A Mathematical Model and Descent Algorithm for Bilevel Traffic Management

Michael Patriksson • R. Tyrrell Rockafellar
Department of Mathematics, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden
Department of Mathematics, University of Washington, Seattle, Washington 98195-4350
mipat@math.chalmers.se • rtr@math.washington.edu

We provide a new mathematical model for strategic traffic management, formulated and analyzed as a mathematical program with equilibrium constraints (MPEC). The model includes two types of control (upper-level) variables, which may be used to describe such traffic management actions as traffic signal setting, network design, and congestion pricing. The lower-level problem of the MPEC describes a traffic equilibrium model in the sense of Wardrop, in which the control variables enter as parameters in the travel costs. We consider a (small) variety of model settings, including fixed or elastic demands, the possible presence of side constraints in the traffic equilibrium system, and representations of traffic flows and management actions in both link-route and link-node space.

For this model, we also propose and analyze a descent algorithm. The algorithm utilizes a new reformulation of the MPEC into a constrained, locally Lipschitz minimization problem in the product space of controls and traffic flows. The reformulation is based on the Minty (1967) parameterization of the graph of the normal cone operator for the traffic flow polyhedron. Two immediate advantages of making use of this reformulation are that the resulting descent algorithm can be operated and established to be convergent without requiring that the travel cost mapping is monotone, and without having to ever solve the lower-level equilibrium problem. We provide example realizations of the algorithm, establish their convergence, and interpret their workings in terms of the traffic network.

1. Introduction

The need for measures to reduce congestion in the metropolitan traffic areas is becoming more serious as citizens cluster in cities with the immediate side effect of an increase in traffic demand. A functioning society depends on the mobility provided by the transportation network to enable its members to participate in essential activities such as production, consumption, communication, and recreation. It is however necessary for society also to introduce congestion-relief measures for the quality of life, the environment, and the safety of the citizens not to deteriorate.

Any well-founded traffic model recognizes the individual network user's right to decide when, where, and how to travel. The criteria by which the user makes these choices are selfish, and are therefore on the aggregate level not entirely in par with society's goals of an efficient and safe utilization of the traffic network. A classical example of this conflict is that the typical traveller can be expected to choose a route between his/her origin and destination such that the combined travel time and cost is minimal given the network conditions when the travel is made; the aggregate effect of these decisions is a network flow that does not minimize the total system costs.

We may model this conflict in the traffic system as a noncooperative Stackelberg game, in which a traffic manager, represented as the leader, takes some action, such as a change in the infrastructure or in the traffic signal plans, so as to achieve some overall

0041-1655/02/3603/0271$05.00
1526-5447 electronic ISSN

management goal with respect to the distribution of the traffic in the network and some measure of network performance. The travellers are then modelled as the followers; they react to the actions of the manager by modifying their behavior, for example, by adjusting their route choices, travel modes, or time of day to travel. If the manager's actions are adequate, then the travellers' response is the desired one. Common means for achieving such a change in the traffic flows are to invest in traffic network capacity, to introduce or adjust traffic controls such as traffic lights, to introduce tolls on some links, or to supply the travellers with information about alternative routes. The common point in all these traffic management models is that the main control variable is the users' travel costs (or, rather, travel cost adjustments), even though they may be influenced indirectly through the actions implemented.

Taking as the starting point a general Stackelberg model of the decision problem, we provide several examples of possible management goals and corresponding model instances. Measures that one can model through this strategy are variations in traffic signal plans, alterations of the network infrastructure, also referred to as network design, and the introduction and setting of tolls in the network. We begin however with a short discussion on the equilibrium model that represents the behavior of the network users given a network infrastructure. We stress already at this stage that we shall be concerned only with the case of static equilibrium conditions, thus ignoring in this work possible time-dependent (dynamic) effects.

2. Wardrop Equilibrium

Let $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ be a transportation network, where $\mathcal{N}$ and $\mathcal{A}$ are the sets of nodes and directed links (arcs), respectively. For certain ordered pairs of nodes, $(p, q) \in \mathcal{C}$, where node p is an origin, node q is a destination, and $\mathcal{C}$ is a subset of $\mathcal{N} \times \mathcal{N}$, there are positive travel demands d_{pq} (which initially shall be assumed fixed) giving rise to a link traffic flow pattern. We assume that the network is strongly connected, that is, that at least one route joins each origin-destination (OD) pair.

Wardrop's user equilibrium principle (1952) states that for every OD pair $(p, q) \in \mathcal{C}$, the travel costs of the routes utilized are equal and minimal. We denote by $\mathcal{R}_{pq}$ the set of simple (loop-free) routes for OD pair (p, q), by h_{pqr} the flow on route $r \in \mathcal{R}_{pq}$, and by $\bar{c}_{pqr} := \bar{c}_{pqr}(h)$ the travel cost on the route given the vector $h \in \Re^{|\mathcal{R}|}$ of route flows, where $|\mathcal{R}|$ denotes the total number of routes in the network; with this notation, an equilibrium flow is defined by the conditions

$$h_{pqr} > 0 \Longrightarrow \bar{c}_{pqr} = \pi_{pq}, \quad r \in \mathcal{R}_{pq}, \quad (p, q) \in \mathcal{C}, \quad \text{(1a)}$$

$$h_{pqr} = 0 \Longrightarrow \bar{c}_{pqr} \geq \pi_{pq}, \quad r \in \mathcal{R}_{pq}, \quad (p, q) \in \mathcal{C}, \quad \text{(1b)}$$

where the value of $\pi_{pq} := \pi_{pq}(h)$ is the minimal (i.e., equilibrium) route cost in OD pair (p, q). By the nonnegativity of the route flows, the system (1) can more compactly be written as the complementarity system

$$0 \leq h_{pqr} \perp (\bar{c}_{pqr} - \pi_{pq}) \geq 0, \quad r \in \mathcal{R}_{pq}, \quad (p, q) \in \mathcal{C}, \quad (2)$$

where $a \perp b$, for two arbitrary vectors $a, b \in \Re^n$, means that $a^{\mathrm{T}} b = 0$. The Wardrop conditions state that an equilibrium state is reached precisely when no traveller can decrease his/her travel cost by unilaterally shifting to another route.

