

$$F(v,\tau)=\sum_{a\in\mathscr{A}} v_a\tau_a \Big/ \sum_{a\in\mathscr{A}} v_a t_a(v_a,\tau_a). \tag{1.8}$$

To solve problem (1.5), two approaches have mainly been used:

The sensitivity analysis which essentially consist of computing the derivative or generalized derivative, when it exists, of the link flow function $v(.)$, after ensuring that it is locally well-defined. Some approximation techniques, like the Taylor expansion theorem, are then used to obtain a local approximation of $v(.)$ which is inserted in the upper level function $F(v,\tau)$, such that the problem to be solved is

$$\begin{aligned} &\text{minimize} \quad F(v(\tau),\tau)\\ &\text{subject to} \quad \tau\in\Gamma. \end{aligned} \tag{1.9}$$

For an extensive review on sensitivity analysis methods used in the literature to obtain the derivative of the link flow $v(.)$ of the traffic assignment problem (1.4), we refer the interested reader to the paper of Lu (2008). Applications of sensitivity analysis techniques to solution methods for bilevel transportation problems include the works of Chiou (2005), Friesz et al. (1990), and Josefsson and Patriksson (2007) for network designing; Yang et al. (1992), and Yang (1995) for the O-D matrix estimation; and Yang and Lam (1996), Yang and Yagar (1994), Yang and Bell (1997), and Patriksson and Rockaffelar (2002) for the road pricing problem. It may be important to mention that Robinson (2006) discovered an error in one of the main results used in the sensitivity analysis technique suggested by Patriksson (2004) and that was applied in Josefsson and Patriksson (2007). The interested reader is referred to Dempe (1993), Ralph and Dempe (1995), and Dempe and Vogel (2001) for sensitivity analysis in more general optimization problems; and to Dempe and Schmidt (1996) for application of a sensitivity analysis technique in general bilevel programming problems. A major draw back of most of the sensitivity analysis approaches are the strong assumptions required in order to have the local uniqueness of the optimal solution of the traffic assignment problem. For instance, it is usually required that for each $a\in\mathscr{A}$, the link cost function t_a is strictly increasing in the first argument; which makes this approach not applicable in the simple case where the lower level disutility function $f(v,\tau)$ is linear or bilinear. Secondly, Netter (1972) proved that in reality it is not always possible to have uniqueness of the optimal flows in the context of multiclass users like for the road accident reduction problem mentioned earlier in this section.

The KKT reformulation which consist in replacing the lower level problem (1.4) represented in (1.5) through the solution set $S(\tau)$ by its Karush-Kuhn-Tucker (KKT) conditions. This approach has been used by Labbée et al. (1998), Dewez et al. (2008), Heilporn et al. (2010), etc., in order to solve the bilevel road pricing problem. In most of these papers, the leader's cost function takes the bilinear form (1.7) and after the KKT reformulation a new transformation is made, to obtain a 0-1 mixed integer programming problem, and heuristics are then developed to solve the latter problem. It may also be of interest to remind that in the latter works the *link-node* model is considered for the traffic assignment problem instead of the *link-route* formulation used in (1.4). For more details on the link-node and link-route formulations of the traffic assignment problem and relations between them, we refer the interested reader to the book of Patriksson (1994). The main difficulty faced by this approach is that the KKT reformulation of the bilevel road pricing problem (1.5) may not be equivalent to the initial problem when local solutions are considered; see Dempe and Dutta (2010). Hence, considering the fact that problem (1.5) is a typical nonconvex optimization problem, as to be discussed in the next section, the idea of computing global optimal solutions is hypothetical.

In the next section, we consider the *optimal value reformulation* of the bilevel road pricing problem which has been given no attention, to the best of our knowledge, although it potentially helps to avoid the main difficulties raised by the first two approaches.

2 The optimal value function approach

The optimal value function of the traffic assignment problem (1.4) is given as

$$\varphi(\tau) := \min\{f(v,\tau)|v \in V\}. \tag{2.1}$$

It can easily be shown that the bilevel road pricing problem (1.5) is globally and locally equivalent to the following problem called the optimal value reformulation:

$$\begin{aligned} &\text{minimize} \quad F(v,\tau)\\ &\text{subject to} \quad \begin{cases} f(v,\tau) \le \varphi(\tau),\\ \tau \in \Gamma,\ v \in V. \end{cases} \end{aligned} \tag{2.2}$$

In contrary to the sensitivity analysis model in the previous section, no requirement of uniqueness of the optimal flows for the traffic assignment problem (1.4) is needed. Another advantage of the optimal value reformulation as compared to the KKT reformulation is that in the process of deriving optimality conditions, the latter reformulation usually requires second order derivatives for the functions involved in the lower level problem. Not only second order derivatives are already demanding for a solution process but this may be even more harmful in the context of problem (1.5) since some essential data could disappear from the optimality conditions, given that the constraints of the traffic assignment problem (1.4) are linear and one may also have situations where there is no congestion in the network, cf. Dempe and Zemkoho (2011a).

The optimal value reformulation (2.2) was introduced in bilevel programming by Outrata (1990) and to the best of our knowledge it was first used in the field of bilevel transportation by Chen (1994). Chen (1994), and Chen and Florian (1998) derived Fritz-John's type optimality conditions for the O-D demand adjustment problem in the reformulation analogous to (2.2). An augmented Lagrangean method was then suggested to solve the latter problem, but with the possibility that the multiplier attached to the leader's cost function could vanish, the Fritz-John's type conditions may not be indicated for this algorithm to work efficiently. For this reason, we will show later in this paper, how KKT type optimality conditions could be derived for the O-D matrix estimation problem. Meng et al. (2001) also used the optimal value function approach to tackle the network design problem. But some strong assumptions were made, including that all the link costs t_a, $a \in \mathscr{A}$ are strictly increasing w.r.t. v_a, in order to obtain the differentiability of the optimal value function. As already mentioned above, such a requirement is not applicable in the case where f is bilinear, that is when there is no congestion in the network. An augmented Lagrangean method was also proposed for this problem, but with different stopping criteria. It may be of interest to remind that, mathematically, the bilevel network design and the road pricing problem have the same structure: the lower level problem is parameterized only in the route users total cost function represented by f in our case. At the difference that instead of the toll as in the road pricing problem, the parameter represents the capacity enhancement for the network design problem.