To cast the Wardrop conditions as a variational inequality problem, we need to decide in which space we wish to represent the flows and the flow feasibility requirements. A general form is obtained by describing the set of feasible, aggregate, link flows as the solution in $f \in \Re^{|\mathcal{A}|}$ to the linear system

$$f = Vv, \quad \text{(3a)}$$

$$Wv = d, \quad \text{(3b)}$$

$$v \geq 0, \quad \text{(3c)}$$

where v is the (disaggregated) vector of the commodity flows, V is an incidence matrix that describes the aggregation of these flows into a corresponding link flow f, and W is an incidence matrix that describes the feasibility requirements with respect to the demand, d, in the commodity flow space.

The most common representation of the Wardrop conditions as a variational inequality problem is in terms of the route flow variables h_{pqr}. We obtain this formulation by identifying $v = h$, $d \in \Re^{|\mathcal{C}|}_{++}$ as the vector of each OD pair's demand, and $W = \Gamma^{\mathrm{T}}$, where

$\Gamma \in \Re^{|\mathcal{R}|\times|\mathcal{C}|}$ is the route-OD pair incidence matrix (i.e., the element γ_{rk} is one if route r joins OD pair $k = (p, q) \in \mathcal{C}$, and zero otherwise). In a *disaggregated* version of the Wardrop conditions, we consider only utilizing the part (3b)–(3c) of the system (3) above, thus describing the (bounded polyhedral) set

$$H := \{h \in \Re^{|\mathcal{R}|}_+ \,|\, \Gamma^{\mathrm{T}} h = d\}$$

of demand-feasible route flows. The condition (1) is equivalent to h satisfying

$$-\bar{c}(h) \in N_H(h), \qquad \text{[VIP-}H\text{]}$$

where $\bar{c} : \Re^{|\mathcal{R}|}_+ \mapsto \Re^{|\mathcal{R}|}_{++}$ is the vector of route travel cost functions, and $N_S(s)$ denotes the normal cone to a nonempty, closed and convex set $S \subseteq \Re^n$ at $s \in \Re^n$, that is, the set

$$N_S(s) := \begin{cases} \{z \in \Re^n \mid z^{\mathrm{T}}(y - s) \leq 0, \quad \forall\, y \in S\}, & s \in S, \\ \varnothing, & s \notin S. \end{cases}$$

To see this equivalence directly, we utilize the notation Γ to rewrite (2) as follows:

$$0 \leq h \perp (\bar{c}(h) - \Gamma\pi) \geq 0. \qquad (4)$$

Together with the feasibility requirement that $\Gamma^{\mathrm{T}} h = d$ must hold, the system (4) describes the optimality conditions for h, solving the linear program to minimize $\bar{c}(h)^{\mathrm{T}} y$ over $y \in H$; this is precisely [VIP-H].

We remark here that the existence of several groups of users or modes of transport is easily modelled within the above framework, by simply creating a copy of the network for each user group and mode, and relating their travel costs, if needed, through the vector $\bar{c}$.

In the case where the travel cost of a route is the sum of the travel costs on the links defining it (i.e., the route costs are additive), then the above Wardrop conditions can be described in terms of link flows. We then further identify $V = \Lambda$ in (3a), where $\Lambda \in \{0, 1\}^{|\mathcal{A}|\times|\mathcal{R}|}$ is the link-route incidence matrix (i.e., the element λ_{ar} equals one if route r utilizes link a, and zero otherwise), and thus the (bounded polyhedral) set of demand-feasible link flows

$$\widehat{F} := \{f \in \Re^{|\mathcal{A}|} \,|\, \exists h \in H \text{ with } f = \Lambda h\}.$$

Then, the problem [VIP-H] can be equivalently written as

$$-\bar{t}(f) \in N_{\widehat{F}}(f), \qquad \text{[VIP-}\widehat{F}\text{]}$$

where $\bar{t} : \Re^{|\mathcal{A}|}_+ \mapsto \Re^{|\mathcal{A}|}_{++}$ is the vector of link travel cost functions. (The link and route costs are related by $\bar{c}(h) = \Lambda^{\mathrm{T}} \bar{t}(f)$, for any pair $(h, f) \in H \times \widehat{F}$.)

The set of feasible link flows can also be described by the OD-specific link flows that satisfy the demand for transportation and flow conservation constraints for all the nodes of the network; this is the second most popular representation of feasible flows. In the system (3), we then identify W as a block-diagonal matrix with $|\mathcal{C}|$ blocks W_k, with $W_k = A$, $A \in \{-1, 0, 1\}^{|\mathcal{N}|\times|\mathcal{A}|}$ being the node-link incidence matrix of the network. Further, d is a $(|\mathcal{C}| \cdot |\mathcal{A}|)$-vector, with $|\mathcal{C}|$ vectors d_k, each being a vector of OD-specific demands, stacked on top of each other. (The elements of d_k sum to zero.) We also identify v as the $(|\mathcal{C}| \cdot |\mathcal{A}|)$-vector of commodity link flows f_{ak}. Hence, (3b) corresponds to the commodity-specific flow conservation constraints

$$Af_k = d_k, \qquad k \in \mathcal{C}.$$

Finally, V is the block-diagonal $(|\mathcal{A}| \times |\mathcal{C}|)$-matrix that describes the aggregation of the commodity link flows f_k into f. Summarizing, then, the system (3) describes the (unbounded polyhedral) set of demand-feasible link flows

$$F := \left\{ f \in \Re^{|\mathcal{A}|} \,\middle|\, \exists f_k \in \Re^{|\mathcal{A}|}_+,\ k \in \mathcal{C}, \text{ with } f = \sum_{k\in\mathcal{C}} f_k \text{ and } Af_k = d_k \right\}.$$

In the present setting, of course, k is identified with an OD pair $(p, q) \in \mathcal{C}$, and, further, each vector d_k has precisely two nonzeros. We may, however, consider k to denote a less disaggregated flow, such as flows from different origins, or different vehicle types, etc. The two representations that we have chosen here are in that sense at the two extremes in terms of level of aggregation. We also note that in more generality, we may consider different networks, that is, different matrices A_k, for each commodity k, or type k of traffic. This will necessarily also lead to a proper modification of the matrix V above.

Note that $\widehat{F} \subset F$ holds because the latter contains cyclic flows, but due to the positivity assumption on $\bar{t}$, no equilibrium flow will utilize any cyclic flow, so this alternative representation is, in that sense, equivalent. In the following, we shall always consider the version [VIP-F] whenever considering a link flow-based equilibrium system.

More general considerations, such as the possible presence of side constraints in the equilibrium system, or that the demand at equilibrium depends on the cost of the trip, is relegated to §7.

3. A General Stackelberg Model

3.1. The Mathematical Model

This section introduces a general Stackelberg model for the society's traffic management problem. We introduce two vectors, ρ and β, of parameters denoting the actions taken by the traffic manager.

The parameter ρ is assumed to enter the travel cost function, leading to the parameterized (and presumed continuous) function $t(\rho, f)$ (in the case of the equilibrium model [VIP-F]), or $c(\rho, h)$ (in the case of the equilibrium model [VIP-H]). Further, ρ is constrained to a polyhedral set, which we denote by $P \subset \Re^p$ and which may be determined by political, practical, environmental, and economical constraints, and possibly other considerations as well. (The assumption that P is polyhedral is not essential to the results presented, but simplifies some parts of the algorithm.)

The parameter β enters the travel cost function as an additive term. So, given actions (ρ, β), the travel cost mapping takes the form $f \mapsto t(\rho, f) + \beta$ (respectively, $h \mapsto c(\rho, h) + \beta$). We allow for no explicit constraints on β because we wish for an equilibrium to always exist whatever the choice of $\rho \in P$. However, one could always include smooth penalties for any constraints one wishes to impose on β into the objective function φ, to be discussed next.

Among the possible actions, the manager optimizes a function, φ, defined over $P \times \Re^{|\mathcal{A}|} \times \Re^{|\mathcal{A}|}$ (respectively, $P \times \Re^{|\mathcal{R}|} \times \Re^{|\mathcal{R}|}$), of the actions and traffic flows. This function may include some further measures of network performance as well as measures of the cost and/or benefits associated with a given action. We shall presume throughout that this function is continuously differentiable on $P \times \Re^{|\mathcal{A}|} \times \Re^{|\mathcal{A}|}$ but remark that in principle piecewise differentiability would suffice.

Taking [VIP-F] as the underlying equilibrium model, the general problem then is to

[MPEC-F]

$$\text{minimize} \quad \varphi(\rho, \beta, f) \tag{5a}$$

$$\text{subject to} \quad \rho \in P, \tag{5b}$$

$$-t(\rho, f) - \beta \in N_F(f). \tag{5c}$$

For further reference, we shall denote the set f of solutions to (5c) by $S(\rho, \beta)$. If the lower-level problem (5c) has unique solutions f, then the problem [MPEC-F] is well defined, but in situations where there is more than one equilibrium solution, it is not clear how to interpret the minimization operation in [MPEC-F] because the value of $\varphi(\rho, \beta, f)$ is then impossible to predict. We next turn to explain our proposal to resolve this issue.

3.2. Sensitivity Analysis and Well Posedness Under Lower-Level Nonuniqueness

In the case where the cost mapping $f \mapsto t(\rho, f) + \beta$ is positive and strictly monotone on F, the solution, f, to (5c) is uniquely determined by (ρ, β), that is, $S(\rho, \beta)$ is a singleton set. We may, in the situation that this is true for every $(\rho, \beta) \in P \times \Re^{|\mathcal{A}|}$, think of the problem [MPEC-F] as that to find the minimum of the function $(\rho, \beta) \mapsto \varphi(\rho, \beta, f(\rho, \beta))$ over $P \times \Re^{|\mathcal{A}|}$, where $f(\rho, \beta)$ denotes the unique solution to (5c). This *implicit* function is continuous on $P \times \Re^{|\mathcal{A}|}$.

For the development of efficient algorithms for finding a minimum of the function $(\rho, \beta) \mapsto \varphi(\rho, \beta, f(\rho, \beta))$ over $P \times \Re^{|\mathcal{A}|}$ it is detrimental that it has stronger differentiability properties. Much is, of course, known about the sensitivity and stability of solutions to perturbed variational problems (see, for example, the monographs in Luo et al. 1996, Rockafellar and Wets 1998, Outrata et al. 1998, and Bonnans and Shapiro 2000), and specialized analyses have been conducted also for the case at hand. It is important to note that in traffic equilibrium models the presence of variables at different levels of aggregation (total link flows together with commodity link flows or route flows) means that some of the traditional techniques in sensitivity analysis, such as the strong regularity results by Robinson (1980), are applicable only by viewing the problem in terms of the aggregated variables (total link flows). In the sensitivity analysis of traffic equilibria, this possibility has most often been overlooked, and much (unnecessary) effort has been spent on the choice of an appropriate disaggregated flow (cf. Tobin and Friesz 1988, Qiu and Magnanti 1989, Yen 1995, Cho et al. 2000). The fact that the sensitivity analysis is independent of any such choice was first demonstrated in Patriksson and Rockafellar (2002) for the case of elastic demands. Patriksson (2002) provides a rather complete analysis of the sensitivity of traffic equilibria and improves and extends the analyses made in Tobin and Friesz (1988), Qiu and Magnanti (1989), Yen (1995), Outrata (1997), Cho et al. (2000), and Patriksson and Rockafellar (2002), including characterizations of the differentiability, and the generation of subgradients of the mapping S at a reference point (ρ^r, β^r). An overview of the sensitivity analysis of the problem at hand follows.