For the rest of the paper, unless otherwise stated, the leader's cost function F and the route users' total cost function f are assumed to be continuously differentiable. One of the major challenges for the optimal value reformulation (2.2) of our bilevel road pricing problem (1.5) is the presence of the optimal value function φ. This function is typically a nonsmooth function. In fact, even in the simplest case where the follower's cost function takes the bilinear form $f(v,\tau)=v^\top\tau$, φ is a piecewise linear function. Secondly, the constraint function $(v,\tau)\to f(v,\tau)-\varphi(\tau)$ may not be convex, even in the situation where f would be convex in (v,τ), in which case we would have a difference of convex functions. Nevertheless we will show in Theorem 2.1 below that φ is locally Lipschitz continuous. Thus problem (2.2) can be considered as a Lipschitz optimization problem with a special constraint $f(v,\tau)-\varphi(v)\leq 0$, called the optimal value constraint. For the investigation of such a problem we need some nonsmooth tools. In this paper, we use the normal cone and subdifferential of Mordukhovich, which are defined respectively as:

$$N_C(\overline{x}) := \{u\in\mathbb{R}^n | \exists u_k\to u, x_k\to\overline{x}(x_k\in C): u_k\in\widehat{N}_C(x_k)\},$$
$$\partial\psi(\overline{x}) := \{u\in\mathbb{R}^n | (u,-1)\in N_{\mathrm{epi}\psi}(\overline{x},\psi(\overline{x}))\},$$

where $\overline{x}\in C$ (closed set) and $\overline{x}\in\mathrm{dom}\,\psi$ ($\psi:\mathbb{R}^n\to\overline{\mathbb{R}}$), respectively. Furthermore, $\widehat{N}_C$ and $\mathrm{epi}\,\psi$ denote the Fréchet normal cone to C and the epigraph of ψ, respectively. It is worth mentioning that $\partial\psi(\overline{x})$ is nonempty and compact when ψ is Lipschitz continuous around $\overline{x}$ and further we have $\partial\psi(\overline{x})=\{\nabla\psi(\overline{x})\}$ when ψ is continuously differentiable. Also let us mention that $\widehat{N}_C(\overline{x})$ and $N_C(\overline{x})$ both coincide with the ordinary normal cone in the sense of convex analysis provided that the set C is convex. In the case where ϕ and ψ are locally Lipschitz continuous around $\overline{x}$, the sum rule

$$\partial(\lambda\phi+\mu\psi)(\overline{x})\subseteq\lambda\partial\phi(\overline{x})+\mu\partial\psi(\overline{x}), \tag{2.3}$$

with λ and μ being nonnegative real numbers, and the convex hull property

$$\mathrm{co}\partial(-\psi)(\overline{x}) = -\mathrm{co}\partial\psi(\overline{x}), \tag{2.4}$$

where $\mathrm{co}A$ represents the convex hull of A, will also be of great utility in the sequel. For more details on these tools, the interested reader is referred to the books of Rockafellar and Wets (1998) and Mordukhovich (2006).

From now on, it will also be important to notice that the set of feasible link flows can be written as the image of the set of route flows Q via the linear application $q\to\Delta q$, that is in other words $V:=\Delta Q$.

Theorem 2.1 (Sensitivity analysis of the traffic assignment value function for the road pricing problem) *The optimal value function φ (2.1) is Lipschitz continuous around any $\overline{\tau}\in\Gamma$, and its Mordukhovich subdifferential is obtained as*

$$\partial\varphi(\overline{\tau})\subseteq\{\nabla_\tau t(\overline{v},\overline{\tau}) | \overline{v}\in S(\overline{\tau})\}, \tag{2.5}$$

where $\nabla_\tau t(\overline{v},\overline{\tau}) := [\int_0^{\overline{v}_a} \frac{\partial t_a}{\partial \tau_a}(s,\overline{\tau})ds]_{a\in\mathscr{A}}$.

Proof We have $Q \subseteq |c|\mathbb{B}$, where $\mathbb{B}$ is the unit ball of $\mathbb{R}^\pi$ and $|c| := \max\{c_i | i = 1,\dots,\pi\}$ (c is the route capacity vector). Hence $V := \Delta Q \subseteq \Delta(|c|\mathbb{B})$ is a closed and bounded set given that $q \to \Delta q$ is a continuous function and Q is also a closed set. In addition to the continuous differentiability of f, it follows from Mordukhovich (2006) that inclusion (2.5) holds true and from Mordukhovich and Nam (2005) that φ is locally Lipschitz continuous. □

Theorem 2.2 (Existence of solution for the bilevel road pricing problem) *Problem* (2.2) *has at least one optimal solution provided that the set of feasible tolls Γ is bounded.*

Proof Since Γ is closed by definition, the feasible set of problem (2.2) is compact, taking into account the compactness of V (see Proof of Theorem 2.1). Moreover, considering the locally Lipschitz continuity of the optimal value function φ (2.1) (cf. Theorem 2.1), the result follows from the famous Weierstraß theorem. □

At this level, it is clear that the bilevel road pricing problem faces many challenges including the nonsmoothness and the nonconvexity as already illustrated. We now add to this list the fact that most of the well-known constraint qualifications (CQs) fail for problem (2.2). For more details on this issue, the interested reader is referred to Dempe and Zemkoho (2011b) and references therein. Nonetheless, it is worth to mention that the failure of the underlined CQs for the optimal value reformulation (2.2) of the bilevel road pricing problem (1.5) is due to the optimal value constraint $f(v,\tau) \leq \varphi(\tau)$. In the next section we will introduce the partial calmness, a CQ that will help move this constraint to the authority's cost function F. Hence it will then be easier to derive KKT type optimality conditions for the bilevel road pricing problem.

3 Optimality conditions

In order to ease the presentation of optimality conditions for the bilevel road pricing problem (1.5), the estimation of the normal cone to the joined upper and lower level feasible set is necessary, i.e. we have to compute $N_{V\times\Gamma}$. From now on we consider the set of feasible tolls as

$$\Gamma := \{\tau \in \mathbb{R}^\alpha | \varsigma \leq \tau \leq \kappa\}, \tag{3.1}$$

where $\varsigma, \kappa \in \mathbb{R}^\alpha$ represent the minimum and the maximum tolls, respectively. Consider a feasible point $(\overline{v},\overline{\tau})$ to problem (2.2), then we have

$$N_{V\times\Gamma}(\overline{v},\overline{\tau}) = N_V(\overline{v}) \times N_\Gamma(\overline{\tau}). \tag{3.2}$$

Now let us denote by

$$\mathscr{A}^\varsigma := \{a \in \mathscr{A} | \overline{\tau}_a = \varsigma_a\}$$

the set of all the links of the network having the minimum toll. We remind that since the aim of the toll setting is to encourage road users to utilize some abandoned or less utilized routes, then we may have $\varsigma_a = 0$, for some $a \in \mathscr{A}$. Hence to correct the unfairness in our model, we may assume that

$$\emptyset \neq \mathscr{A}^o := \{a \in \mathscr{A} | \overline{\tau}_a = \varsigma_a = 0\} \subset \mathscr{A}^\varsigma;$$

thus, allowing some links to be toll-free. We further define the set

$$\mathscr{A}^{\kappa} := \{a \in \mathscr{A} | \overline{\tau}_a = \kappa_a\}$$

of links with maximum tolls. It is worth mentioning that the restriction that the tolls should not exceed some certain amount is of great importance for social considerations since the road users and the community in general should not have the feeling that the road authority just intends to make as much money as possible. Finally, let

$$\mathscr{A}^{\gamma} := \{a \in \mathscr{A} | \varsigma_a < \overline{\tau}_a < \kappa_a\};$$

then, $\mathscr{A}^{\varsigma}$, $\mathscr{A}^{\kappa}$ and $\mathscr{A}^{\gamma}$ form a partition of $\mathscr{A}$. Thus, $\mathscr{A} = \mathscr{A}^{\varsigma} \cup \mathscr{A}^{\kappa} \cup \mathscr{A}^{\gamma}$.

To make the further explanations more clear, we make the following technical assumption: We assume that $\mathscr{A}$ is an ordered set; hence, each link $a \in \mathscr{A}$ is associated with an index $|a| \in \mathbb{N}$ and we define the α-dimensional vector

$$e^a := (0, \dots, 0, 1, 0, \dots, 0)^{\top}, \tag{3.3}$$

where 1 is at position $|a|$, in order to symbolize the utilization of the corresponding link by a road user. Then it follows from Rockafellar and Wets (1998, Theorem 6.46) that

$$N_{\Gamma}(\overline{\tau}) = \left\{ -\sum_{a \in \mathscr{A}^{\varsigma}} \lambda_a^{\varsigma} e^a + \sum_{a \in \mathscr{A}^{\kappa}} \lambda_a^{\kappa} e^a \middle| (\lambda_a^{\varsigma})_{a \in \mathscr{A}^{\varsigma}} \geq 0, (\lambda_a^{\kappa})_{a \in \mathscr{A}^{\kappa}} \geq 0 \right\}. \tag{3.4}$$

Next, we compute $N_V(\overline{v})$; for this, it may be preferable to first compute $N_Q(\overline{q})$, where $\overline{q}$ is a feasible route flow verifying $\overline{v} = \Delta \overline{q}$. We recall that Q is defined as in (1.1). We also assume that $\mathscr{P}$ is an ordered set such that for a route $r \in \mathscr{P}$, we associate an index $|r| \in \mathbb{N}$ and we define the π-dimensional vector e^r as in (3.3). We consider the set

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

of unused routes of the network and the set

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

of routes used at their full capacity. Then $\mathscr{P}$ can be partitioned into $\mathscr{P}^0$, $\mathscr{P}^c$ and $\mathscr{P}^u$, where $\mathscr{P}^u$ is the set of routes used but which are not at full capacity. Thus, $\mathscr{P} = \mathscr{P}^o \cup \mathscr{P}^c \cup \mathscr{P}^u$.

We now consider the collection $[\Lambda_w]_{w \in \mathscr{W}}$ of rows of the O-D-route incidence matrix Λ. Then, we have the equality

$$N_Q(\overline{q}) = \left\{ \sum_{r \in \mathscr{P}^c} \lambda_r^c e^r - \sum_{r \in \mathscr{P}^o} \lambda_r^o e^r + \sum_{w \in \mathscr{W}} \lambda_w \Lambda_w^{\top} \middle| (\lambda_r^o)_{r \in \mathscr{P}^o} \geq 0, (\lambda_r^c)_{r \in \mathscr{P}^c} \geq 0 \right\}, \tag{3.5}$$

from Rockafellar and Wets (1998, Theorem 6.46). On the other hand, $V = \Delta Q$, with $q \to \Delta q$ being a linear function and Q a convex set. Hence, it follows from Rockafellar and Wets (1998, Theorem 6.43) that

$$N_V(\overline{v}) = \left\{ u \in \mathbb{R}^\alpha \,\middle|\, \Delta^\top u = \sum_{r\in\mathscr{P}^c} \lambda_r^c e^r - \sum_{r\in\mathscr{P}^o} \lambda_r^o e^r + \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top, \right.$$
$$\left. (\lambda_r^o)_{r\in\mathscr{P}^o} \geq 0, (\lambda_r^c)_{r\in\mathscr{P}^c} \geq 0 \right\}. \tag{3.6}$$

The computation of the normal cone $N_{V\times\Gamma}(\overline{v},\overline{\tau})$ can be summarized in the following result while considering equality (3.2).

Lemma 3.1 (Normal cone to the joined feasible set of the bilevel road pricing problem) $(v^*,\tau^*) \in N_{V\times\Gamma}(\overline{v},\overline{\tau})$ *if and only if there exists* $\overline{q} \in \mathbb{R}^\pi$, *with* $0 \leq \overline{q} \leq c$, $\Lambda\overline{q} = d$ *and* $\Delta\overline{q} = \overline{v}$ *such that* $(\Delta^\top v^*, \tau^*) \in N_Q(\overline{q}) \times N_\Gamma(\overline{\tau})$.

As already mentioned in the previous section, we will use the partial calmness condition, as a CQ to derive KKT type necessary optimality conditions for the bilevel road pricing problem. This condition was introduced by Ye and Zhu (1995).

Definition 3.1 (The partial calmness concept) Problem (2.2) will be said to be *partially calm* at one of its optimal solutions $(\overline{v},\overline{\tau})$ if and only if there exists $\mu > 0$ such that $(\overline{v},\overline{\tau})$ is an optimal solution to the partially penalized problem to

$$\begin{aligned} &\text{minimize} \quad F(v,\tau) + \mu(f(v,\tau) - \varphi(\tau)) \\ &\text{subject to} \quad \tau \in \Gamma, \quad v \in V. \end{aligned} \tag{3.7}$$

Theorem 3.1 (Necessary optimality conditions for the bilevel road pricing problem) *Let* $(\overline{v},\overline{\tau})$ *be a local optimal solution to problem* (2.2), *which is assumed to be partially calm at* $(\overline{v},\overline{\tau})$. *Then there exist* $\mu > 0$, $(\lambda^\varsigma, \lambda^\kappa, \lambda^o, \lambda^c, \lambda^\omega)$, *and* $\overline{q} \in \mathbb{R}^\pi$, $v_s \in S(\overline{\tau})$, $\eta_s \geq 0, s = 1,\dots,\alpha+1$ *with* $\sum_{s=1}^{\alpha+1} \eta_s = 1$ *such that*

$$\nabla_\tau F(\overline{v},\overline{\tau}) + \mu\nabla_\tau t(\overline{v},\overline{\tau}) - \mu\sum_{s=1}^{n+1} \eta_s \nabla_\tau t(v_s,\overline{\tau}) = \sum_{a\in\mathscr{A}^\varsigma} \lambda_a^\varsigma e^a - \sum_{a\in\mathscr{A}^\kappa} \lambda_a^\kappa e^a, \tag{3.8}$$

$$\Delta^\top(\nabla_v F(\overline{v},\overline{\tau}) + \mu t(\overline{v},\overline{\tau})) = \sum_{r\in\mathscr{P}^o} \lambda_r^o e^r - \sum_{r\in\mathscr{P}^c} \lambda_r^c e^r - \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top, \tag{3.9}$$

$$0 \leq \overline{q} \leq c, \quad \Lambda\overline{q} = d, \quad \Delta\overline{q} = \overline{v}, \tag{3.10}$$

$$\lambda^\varsigma = (\lambda_a^\varsigma) \geq 0, \quad \lambda^\kappa = (\lambda_a^\kappa) \geq 0, \quad \lambda^o = (\lambda_r^o) \geq 0, \quad \lambda^c = (\lambda_r^c) \geq 0, \quad \lambda^\omega = (\lambda_w). \tag{3.11}$$

Proof Since problem (2.2) is partially calm at $(\overline{v},\overline{\tau})$, there exists $\mu > 0$ such that $(\overline{v},\overline{\tau})$ solves problem (3.7). Since F and f are continuously differentiable and φ is locally Lipschitz continuous (see Theorem 2.1) then it follows from Mordukhovich (2006, Proposition 5.3) that

$$0 \in \partial\big(F + \mu(f - \varphi)\big)(\overline{v},\overline{\tau}) + N_{V\times\Gamma}(\overline{v},\overline{\tau}).$$