Assume that the parameterization is rich enough so that the rank of the Jacobian matrix $\nabla_\rho t(\rho^r, f^r)$ is full (that is, p). (This condition can always be fulfilled through the introduction of dummy parameters.) We introduce the *sensitivity problem* as that of, given a perturbation $(\delta\rho, \delta\beta)$, finding a solution to the variational inequality

$$DS((\rho^r, \beta^r) \mid f^r)(\delta\rho, \delta\beta) := \left\{\delta f \in \Re^{|\mathcal{A}|} \,\middle|\, -r(\delta\rho, \delta\beta, \delta f) \in N_K(\delta f)\right\}, \quad (6)$$

where

$$r(\delta\rho, \delta\beta, \delta f) := \nabla_\rho t(\rho^r, f^r)\delta\rho + \delta\beta + \nabla_f t(\rho^r, f^r)\delta f \quad (7)$$

is the problem mapping, and where its feasible set is the critical cone,

$$K := T_F(f^r) \cap [t(\rho^r, f^r) + \beta^r]^\perp, \quad (8)$$

where $T_F(f^r)$ is the tangent cone to F at f^r. Further, for any vector $z \in \Re^n$, $z^\perp := \{y \in \Re^n \mid z^{\mathrm{T}} y = 0\}$ is the orthogonal subspace associated with the vector z. The problem (6) amounts to solving an affine variational inequality defined such that we retain first-order optimality and feasibility in the original model. (In Qiu and Magnanti 1989, Patriksson and Rockafellar 2002, and Patriksson 2002 it is shown that these types of problems are special affine traffic equilibrium problems over variations of the original traffic network.) The first result states that the mapping S is *strongly regular* in the sense of Robinson (1980 and 1985), that is, single-valued and locally Lipschitz continuous, at (ρ^r, β^r), if, and only if, the solution δf to the sensitivity problem (6) is unique for each choice of perturbation. Moreover, when this condition is satisfied, this unique value is the directional derivative of the equilibrium link flow solution at (ρ^r, β^r) in the direction of $(\delta\rho, \delta\beta)$, and the mapping S is B-differentiable (in the sense of Robinson 1985), and, equivalently, semidifferentiable (in the sense of Rockafellar and Wets 1998). This property is important, in that it is exactly what is needed to apply a Newton-type algorithm for the problem to minimize $(\rho, \beta) \mapsto \varphi(\rho, \beta, f(\rho, \beta))$ over $P \times \Re^{|\mathcal{A}|}$, a subject which has yet not found application in the context of traffic management. (See further the references Pang and Qi 1993, Qi 1993, and Qi and Sun 1993, for more information about Newton methods for semi-smooth functions.) It is also enough to be able to devise bundle subgradient algorithms for the problem; see the further research section and Patriksson (2002) for further details on that subject.

A sufficient, but not necessary, condition for strong regularity is that the partial Jacobian $\nabla_f t(\rho^r, f^r)$ satisfies the condition that

$$s^{\mathrm{T}} \nabla_f t(\rho^r, f^r) s > 0, \qquad s \in K - K, \quad (9)$$

that is, a positive definiteness condition on the *critical subspace* associated with the problem (5c) at (ρ^r, β^r, f^r). This type of condition has been utilized in Qiu and Magnanti (1989), Yen (1995), and Outrata (1997).

The mapping S is moreover differentiable at (ρ^r, β^r) if, and only if, for every choice of perturbation vector $(\delta\rho, \delta\beta)$ it holds that if a route r or a link for a specific commodity is such that its flow in every equilibrium solution is zero, then it remains zero in every solution to the sensitivity problem. This result in Patriksson (2002) improves on those previously stated in Tobin and Friesz (1988), Cho et al. (2000), which assume that there is a strictly complementary equilibrium link flow. Not only is it shown in Patriksson (2002) that the strict complementarity condition is stronger than necessary, but that the computational formula in Tobin and Friesz (1988), Cho et al. (2000) may fail to produce a gradient value even if one exists, or even provide a value when no gradient exists. Patriksson (2002) also supplies a calculus formula for a subgradient of S at (ρ^r, β^r) in the absence of a gradient; the problem solved to obtain a subgradient is similar to that of calculating a directional derivative in each coordinate direction, but it contains only equality constraints.

In all events, the technical conditions stated above amount to some form of positive definiteness of the Jacobian of the travel cost function, which is naturally implied by some strict (or, strong) monotonicity assumption with respect to $t(\rho, \cdot)$ on F for every $\rho \in P$. The assumption that the equilibrium solution is unique (which it will boil down to of course) is, however, often too strong to be accepted easily. We mention two such cases. If we wish to consider the underlying traffic equilibrium model [VIP-H] in our development of traffic management instruments, we must note that the equilibrium solution h is likely to *never* be uniquely determined by (ρ, β) even if f is because a link flow is not uniquely decomposable into route flows in general. (A counterexample is however the stochastic user equilibrium (SUE) model of Fisk (1980), in whose solution the route flows are unique; a necessary condition for this to hold is that the route costs are not additive. See Davis (1994) for an application of the SUE model in bilevel network design and Patriksson (2002) for a characterization of the gradient of the equilibrium link flow as an asymptotic result of SUE sensitivity analysis.) Further, if we model cost interactions between links, particularly for links joining the same intersection, it has been demonstrated (e.g., Heydecker 1983) that the appropriate travel cost mapping t will not even be monotone, whence the equilibrium link flow solution determined by the Wardrop conditions will not necessarily be unique either. (The same is true for multiclass user traffic equilibrium models; see Toint and Wynter 1996.) In our continued development, we will presume that the cost mappings t and c are *continuously differentiable* on their respective domains F and H, but we make *no assumption about their monotonicity*.

The effect of a nonuniqueness in the lower-level problem is of course that the value of φ becomes unpredictable (and may also in some cases imply the nonexistence of optimal solutions to the bilevel problem altogether; see, e.g., Bard and Fulk 1982). We therefore need a finer rule for choosing *one* element in the set of solutions, $S(\rho, \beta)$, to the equilibrium system (5c). (In the literature of Stackelberg games, this set is known as the rational reaction, or response, set.) There are several approaches to this problem (see, e.g., Lordin and Morgan 1992 and 1996, Dempe and Schmidt 1996, Dempe 2000). The two most common ones are usually referred to as the *optimistic* or, strong or cooperative) approach and the *pessimistic* (or, weak or noncooperative) approach. The optimistic approach is to assume that the followers (travellers) in the game establish (or, choose) one equilibrium that minimizes $\varphi(\rho, \beta, \cdot)$ over the set $S(\rho, \beta)$, thereby assuming a kind of cooperation on the part of the followers. The resulting objective value for [MPEC-F] is then

$$\hat{\varphi}(\rho, \beta) := \min_{f \in S(\rho, \beta)} \varphi(\rho, \beta, f),$$

whenever the minimum is attained. The pessimistic approach is precisely the opposite assumption, leading to a kind of worst-case optimal solution wherein the damage resulting from an unwelcome choice of the followers is minimized. A third alternative (Dempe 1997) is to introduce a perturbation of the optimistic solution to better try to reflect the behavior of the followers. Finally, a completely different way out (Dempe and Schmidt 1996, Dempe 2000) is to introduce a strictly or strongly monotone regularizing term in the lower-level cost mapping, making the lower-level solution uniquely determined, and the associated positive scaling factor is forced to zero to approach the original equilibrium problem.