Hence, from the sum rule (2.3) and the convex hull property (2.4), there exists $(v^*,\tau^*) \in N_{V\times\Gamma}(\overline{v},\overline{\tau})$ such that

$$\nabla F(\overline{v},\overline{\tau}) + \mu\nabla f(\overline{v},\overline{\tau}) + (v^*,\tau^*) \in \{0\} \times \mu\,\mathrm{co}\,\partial\varphi(\overline{\tau}). \tag{3.12}$$

Applying Lemma 3.1, we can also find $\tau^* \in \operatorname{co}\partial\varphi(\overline{\tau})$ and $(\lambda^\varsigma, \lambda^\kappa, \lambda^o, \lambda^c, \lambda^\omega)$, $\overline{q} \in \mathbb{R}^\pi$ (cf. (3.4) and (3.5)) satisfying (3.10)–(3.11) such that

$$\nabla_v F(\overline{v}, \overline{\tau}) + \mu \nabla_v f(\overline{v}, \overline{\tau}) + v^* = 0, \tag{3.13}$$

$$\Delta^\top v^* = \sum_{r \in \mathscr{P}^c} \lambda_r^c e^r - \sum_{r \in \mathscr{P}^o} \lambda_r^o e^r + \sum_{w \in \mathscr{W}} \lambda_w \Lambda_w^\top, \tag{3.14}$$

$$\nabla_\tau F(\overline{v}, \overline{\tau}) + \mu \nabla_\tau f(\overline{v}, \overline{\tau}) - \sum_{a \in \mathscr{A}^\varsigma} \lambda_a^\varsigma e^a + \sum_{a \in \mathscr{A}^\kappa} \lambda_a^\kappa e^a = \mu \tau^*. \tag{3.15}$$

Applying Caratheodory's theorem (see for example Rockafellar and Wets 1998), $\tau^* \in \operatorname{co}\partial\varphi(\overline{\tau})$ implies the existence of $v_s \in S(\overline{\tau})$ (cf. Theorem 2.1) and $\eta_s \geq 0$, $s = 1, \ldots, \alpha + 1$ with $\sum_{s=1}^{\alpha+1} \eta_s = 1$ such that

$$\tau^* = \sum_{s=1}^{\alpha+1} \eta_s \nabla_\tau f(v_s, \overline{\tau}).$$

Combining (3.13)–(3.15) and the latter equality, we have the result. □

Next, we derive KKT conditions for problem (1.5) without the convex combination on the Mordukhovich subdifferential of the value function (2.1).

Corollary 3.1 (Optimality conditions without the convex combination) *Let* $(\overline{v}, \overline{\tau})$ *be a local optimal solution to problem* (2.2), *which is assumed to be partially calm at* $(\overline{v}, \overline{\tau})$, *with* $S(\overline{\tau}) = \{\overline{v}\}$. *Then there exist* $\mu > 0$, $(\lambda^\varsigma, \lambda^\kappa, \lambda^o, \lambda^c, \lambda^\omega)$, *and* $\overline{q} \in \mathbb{R}^\pi$ *such that relationships* (3.9)–(3.11), *together with the following condition are satisfied*:

$$\nabla_\tau F(\overline{v}, \overline{\tau}) = \sum_{a \in \mathscr{A}^\varsigma} \lambda_a^\varsigma e^a - \sum_{a \in \mathscr{A}^\kappa} \lambda_a^\kappa e^a.$$

Proof It from the proof of the previous theorem by noting that with $S(\overline{\tau}) = \{\overline{v}\}$, we have from Theorem 2.1 that

$$\operatorname{co}\partial\varphi(\overline{\tau}) = \{\nabla_\tau f(\overline{v}, \overline{\tau})\}$$

which implies the result by substituting the latter expression of $\operatorname{co}\partial\varphi(\overline{\tau})$ in (3.12). □

It should be clear that the condition $S(\overline{\tau}) = \{\overline{v}\}$ imposed in this corollary is far away from the usual strong assumptions made in the sensitivity analysis approaches mentioned in the Introduction. In fact, in the latter cases, it is usually required that the traffic assignment problem admits a unique optimal solution in a certain neighborhood. In particular, the strict monotonicity of the link costs $t_a (a \in \mathscr{A})$, often needed is not satisfied in the framework of Theorem 3.2 below.

For the next result, we assume that the cost function of the traffic assignment problem is bilinear, i.e. $f(v, \tau) = v^\top \tau$ and the set of feasible tolls coincides with the whole space, i.e. $\Gamma = \mathbb{R}^\alpha$. This definition of f corresponds to the ideal case, where there is no congestion in the network. This framework has been considered by many authors; see for example Labbée et al. (1998), Dewez et al. (2008) and Heilporn et al. (2010).

Theorem 3.2 (Optimality conditions in the case of no congestion) *Let* $(\overline{v}, \overline{\tau})$ *be a local optimal solution to problem* (2.2), *then there exist* $\widetilde{v} \in S(\overline{\tau})$, $\mu > 0$, $(\lambda^o, \lambda^c, \lambda^\omega)$ *and* $\overline{q} \in \mathbb{R}^\pi$ *such that* (3.10), *together with the following conditions are satisfied*:

$$\nabla_\tau F(\overline{v}, \overline{\tau}) + \mu(\overline{v} - \widetilde{v}) = 0, \tag{3.16}$$

$$\Delta^\top(\nabla_v F(\overline{v}, \overline{\tau}) + \mu\overline{\tau}) = \sum_{r \in \mathscr{P}^o} \lambda_r^o e^r - \sum_{r \in \mathscr{P}^c} \lambda_r^c e^r - \sum_{w \in \mathscr{W}} \lambda_w \Lambda_w^\top, \tag{3.17}$$

$$\lambda^o = (\lambda_r^o) \geq 0, \qquad \lambda^c = (\lambda_r^c) \geq 0, \lambda^\omega = (\lambda_w). \tag{3.18}$$

Proof We recall from Dempe and Zemkoho (2011a, Theorem 4.2) that with $f(v, \tau) = v^\top \tau$ and $\Gamma = \mathbb{R}^\alpha$, problem (2.2) is partially calm at $(\overline{v}, \overline{\tau})$. Hence, from the proof of Theorem 3.1, we have that there exists $(v^*, \tau^*) \in N_{V \times \Gamma}(\overline{v}, \overline{\tau})$ such that inclusion (3.12) holds. Considering inclusion (2.5) of Theorem 2.1, we have

$$\partial\varphi(\overline{\tau}) \subseteq S(\overline{\tau}) = \mathrm{co}\,\partial S(\overline{\tau}) \tag{3.19}$$

given that $S(\overline{\tau})$ is convex in this case. The result then follows by applying Lemma 3.1 to (3.12) while considering the last equality of (3.19) and the fact that

$$N_{V \times \Gamma}(\overline{v}, \overline{\tau}) = N_V(\overline{v}) \times \{0\}$$

since $\Gamma = \mathbb{R}^\alpha$. □

This result can easily be extended to the case where $f(v, \tau) = u(\tau)^\top v$, with $u : \mathbb{R}^\alpha \to \mathbb{R}^\alpha$, considering Dempe and Zemkoho (2011b, Theorem 4.2) insuring the partial calmness of problem (1.5) in this situation at every local optimal solution.