The Minty parameterization of the equilibrium system (5c) provided in the next section leads to a one-level optimization problem, which we will show is equivalent to [MPEC-F] (in the sense that they have the same set of locally optimal solutions), provided that we take the optimistic approach but not necessarily otherwise. In that section, we will also complete the discussion on the possible nonuniqueness of lower-level solutions with some remarks on its consequences for decentralized traffic control through link tolls.

3.3. Instances

We next illustrate the scope of the Stackelberg model [MPEC-F]. (An overview of bilevel optimization models in the field of transportation is found in Migdalas 1995.)

Example 1 (Network Design). A familiar form of the *equilibrium network design problem* (LeBlanc and Abdulaal 1979, Marcotte 1986) is an instance of [MPEC-F]. Let ρ_a denote an investment in network capacity on link a; the effect of an investment is that of a reduced travel time; its form is often taken to be $t_a(\rho, f) := \bar{t}_a(f_a/\rho_a)$. An investment ρ_a is associated with an investment cost, $\psi_a(\rho_a)$. The goal is to minimize the total travel time, at a user equilibrium flow (that is, $\varphi(\rho, f) := \sum_{a\in\mathcal{A}} c_a(\rho, f) f_a$) while satisfying budget constraints on the investments made, $\rho \in P := \{\, \rho \in \Re_+^{|\mathcal{A}|} \mid \ell \le \rho \le u;\ \psi(\rho) \le b;\ \sum_{a\in\mathcal{A}} \rho_a \le U \,\}$.

The parameters ρ_a may also be associated with the lowering of capacity of a link, such as when a lane is narrowed to allow for the construction of a bicycle lane. The lowering of capacity on certain links then acts as an influence on the travellers to choose other routes, other modes, etc. If the lower-level model (5c) is a multimodel model that allows for the demand d to differentiate between different modes of transport, then [MPEC-F] may be used, for example, to model an influence on the travellers to utilize public transport or the bicycle alternative through inducing an additional delay for cars.

Example 2 (Signal Control). A problem of a form similar to the equilibrium network design problem is the *signal setting problem*. The solution of this problem aims at finding a set of signal control parameter values that, under user equilibrium conditions, optimizes some measure of the performance of the network, such as the total queueing delay, but without altering the traffic infrastructure. In this case, the variables ρ are the signal control parameters, for example the portion of *green times* allocated to the signal controls, and the parameterized travel cost mapping $f \mapsto t(\rho, f)$ measures the sum of travel times and delays at intersections. (See Cantarella and Sforza 1986, and Smith and Van Vuren 1993, and references therein for examples of traffic control policies and mathematical models.) In this case, the set P is the unit simplex.

We note that the generality of the model allows for the introduction of queueing delays for private vehicles only to favour public transport.

Example 3 (Toll Optimization). The actions discussed in the above examples lead indirectly to an adjustment in the travellers' cost perception, through the increase or decrease in queueing delays, for example. It is also possible to associate the parameters with monetary expenses (although properly measured in time equivalents), such as link or route tolls. In such cases, we could let $t(\beta, f) := \bar{t}(f) + \beta$. Larsson and Patriksson (1998) discuss several alternative interpretations and uses of such a model, for example to induce mode changes through changes in ticket prices, and indirect derivations of β so as to satisfy some flow-side constraints in equilibrium. As an example, let φ denote the total travel costs, and let the equilibrium problem be [VIP-F]. The well-known marginal cost pricing solution (e.g., Dafermos and Sparrow 1971) is one optimal solution to this problem. (See Larsson and Patriksson 1998 and Bergendorff et al. 1996, for discussions about alternative pricing solutions.) In the case where the underlying traffic model is an elastic demand model (cf. the model [EVIP-H_d] of §7.1), the optimal solution is the zero flow. We then note that toll optimization under constraints can be modelled through the framework of this paper by letting $t(\rho, f) := \bar{t}(f) + \rho$ and by adding penalties to avoid a nonzero β.

4. A Reformulation Based on a Minty Parameterization

4.1. The Minty Parameterization

We will consider a reformulation of the problem [MPEC-F] into an equivalent (one-level) optimization problem in the space (ρ, f). The conversion is based on the Minty (1967) parameterization of the graph of the normal cone operator N_F.

That f solves (5c) is equivalent to the existence of a vector $v \in N_F(f)$ such that $v = -[t(\rho, f) + \beta]$. Recall the definition of the graph of N_F:

$$\operatorname{graph} N_F = \{\, (v, f) \mid v \in N_F(f) \,\}.$$

Because N_F is maximal monotone, it has a Minty parameterization, that is,

$$(v, f) \in \operatorname{graph} N_F \Longleftrightarrow \exists \hat{f} \in \Re^{|\mathcal{A}|}$$

with

$$f = \operatorname{Proj}_F(\hat{f}) \text{ and } v = (I - \operatorname{Proj}_F)[\hat{f}] = \hat{f} - f,$$

where Proj_F is the Euclidean projection operator for the convex set F. (This identity is in fact a very simple form of the Minty parameterization that utilizes the relation $v \in N_F(f) \Longleftrightarrow \operatorname{Proj}_F(f + v) = f$.)