If $S(\overline{\tau}) = \{\overline{v}\}$, it follows from (3.16) that $\nabla_\tau F(\overline{v}, \overline{\tau}) = 0$ which is a natural optimality condition for the road authority's problem to

$$\text{minimize } F(\overline{v}, \tau) \quad \text{subject to } \tau \in \mathbb{R}^\alpha$$

provided that $\overline{v}$ is the unique optimal solution of the traffic assignment problem (1.4) where the toll is fixed at $\tau = \overline{\tau}$ satisfying (3.10), (3.17) and (3.18). This observation could be of a great utility in designing an algorithm for the bilevel road pricing problem in the case where the total road cost is bilinear.

To conclude this section, we give a sufficient condition for problem (2.2) to be partially calm when the total road cost takes a more general form as illustrated in the traffic assignment problem (1.4).

Theorem 3.3 (A sufficient condition ensuring the partial calmness) *Let $(\overline{v}, \overline{\tau})$ be a feasible point to (2.2). We assume that for all $a \in \mathscr{A}$, t_a is increasing with respect to its first argument. Then problem (2.2) is partially calm at $(\overline{v}, \overline{\tau})$ provided that there exists $\alpha > 0$ such that*

$$\sum_{a \in \mathscr{A}} t_a(v_a, \tau) h_a \geq \alpha \|h\|, \quad \forall v \in S(\tau),\ h \in T_V(v) \cap N_{S(\tau)}(v),\ \tau \in \Gamma. \tag{3.20}$$

Proof Since for all $a \in \mathscr{A}$, t_a is increasing with respect to its first argument, then the function f is convex with respect to the link flow v. In addition to the convexity of V, it follows from Ye (1998, Theorem 3.3) that there exists $\alpha > 0$ such that

$$d_{S(\tau)}(\tau) \leq \alpha^{-1}(f(v, \tau) - \varphi(\tau)), \forall (v, \tau) \in V \times \Gamma,$$

provided that (3.20) holds true. Hence, the result follows from Ye and Zhu (1995). □

For more on the characterization of partial calmness, we refer the interested reader to Dempe and Zemkoho (2010, 2011b), Ye and Zhu (1995) and Ye (1998).

4 Estimation of the O-D matrix

The (fixed) demand vector d needed in the road pricing problem and particularly in the traffic assignment problem (1.4), is a crucial datum given that a good decision process highly depends on how accurate it is estimated. The origin-destination (O-D) matrix estimation or O-D demand adjustment problem (DAP) is important not only for the road pricing problem, but also for many other decision-making frameworks of transportation planing. The modeling of this problem has evolved over the years, see Abrahamsson (1998), and Chen and Florian (1998) for extensive reviews. The bilevel formulation was pioneered by Fisk (1988). Since then, many researchers have adopted this model which usually takes the form:

$$\begin{aligned} &\text{minimize} \quad F(d,v) \\ &\text{subject to} \quad d \in D, \ v \in S(d), \end{aligned} \tag{4.1}$$

where $D \subseteq \mathbb{R}^\omega$ is a closed set and $S(d)$ is the solution set of the traffic assignment problem

$$\begin{aligned} &\text{minimize} \quad f(d,v) := \sum_{a\in\mathscr{A}} \int_0^{v_a} t_a(s)ds \\ &\text{subject to} \quad v \in V(d), \end{aligned} \tag{4.2}$$

parameterized by d, also called O-D demand and representing the O-D matrix organized as a vector. The set-valued mappings

$$V(d) := \{v \in \mathbb{R}^\alpha | \exists q \in Q(d), \varDelta q = v\} \quad \text{and} \quad Q(d) := \{q \in \mathbb{R}^\pi_+ | q \le c, \varLambda q = d\}$$

denote the set of feasible link flows and feasible route flows, respectively, for a given demand vector d. The upper level objective function F is usually of the form

$$F(d,v) := \gamma_1 F_1(d,\hat{d}) + \gamma_2 F_2(v,\hat{v}),$$

where $\hat{d}$ represents the target O-D matrix that may be obtained from sample surveys, and $\hat{v}$ denotes the vector of flows observed on some links of the network. The function $F_1(d,\hat{d})$ represents the error measurement between the target O-D matrix $\hat{d}$ and the estimated matrix d, while $F_2(v,\hat{v})$ denotes the error measurement between the observed link flow $\hat{v}$ and the estimated flow v. The parameters γ_1 and γ_2 represent the uncertainty in the information contained in $\hat{d}$ and $\hat{v}$, respectively. As for the total road cost function f, the expression in (4.2) is mainly considered for illustrative purpose, since many other cost function models exist in the literature, see Patriksson (1994) for details. It is however important to mention that the function f in (4.2) is a convex function in (d,v), which appears to be an important property for most of the results in this section. In the line of Migdalas (1995), the set D can be considered analogously to $\varGamma$ (3.1).

As mentioned in the Introduction of the paper, the sensitivity analysis has also been used to tackle the O-D matrix estimation problem. We refer the interested reader to the papers of Abrahamsson (1998), Codina and Montero (2006), Lundgren and Peterson (2008), and Noriega and Florian (2009) for various methodological approaches in solving the problem. Problem (4.1) can be reformulated as

$$\begin{aligned} &\text{minimize} \quad F(d,v) \\ &\text{subject to} \quad \begin{cases} f(d,v) - \varphi(d) \le 0, \\ d \in D, v \in V(d), \end{cases} \end{aligned} \tag{4.3}$$

with φ being the optimal value function of the traffic assignment problem (4.2). Chen (1994) and Chen and Florian (1998) considered $D := \{d \in \mathbb{R}^\pi \,|\, d \geq 0\}$ and did not impose capacities on the route flows. For further simplification, they considered the constraint $v = \Delta q$ as exogenous, hence the simplified problem:

$$\begin{aligned} &\text{minimize} \quad F(d, v) \\ &\text{subject to} \quad \begin{cases} f(d, v) - \varphi(d) \leq 0 \\ d, q \geq 0, \;\; \Lambda q = d. \end{cases} \end{aligned} \tag{4.4}$$

Fritz John's type optimality conditions were then derived for (4.4). Our aim here is to suggest Karush-Kuhn-Tucker (KKT) type optimality conditions for the more general problem with the flow conservation constraint $v = \Delta q$ being fully part of the feasible set of the traffic assignment problem (4.2).

In order to write the optimality condition of (4.3) in a detailed form, we should be able to compute or at least give an upper estimation of the Mordukhovich normal cone $N_{gph V}(\overline{d}, \overline{v})$ (recall that for a set-valued mapping M, $(x, y) \in gph\, M$ if and only if $y \in M(x)$) and subdifferential $\partial\varphi(\overline{d})$, respectively. That is exactly what is done in Lemma 4.1 and Lemma 4.2, respectively. These results were established by Dempe and Zemkoho (2010). We will not present the proofs here since they require some sophisticated mathematical tools that we do not intend to present here, notably the coderivative of Mordukhovich and some related calculus rules. Let us just mention that these proofs fully take advantage of the structure of the feasible set of the traffic assignment problem, that is, the fact that for a given demand d, the set of feasible link flows $V(d)$ is obtained as the image of the set of feasible route flows $Q(d)$ via the mapping $q \to \Delta q$.

Lemma 4.1 (Normal cone to the graph of V) *For any $(\overline{d}, \overline{v}) \in gph\, V$, we have*

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

where $\mathscr{H}(\overline{d}, \overline{v})$ and $N_{gphQ}(\overline{d}, \overline{q})$ are given respectively as:

$$\begin{cases} \mathscr{H}(\overline{d}, \overline{v}) := \{q \in \mathbb{R}^\pi \,|\, \Delta q = \overline{v}, (\overline{d}, q) \in gphQ\}, \\ N_{gphQ}(\overline{d}, \overline{q}) = \big\{\big(-\sum w \in \mathscr{W} \lambda_w e^w, \sum r \in \mathscr{P}^c(\overline{q}) \lambda_r^c e^r \\ \qquad - \sum r \in \mathscr{P}^o(\overline{q}) \lambda_r^o e^r + \sum w \in \mathscr{W} \lambda_w \Lambda_w^\top\big) : \\ \qquad (\lambda_w)_{w \in \mathscr{W}} \in \mathbb{R}^\omega, \; (\lambda_r^o)_{r \in \mathscr{P}^o(\overline{q})}, (\lambda_r^c)_{r \in \mathscr{P}^c(\overline{q})} \geq 0\big\}. \end{cases} \tag{4.5}$$

Here, e^r and Λ_w are defined as in the previous section, while e^w is the analog of e^a (cf. previous section) and $\mathscr{H}(\overline{d}, \overline{v})$ denotes the set of route flows corresponding to the feasible demand-link flow couple $(\overline{d}, \overline{v})$. Next, we provide an estimate for the subdifferential of the optimal value function of the traffic assignment problem.

Lemma 4.2 (Sensitivity analysis of the traffic assignment value function of the DAP) *Assume that the total road cost function f is convex in (d, v). Then, for every $(\overline{d}, \overline{v}) \in gph\Psi$, the optimal value function φ of the traffic assignment problem* (4.2) *is Lipschitz continuous around $\overline{d}$ and*

$$\partial\varphi(\overline{d}) \subseteq \bigcup_{\overline{q} \in \mathscr{H}(\overline{d}, \overline{v})} \; \bigcup_{(\lambda^\omega, \lambda^c, \lambda^0) \in \Lambda(\overline{d}, \overline{q})} \Big\{-\sum_{w \in \mathscr{W}} \lambda_w e_w^\top + \nabla_d f(\overline{d}, \overline{v})\Big\},$$

where $[e_w]_{w\in\mathscr{W}}$ *is the collection of rows of the identity matrix of* $\mathbb{R}^{\omega\times\omega}$, *whereas the set* $\Lambda(\overline{d},\overline{q})$ *of Lagrange multipliers for the traffic assignment problem* (4.2) *is given by*:

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

As already mentioned above, the convexity assumption on the total road cost function f is automatically satisfied for the expression in (4.2).

Before we give the KKT type optimality conditions for the O-D matrix estimation problem, in the classical form, we remind that the characterization of partial calmness given in Theorem 3.1 is a general result for bilevel programming, hence remains true for problem (4.3). We first derive the optimality conditions of the latter problem in the simple case where the road authority does not impose any constraint on the demand, that is, $D:=\mathbb{R}^\omega$.

Theorem 4.1 (Optimality conditions for the DAP without upper level constraints) *Let* $(\overline{d},\overline{v})$ *be a local optimal solution to problem* (4.3), *where* f *is convex in* (d,v) *and* $D:=\mathbb{R}^\omega$. *Assume that the problem is partially calm at* $(\overline{d},\overline{v})$. *Then there exist* $\mu>0$, $\overline{q},\widetilde{q}\in\mathbb{R}^\pi$, $(\lambda^\omega,\lambda^c,\lambda^o)$ *and* $(\widetilde{\lambda}^\omega,\widetilde{\lambda}^c,\widetilde{\lambda}^o)$ *such that*:

$$\nabla_d F(\overline{d},\overline{v})-\sum_{w\in\mathscr{W}}(\lambda_w-\mu\widetilde{\lambda}_w)e_w^\top=0, \tag{4.6}$$

$$\Delta^\top\Big(\nabla_v F(\overline{d},\overline{v})+\mu\nabla_v f(\overline{d},\overline{v})\Big)=\sum_{r\in\mathscr{P}^o(\overline{q})}\lambda_r^o e^r-\sum_{r\in\mathscr{P}^c(\overline{q})}\lambda_r^c e^r-\sum_{w\in\mathscr{W}}\lambda_w\Lambda_w^\top, \tag{4.7}$$

$$\Delta^\top\nabla_v f(\overline{d},\overline{v})=-\sum_{r\in\mathscr{P}^c(\widetilde{q})}\widetilde{\lambda}_r^c e^r+\sum_{r\in\mathscr{P}^o(\widetilde{q})}\widetilde{\lambda}_r^o e^r-\sum_{w\in\mathscr{W}}\widetilde{\lambda}_w\Lambda_w^\top, \tag{4.8}$$

$$0\leq\overline{q}\leq c,\quad \Lambda\overline{q}=\overline{d},\quad \Delta\overline{q}=\overline{v}, \tag{4.9}$$

$$0\leq\widetilde{q}\leq c,\quad \Lambda\widetilde{q}=\overline{d},\quad \Delta\widetilde{q}=\widetilde{v}, \tag{4.10}$$

$$\lambda^\omega=(\lambda_w),\quad \lambda^c=(\lambda_r^c)\geq 0,\quad \lambda^o=(\lambda_r^o)\geq 0, \tag{4.11}$$

$$\widetilde{\lambda}^\omega=(\widetilde{\lambda}_w),\quad \widetilde{\lambda}^c=(\widetilde{\lambda}_r^c)\geq 0,\quad \widetilde{\lambda}^o=(\widetilde{\lambda}_r^o)\geq 0. \tag{4.12}$$

Proof Under the partial calmness of problem (4.3) at $(\overline{d},\overline{v})$, there exists $\mu>0$ such that $(\overline{d},\overline{v})$ solves

$$\begin{array}{ll}\text{minimize} & F(d,v)+\mu(f(d,v)-\varphi(d))\\ \text{subject to} & (d,v)\in\text{gph}\,V.\end{array} \tag{4.13}$$

Applying Mordukhovich (2006, Proposition 5.3) to the latter problem, one gets

$$0\in\nabla F(\overline{d},\overline{v})+\mu\nabla f(\overline{d},\overline{v})+\partial(-\varphi)(\overline{d})\times\{0\}+N_{gphV}(\overline{d},\overline{v}). \tag{4.14}$$

Applying Lemma 4.1, it then follows that there exist $\overline{q}\in\mathbb{R}^\pi$, $(\lambda^\omega,\lambda^c,\lambda^o)$ satisfying (4.9) and (4.11), respectively, and $v^*\in\mathbb{R}^\alpha$ such that:

$$-\sum_{w\in\mathscr{W}}\lambda_w e_w^\top+\nabla_d F(\overline{d},\overline{v})+\mu\nabla_d f(\overline{d},\overline{v})\in\mu\text{co}\partial\varphi(\overline{d}), \tag{4.15}$$

$$\Delta^\top v^* = \sum_{w\in\mathscr{W}} \lambda_w \Lambda_w^\top + \sum_{r\in\mathscr{P}^c(\overline{q})} \lambda_r^c e^r - \sum_{r\in\mathscr{P}^o(\overline{q})} \lambda_r^o e^r, \tag{4.16}$$

$$v^* = -\nabla_v F(\overline{d},\overline{v}) - \mu\nabla_v f(\overline{d},\overline{v}). \tag{4.17}$$

Considering the convexity of the optimal value function φ we have $\mathrm{co}\partial\varphi(\overline{d}) = \partial\varphi(\overline{d})$. Then applying Lemma 4.2 to (4.15), there exist $\widetilde{q}\in\mathbb{R}^\pi$, and $(\widetilde{\lambda}^\omega,\widetilde{\lambda}^c,\widetilde{\lambda}^o)$ satisfying (4.8), (4.10) and (4.12), respectively, such that (4.6) holds. Finally, (4.8) is obtained by inserting (4.17) in equality (4.16). □

In the next theorem, we address the case where the demand vector d is nonnegative, as considered by Chen (1994) (also see Chen and Florian 1998). It appears that the proof of the optimality conditions result becomes a little bit complicate.