We will utilize the Minty parameterization as follows. The condition (5c) may be viewed as the existence of a vector $\hat{f} \in \Re^{|\mathcal{A}|}$ such that $-t(\rho, \operatorname{Proj}_F(\hat{f})) - \beta = \hat{f} - \operatorname{Proj}_F(\hat{f})$, in other words,

$$\beta = \operatorname{Proj}_F(\hat{f}) - \hat{f} - t(\rho, \operatorname{Proj}_F(\hat{f})). \tag{10}$$

We define the mapping $\Psi : \Re^p \times \Re^{|\mathcal{A}|} \mapsto \Re^{|\mathcal{A}|}$ by

$$\Psi(\rho, \hat{f}) = \operatorname{Proj}_F(\hat{f}) - \hat{f} - t(\rho, \operatorname{Proj}_F(\hat{f})). \tag{11}$$

(This is a mapping that induces $\operatorname{Proj}_F(\hat{f})$ to become a traffic equilibrium by adjusting the value of β.) Thus, we recast the problem [MPEC-F] as the problem to

[P-F]

$$\text{minimize} \quad \Phi(\rho, \hat{f}) := \varphi\big(\rho, \Psi(\rho, \hat{f}), \operatorname{Proj}_F(\hat{f})\big), \tag{12a}$$

$$\text{subject to} \quad \rho \in P, \tag{12b}$$

$$\hat{f} \in \Re^{|\mathcal{A}|}. \tag{12c}$$

The corresponding parameterization of the normal cone operator N_H leads to a problem, [P-H], based on the equilibrium system [VIP-H]. We remark that the vector $\hat{f}$ does not correspond to a network flow in general, whereas its projection onto F, $\operatorname{Proj}_F(\hat{f})$, certainly does.

4.2. Properties of the Equivalent Problem

We next turn to look at the equivalence between the two problems [MPEC-F] and [P-F] and the basic properties of the latter.

First, consider any feasible triple $(\bar{\rho}, \bar{\beta}, \bar{f})$ in [MPEC-F], that is, a triple that satisfies (5c), $\hat{\varphi}(\bar{\rho}, \bar{\beta}) = \varphi(\bar{\rho}, \bar{\beta}, \bar{f})$, and is such that $\bar{\rho} \in P$. Let $\hat{f} := \bar{f} - t(\bar{\rho}, \bar{f}) - \bar{\beta}$. From the Minty parameterization, it follows immediately that $\bar{f} = \operatorname{Proj}_F(\hat{f})$ holds, and so from (11)

$$\Phi(\bar{\rho}, \hat{f}) = \varphi(\bar{\rho}, \bar{\beta}, \bar{f}) = \hat{\varphi}(\bar{\rho}, \bar{\beta}).$$

We conclude that every feasible solution to [MPEC-F] corresponds to a feasible solution to [P-F] with the same objective function value.

Second, consider any feasible pair $(\bar{\rho}, \hat{f})$ in [P-F], that is, one with $\bar{\rho} \in P$, and let the pair $(\bar{\beta}, \bar{f})$ be given by $\bar{\beta} := \Psi(\bar{\rho}, \hat{f})$ and $\bar{f} := \operatorname{Proj}_F(\hat{f})$. Then, again from the Minty parameterization, the triple $(\bar{\rho}, \bar{\beta}, \bar{f})$ satisfies (5c), and so

$$\hat{\varphi}(\bar{\rho}, \bar{\beta}) \le \varphi(\bar{\rho}, \bar{\beta}, \bar{f}) = \Phi(\bar{\rho}, \hat{f})$$

holds, where the inequality follows from the fact that the equilibrium flow $\bar{f}$ is not determined through any optimization over the set $S(\bar{\rho}, \bar{\beta})$. (Equality holds however if $S(\bar{\rho}, \bar{\beta})$ is a singleton set.) Because, for some values of the parameters $(\bar{\rho}, \bar{\beta})$, the two models may have different objective values, the two models are not equivalent in that sense. The optimistic approach is inherent in the setup of the problem [P-F], so embracing this approach in [MPEC-F] becomes necessary to achieve an equivalence at locally optimal solutions. We establish below that local minimizers of Φ do constitute local minimizers for $\hat{\varphi}$.

Proposition 4 (The Locally Optimal Solutions to [MPEC-F] and [P-F] Coincide). *The sets of constrained, locally optimal solutions to* [MPEC-F] *and* [P-F] *are the same.*

Proof. We establish that locally optimal solutions $(\rho^*, \hat{f}^*)$ to the problem [P-F] translate to locally optimal solutions (ρ^*, β^*, f^*) to [MPEC-F]. The converse follows immediately from the correspondence between feasible solutions of [MPEC-F] and [P-F] established above.

Consider a locally optimal solution $(\rho^*, \hat{f}^*)$ to the problem [P-F], and let the triple (ρ^*, β^*, f^*) be given by $\beta^* := \Psi(\rho^*, \hat{f}^*)$ and $f^* := \operatorname{Proj}_F(\hat{f}^*)$. Arguing by contradiction, we assume that there exists a triple $(\tilde{\rho}, \tilde{\beta}, \tilde{f})$, arbitrarily close to (ρ^*, β^*, f^*), and satisfying $\tilde{\rho} \in P$, (5c) and $\varphi(\tilde{\rho}, \tilde{\beta}, \tilde{f}) < \varphi(\rho^*, \beta^*, f^*)$.

Let $\bar{f} := \tilde{f} - t(\tilde{\rho}, \tilde{f}) - \tilde{\beta}$ (which likewise can be made arbitrarily close to $\hat{f}^*$). Through the Minty parameterization, we then obtain that

$$\begin{aligned}\Phi(\tilde{\rho}, \bar{f}) &:= \varphi(\tilde{\rho}, \Psi(\tilde{\rho}, \bar{f}), \mathrm{Proj}_F(\bar{f})) \\ &= \varphi(\tilde{\rho}, \tilde{\beta}, \tilde{f}) < \varphi(\rho^*, \beta^*, f^*) \\ &= \Phi(\rho^*, \hat{f}^*).\end{aligned}$$

We have therefore reached a contradiction to the local optimality of $(\rho^*, \hat{f}^*)$ in [P-F]. This completes the proof. □

This result has the interesting consequence that when we have a locally optimal solution at hand, $(\rho^*, \hat{f}^*)$, to the problem [P-F], a vector (ρ^*, β^*, f^*) that locally minimizes $\hat{\varphi}$ is (rather immediately) available. A final note on the nonuniqueness issue in association with the optimistic approach is the following: If we do not work under the assumption of an optimistic condition, then the optimal value $\hat{\varphi}(\rho^*, \beta^*)$ may not be achieved when the control (ρ^*, β^*) is implemented because the equilibrium solution f^* *actually* reached by the travellers then may be such that $\varphi(\rho^*, \beta^*, f^*) > \hat{\varphi}(\rho^*, \beta^*)$.