Theorem 4.2 (Optimality conditions for the DAP with upper level constraints) *Let* $(\overline{d},\overline{v})$ *be a local optimal solution to problem* (4.3), *where* f *is convex in* (d,v) *and* $D := \{d\in\mathbb{R}^\omega | d \geq 0\}$. *Assume that the problem is partially calm at* $(\overline{d},\overline{v})$. *Then there exist* $\mu>0$, $\overline{q},\widetilde{q}\in\mathbb{R}^\pi$, $(\lambda^\omega,\lambda^c,\lambda^o)$ *and* $(\widetilde{\lambda}^\omega,\widetilde{\lambda}^c,\widetilde{\lambda}^o)$ *such that relationships* (4.7)–(4.12), *together with the following conditions are satisfied*:

$$\nabla_d F(\overline{d},\overline{v}) - \sum_{w\in\mathscr{W}} (\lambda_w - \mu\widetilde{\lambda}_w)e_w^\top \geq 0, \tag{4.18}$$

$$\overline{d}^\top\left(\nabla_d F(\overline{d},\overline{v}) - \sum_{w\in\mathscr{W}} (\lambda_w - \mu\widetilde{\lambda}_w)e_w^\top\right) = 0. \tag{4.19}$$

Proof Under the partial calmness of problem (4.3) at $(\overline{d},\overline{v})$, there exists $\mu>0$ such that $(\overline{d},\overline{v})$ solves the counterpart of problem (4.13), where inclusion $d\in D$ is part of the constraints. Now consider the set-valued mapping

$$\Phi(d',v') := \{(d,v)\in D\times\mathbb{R}^\alpha |\, (d+d', v+v')\in \mathrm{gph}\, V\}.$$

Considering the expression of $V(d) := \Delta Q(d)$, the graph of Φ is obtained as:

$$\begin{aligned}\mathrm{gph}\Phi &= \{(d,v,d',v') | \exists q : 0\leq q\leq c,\ d\geq 0,\ \Delta q = v+v',\ \Lambda q = d+d'\}\\ &= \textstyle\prod_{1,2,3,4}\{(d,v,d',v',q) | 0\leq q\leq c,\ d\geq 0,\ \Delta q = v+v',\ \Lambda q = d+d'\},\end{aligned}$$

where $\prod_{1,2,3,4}$ denotes the canonical projection from $\mathbb{R}^\omega\times\mathbb{R}^\alpha\times\mathbb{R}^\omega\times\mathbb{R}^\alpha\times\mathbb{R}^\pi$ to $\mathbb{R}^\omega\times\mathbb{R}^\alpha\times\mathbb{R}^\omega\times\mathbb{R}^\alpha$. Clearly, Φ is a polyhedral set-valued mapping in the sense of Robinson (1981). Thus, Φ is calm (see e.g. Henrion et al. (2002) for the definition) at $(0,\overline{d},\overline{v})$. Hence, combining Mordukhovich (2006, Proposition 5.3) and Henrion et al. (2002, Corollary 4.2), we have:

$$0\in\nabla F(\overline{d},\overline{v}) + \mu\nabla f(\overline{d},\overline{v}) + \mu\partial(-\varphi)(\overline{d})\times\{0\} + N_{gphV}(\overline{d},\overline{v}) + N_{D\times\mathbb{R}^\alpha}(\overline{d},\overline{v}). \tag{4.20}$$

Taking into account the expressions of $N_{gphV}(\overline{d},\overline{v})$ and $\partial\varphi(\overline{d})$ from Lemma 4.1 and Lemma 4.2, respectively, there exist $\overline{q},\widetilde{q}$, $(\lambda^\omega,\lambda^c,\lambda^o)$ and $(\widetilde{\lambda}^\omega,\widetilde{\lambda}^c,\widetilde{\lambda}^o)$ satisfying (4.9)–(4.11) and (4.12), respectively, such that relationships (4.16)–(4.17), together with the following condition are satisfied:

$$-\nabla_d F(\overline{d},\overline{v}) + \sum_{w\in\mathscr{W}} (\lambda_w - \mu\widetilde{\lambda}_w)e_w^\top \in N_D(\overline{d}). \tag{4.21}$$

As in the previous theorem, (4.7) follows from the combination of (4.16) and (4.17) while (4.21) yields (4.18) and (4.19) considering the fact that $N_D(\overline{d}) = \{\lambda | \lambda \leq 0,\ \overline{d}^\top \lambda = 0\}$. □

Although the framework of this result is close to that of Chen (1994), as we set $D := \{d \in \mathbb{R}^\omega | d \geq 0\}$, the flow conservation constraint $v = \Delta q$ is fully part of the feasibility requirements of the traffic assignment problem in our result, which is the case for the aforementioned works. Moreover, they imposed a framework ensuring the differentiability of the value function φ of the traffic assignment problem (4.2). In fact, such strong conditions would not change the outcome of our results above. Finally, it may be important to mention that Chen (1994) substituted the components of the route flow q_p by $l_p d_w$, where l_p is a path (route) flow proportion, which allows them to almost eliminate the demand vector d from the feasible set of the traffic assignment problem.

5 Final remarks

The study made in this paper for the bilevel road pricing problem (1.5) with the functions F and f being continuously differentiable can easily be extended to the case where these functions are Lipschitz continuous. The set of feasible tolls Γ can also be replaced by any convex set and, given that for the optimality conditions only an upper estimation of the normal cone is needed, then results in the book by Rockafellar and Wets (1998) can be used for this purpose, provided that Γ is defined by functions satisfying an appropriate constraint qualification.

As already mentioned in Sect. 2, the network design and road pricing problems have the same bilevel structure; that is, the parameter controlled by the leader appears in the cost function of the follower's problem. Hence, the same techniques presented above can be applied for the network design problem and similar results will be obtained for the corresponding optimal value function, the constraint qualifications and the optimality conditions.

Acknowledgements We would like to thank the two anonymous referees for carefully reading our manuscript. Their comments and suggestions have been of a great help for improvements in the paper. The work of A.B. Zemkoho was supported by the Deutscher Akademischer Austausch Dienst (DAAD).

References

Abrahamsson, T. (1998). *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.

Beckmann, M., Mcguire, C. B., & Winsten, C. B. (1956). *Studies in the economics of transportation*. New Haven: Yale University Press.