The existence of optimal solutions to [MPEC-F] (and, simultaneously, to [P-F]) follows from standard existence results for nonlinear programs. We first recall an abstract result; the corollary then translates it into our problem setting. The abstract (cooperative) MPEC problem is given as follows:

[MPEC]

$$\text{minimize} \quad \omega(x, y) \tag{13a}$$

$$\text{subject to} \quad (x, y) \in \mathscr{Z}, \tag{13b}$$

$$y \in S(x), \tag{13c}$$

where $\omega : \Re^n \times \Re^m \mapsto \Re \cup \{\infty\}$, $\mathscr{Z} \subseteq \Re^n \times \Re^m$, and $S : \Re^n \mapsto 2^{\Re^m}$.

We recall that a function $\omega : \Re^p \mapsto \Re \cup \{\infty\}$ is *proper* if $\omega(x) > -\infty$ for every $x \in \mathrm{dom}\, \omega$, and finite for at least one x. Let

$$\mathrm{lev}_\alpha\, \omega := \{x \in \Re^p \mid \omega(x) \leq \alpha\}$$

denote the *lower level set* of ω for the level α. We then say that ω is *weakly coercive* (or, *level-bounded*) if $\mathrm{lev}_\alpha\, \omega$ is bounded for every $\alpha \in \Re$. (This property is equivalent to $\lim_{\|x\| \to \infty} \omega(x) = \infty$.) Next, let the function δ_X denote the *indicator function* for $X \subseteq \Re^p$, that is, $\delta_X(x)$ equals zero if $x \in X$, and ∞ otherwise. We then say that the function ω is *inf-compact* relative to the set X if $\omega + \delta_X$ is lsc, proper, and weakly coercive. The below result extends the famous Weierstrass Theorem and has been established, for example, in Zhang (1994, Proposition 2.3).

Proposition 5 (Existence of Optimal Solutions to [MPEC]). *Let*

$$\mathrm{graph}\, S := \{(x, y) \in \Re^n \times \Re^m \mid y \in S(x)\}$$

denote the graph of the mapping S. Suppose that the objective function ω is inf-compact relative to the feasible set $\mathscr{Z} \cap \mathrm{graph}\, S$. Then, there exist globally optimal solutions to the problem [MPEC].

Corollary 6 (Existence of Optimal Solutions to [MPEC-F]). *Let $\mathscr{Z} := P \times \Re^{|\mathscr{A}|} \times \Re^{|\mathscr{A}|}$ and*

$$\begin{aligned}\mathrm{graph}\, S := \{\, (\rho, \beta, f) \in \Re^p \times \Re^{|\mathscr{A}|} \times \Re^{|\mathscr{A}|} \mid \\ -\, t(\rho, f) - \beta \in N_F(f)\,\}.\end{aligned} \tag{14}$$

Suppose that the function $\varphi : P \times \Re^{|\mathscr{A}|} \times \Re^{|\mathscr{A}|}$ is weakly coercive relative to the feasible set $\mathscr{Z} \cap \mathrm{graph}\, S$. Then, there exist globally optimal solutions to the problem [MPEC-F].

Proof. Because F is polyhedral and t is continuous on $\Re^p \times \Re^{|\mathscr{A}|}_+$, $\mathrm{graph}\, S$ is closed. Further, because $N_F(f)$ is nonempty for every $f \in F$, it is clear that for every pair $(\rho, f) \in P \times F$, we can choose a vector β such that $(\rho, \beta, f) \in \mathrm{graph}\, S$. Because P was assumed to be polyhedral, we may conclude that the feasible set $\mathscr{Z} \cap \mathrm{graph}\, S$ is closed as well as nonempty.

The function φ is in C^1 on $\mathscr{Z}$ and is hence both lsc and proper on the set $\mathscr{Z} \cap \mathrm{graph}\, S$. The property that remains to be ascertained to be able to invoke Proposition 5 is the weak coercivity of φ on $\mathscr{Z} \cap \mathrm{graph}\, S$, but this follows by the assumption. □

The weak coercivity assumption is rather natural. First, it is natural to assume that the controls ρ are confined to a bounded set P. (It is in particular true for Examples 1 and 2 in §3.3.) Second, it is also reasonable that the function φ is such that infinitely large (positive or negative) tolls β, or flows in infinite cycles, are discouraged in the minimization (as discussed in Example 3 in the same section). We remark that in the case of the problem [MPEC-H], the boundedness of H implies that φ is always weakly coercive in the flow space.

The problem [P-F] is obviously both nonsmooth and nonconvex. Because φ and t are smooth and the mapping Proj_F is piecewise affine (e.g., Rockafellar and Wets 1998, Proposition 12.30), [P-F] however is a problem of minimizing the piecewise smooth and, in particular, locally Lipschitz continuous and subdifferentially regular (Rockafellar and Wets 1998, Definition 7.25), hence semidifferentiable (or, B-differentiable) function Ψ over the polyhedral set $\mathscr{X}$, for which descent methods can be devised. We shall study such a method in the next section.

5. A Descent Algorithm for the Traffic Management Model

5.1. Introduction

The algorithm to be developed starting in this section produces a stationary point for the problem [P-F], which in this context is a feasible point $(\rho^*, \hat{f}^*)$ such that

$$\Phi'(\rho^*, \hat{f}^*; \delta\rho, \delta\hat{f}) \geq 0, \qquad \delta\rho \in T_P(\rho^*),\ \delta\hat{f} \in \Re^{|\mathscr{A}|},$$

where T_P denotes the tangent cone mapping for the set P.