Chen, Y. (1994). *Bilevel programming problems: analysis, algorithms and applications*. PhD Thesis, Centre de Recherche sur les Transports, Université de Montréal, CRT-984.

Chen, Y., & Florian, M. (1998). Congested O-D trip demand adjustment problem: bilevel programming formulation and optimality condtions. In A. Migdalas, et al. (Eds.), *Multilevel optimization: algorithms and aplications* (pp. 1–22). Dordrecht: Kluwer Academic.

Chiou, S.-W. (2005). Bilevel programming for the continuous transport network design problem. *Transportation Research Part B*, *39*, 361–383.

Codina, E., & Montero, L. (2006). Approximation of the steepest descent direction for the O-D matrix adjustment problem. *Annals of Operations Research*, *144*, 329–362.

Dempe, S. (1993). Directional differentiability of optimal solutions under Slater's condition. *Mathematical Programming*, *59*, 49–69.

Dempe, S., & Dutta, J. (2010). Is bilevel programming a special case of mathematical programming with equilibrium constraints? *Mathematical Programming*. doi:10.1007/s10107-010-0342-1.

Dempe, S., & Schmidt, H. (1996). On an algorithm solving two-level programming problems with nonunique lower level solutions. *Computational Optimization and Applications*, *6*, 227–249.

Dempe, S., & Vogel, S. (2001). The generalized Jacobian of the optimal solution in parametric optimization. *Optimization*, *50*, 387–405.

Dempe, S., & Zemkoho, A. B. (2008). *A bilevel approach for traffic management in capacitated networks*. Preprint 2008-05, Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg.

Dempe, S., & Zemkoho, A. B. (2010, submitted). *The bilevel programming problem: reformulations, constraint qualifications and optimality conditions*. Preprint 2010-03, Fakultät für Mathematik und Informatik, TU Bergakademie Freiberg.

Dempe, S., & Zemkoho, A. B. (2011a). On the Karush-Kuhn-Tucker reformulation of the bilevel optimization problem. *Nonlinear Analysis: Theory, Methods & Applications*. doi:10.1016/j.na.2011.05.097.

Dempe, S., & Zemkoho, A. B. (2011b). The generalized Mangasarian-Fromowitz constraint qualification and optimality condtions for bilevel programs. *Journal of Optimization Theory and Applications*, *148*, 46–68.

Dewez, S., Labbée, M., Marcotte, P., & Savard, G. (2008). New formulations and valid inequalities for a bilevel pricing problem. *Operations Research Letters*, *36*, 141–149.

Fisk, C. S. (1988). On combining maximum entropy trip matrix estimation with user optimal assignment. *Transportation Research B*, *22*, 66–79.

Friesz, T. L., Tobin, R. L., Cho, H. J., & Mehta, N. J. (1990). Sensitivity analysis based heuristic algorithms for mathematical programs with variational inequality constraints. *Mathematical Programming*, *48*, 265–284.

Heilporn, G., Labbé, M., Marcotte, P., & Savard, G. (2010). A polyhedral study of the network pricing problem with connected toll arcs. *Networks*, *55*(3), 234–246.

Henrion, R., Jourani, A., & Outrata, J. (2002). On the calmness of a class of multifunctions. *SIAM Journal on Optimization*, *13*(2), 603–618.

Josefsson, M., & Patriksson, M. (2007). Sensivity analysis of separable traffic equilibrium equilibria with application to bilevel optimization in network design. *Transportation Research Part B*, *41*(1), 4–31.

Labbée, S., Marcotte, P.M., & Savard, G. (1998). A bilevel model of taxation and its application to optimal highway pricing. *Management Science*, *44*(12), 1608–1622.

Lu, S. (2008). Sensitivity of static traffic user equilibria with perturbations in arc cost function and travel demand. *Transportation Science*, *42*(1), 105–123.

Lundgren, J. T., & Peterson, A. (2008). A heuristic for the bilevel origin-destination matrix estimation problem. *Transportation Research Part B*, *42*(4), 339–354.

Meng, Q., Yang, H., & Bell, M. G. H. (2001). An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem. *Transportation Research B*, *35*, 83–105.

Migdalas, A. (1995). Bilevel programming in traffic planning: models, methods and challenge. *Journal of Global Optimization*, *7*, 381–405.

Mordukhovich, B. S. (2006). *Variational analysis and generalized differentiation I/II. Grundlehren der mathematischen Wissenschaften*. Berlin: Springer.

Mordukhovich, B. S., & Nam, N. M. (2005). Variational stability and marginal functions via generalized differentiation. *Mathematics of Operational Research*, *30*(4), 800–816.

Netter, M. (1972). Affectations de traffic et tarification au coût marginal social: critiques de quelques idées admises. *Transportation Research*, *6*, 411–429.

Noriega, Y., & Florian, M. (2009). Some enhancements of the gradient method for the O-D matrix adjustment. In *CIRRELT-2009-04*.

Outrata, J. (1990). On the numerical solution of a class of Stackelberg problems. *ZOR-Methods and Models of Operations Research*, *34*, 255–277.

Patriksson, M. (1994). *The traffic assignment problem – models and methods, topics in transportation*. Utrecht: VSP BV.

Patriksson, M. (2004). Sensitivity analysis of traffic equilibria. *Transportation Science*, *38*, 258–281.

Patriksson, M., & Rockaffelar, R. T. (2002). A mathematical model and descent algorithm for bilevel traffic management. *Transport Science*, *36*(3), 271–291.

Ralph, D., & Dempe, S. (1995). Directional derivatives of solution of parametric nonlinear program. *Mathematical Programming*, *68*, 159–172.

Robinson, S. M. (1981). Some continuity properties of polyhedral multifunctions. *Mathematical Programming Study*, *14*, 206–214.

Robinson, S. M. (2006). Strong regularity and the sensitivity analysis of traffic equilibria: A comment. *Transportation Science*, *40*(4), 540–542.

Rockafellar, R. T., & Wets, R. J.-B. (1998). *Variational analysis*. *Grundlehren der Mathematische Wissenschaften*. Berlin: Springer.

Wardrop, J. G. (1952). Some theoretical aspects of road traffic research. In *Proceedings of the Institute of Civil, Engeneers, Part II* (pp. 325–378).

Yang, H. (1995). Heuristic algorithms for the bilevel origin-destination matrix estimation problem. *Transportation Research B*, *29*(4), 231–242.

Yang, H. & Bell, M. G. H. (1997). Traffic restraint, road pricing and network equilibrium. *Transportation Research B*, *31*(4), 303–314.

Yang, H., & Lam, W. H. K. (1996). Optimal road tolls under conditions of queueing and congestion. *Transportation Research A*, *30*(5), 319–332.

Yang, H., Sasaki, T., Iida, Y., & Asakura, Y. (1992). Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research B*, *26*(6), 417–434.

Yang, H., & Yagar, S. (1994). Traffic assignment and traffic control in general freeway-arterial corridor systems. *Transportation Research B*, *28*(6), 463–486.

Ye, J. J. (1998). New uniform parametric error bounds. *Journal of Optimization Theory and Applications*, *98*(1), 197–219.

Ye, J. J., & Zhu, D. L. (1995). Optimality conditions for bilevel programming problems. *Optimization*, *33*(1), 9–27 (with Erratum in (1997). *Optimization, 39*(4), 361–366).