Previous solution methods for traffic management models have mostly been heuristic (see, e.g., Migdalas 1995, Ferrari 1997, Larsson and Patriksson 1998). In some cases (e.g., Friesz et al. 1990, Yang and Lam 1996) a heuristic type of sensitivity analysis is applied to the solution $f(\rho, \beta)$ to (5c) to find profitable search directions for the implicit objective function $(\rho, \beta) \mapsto \varphi(\rho, \beta, f(\rho, \beta))$. Such a strategy—known as the *implicit approach*—as well as most of the other heuristics that have been proposed in the literature force one to solve for an equilibrium in each iteration, which is numerically challenging. It further presumes that the equilibrium link flow solution is uniquely determined, that is, that the link cost function $t(\rho, \cdot)$ is (at least) strictly monotone.

Other complications that may arise, and which in some cases have been ignored in the construction of descent algorithms based on the calculation of "gradients," stem from the fact that the equilibrium *commodity* flow (that is, commodity link flow or route flow) is not unique (cf. the discussion in §3.2) but, more importantly, does not necessarily satisfy the conditions for differentiability, especially not those based on satisfying the Wardrop conditions (1) with strict complementarity (that is, with "$>$" in (1b)); the calculus rules for the "directional derivatives" or "gradients" most often used include a procedure for the selection of a proper representative commodity flow in trying to achieve this. See Tobin and Friesz (1988), Qiu and Magnanti (1989), Outrata (1997) and Cho et al. (2000) for some such attempts, all of which require a positive definiteness property of the Jacobian of $t(\rho, \cdot)$ at the equilibrium.

In contrast, our scheme for calculating descent directions for Φ relies not on the solution of traffic equilibrium problems but rather on the (simpler) solution of strictly convex *quadratic* network flow problems (in fact, projections onto either flow polyhedra or subsets of their circulation flow subspaces), and no monotonicity requirements are made on the travel cost function.

We note that the reformulation of a variational inequality problem into a system of nonsmooth equations through the application of the projection operation has been utilized in sensitivity analyses of parametric nonlinear programs, in variational inequality problems (e.g., Robinson 1992, Luo et al. 1996, Pang and Ralph 1996 and references therein), and in algorithms for the solution of variational inequality problems (e.g., Ferris and Ralph 1995), but as far as we are aware the Minty parameterization has not previously been used directly to devise an algorithm for an MPEC problem.

In the next subsection, we investigate how to compute the directional derivative of Φ for the case where the underlying equilibrium problem is [VIP-F] or [VIP-H]. The main effort is to solve a strictly convex quadratic flow circulation problem over a subnetwork. Then, we consider the generation of a descent direction for Φ based, essentially, in the minimization of a quadratic regularization of this derivative over all feasible directions and discuss how its computation can be performed. The main complication here is that the directional derivative is only piecewise linear as a function of the search direction, and the search direction problem is in fact a *linear complementarity* (LCP) constrained strictly convex quadratic optimization problem. (In the case that an iteration point is differentiable, however, it reduces to a system of equations.) In the following section, we formalize the algorithm and establish its convergence to a stationary point of [P-F] and [P-H] under some additional, technical assumptions. The algorithm and its convergence conditions are adapted from Pang et al. (1991) (see also Luo et al. 1996, §§4.2 and 6.3). (Of course, other approaches are also possible to apply; we refer to the MPEC text books Luo et al. 1996, Bard 1998, and Outrata et al. 1998 for examples and references to other algorithms and to the last section for a brief discussion on one such example.) All along, we discuss simultaneously how to perform computations and how to meet the technical conditions in practice, thus leading to a realization of the algorithm.

5.2. Computation of Directional Derivatives

5.2.1. The Case of [P-F]**.** To compute the directional derivative of Φ at $(\rho, \hat{f})$ in the direction of $(\delta\rho, \delta\hat{f})$, we need to analyze the derivative of the projection operator Proj_F. In terms of $f := \mathrm{Proj}_F(\hat{f})$, define the set

$$\delta F := \{\delta f \in T_F(f) \mid \delta f \perp (f - \hat{f})\}.$$

This set is a subset of the flow circulation subspace for the multicommodity network, wherein some arcs are restricted in sign or direction for certain commodities. (Formulas for computing the cones $T_F(f)$ and $N_F(f)$ are discussed in detail in Patriksson 2002 and Patriksson and Rockafellar 2002 and will be used henceforth.)

According to Rockafellar and Wets (1998, Corollary 13.43), we have

$$\mathrm{Proj}'_F(\hat{f}; \delta\hat{f}) = \mathrm{Proj}_{\delta F}(\delta\hat{f}). \tag{15}$$

The directional derivative of Φ at $(\rho, \hat{f})$ in the direction of $(\delta\rho, \delta\hat{f})$ is then

$$\begin{aligned}
&\Phi'(\rho, \hat{f}; \delta\rho, \delta\hat{f}) \\
&\quad = \lim_{t\downarrow 0} \frac{1}{t}[\Phi(\rho + t\delta\rho, \hat{f} + t\delta\hat{f}) - \Phi(\rho, \hat{f})] \qquad (16a) \\
&\quad = \nabla_\rho \varphi(\rho, \beta, f)^{\mathrm{T}} \delta\rho + \nabla_\beta \varphi(\rho, \beta, f)^{\mathrm{T}} \delta\beta \\
&\qquad + \nabla_f \varphi(\rho, \beta, f)^{\mathrm{T}} \mathrm{Proj}_{\delta F}(\delta\hat{f}), \qquad (16b)
\end{aligned}$$

where β is given by the formula (10), and where

$$\begin{aligned}
\delta\beta &= \nabla_\rho \Psi(\rho, \hat{f})\delta\rho + \nabla_{\hat{f}} \Psi(\rho, \hat{f})\delta\hat{f} \qquad (16c) \\
&= -\nabla_\rho t(\rho, f)\delta\rho + \mathrm{Proj}_{\delta F}(\delta\hat{f}) \\
&\quad - \delta\hat{f} - \nabla_f t(\rho, f)\mathrm{Proj}_{\delta F}(\delta\hat{f}). \qquad (16d)
\end{aligned}$$

We note that the calculation of $\Phi'(\rho, \hat{f}; \delta\rho, \delta\hat{f})$ separates into simple calculations for each component of $\delta\rho$ and that it is linear in this vector. The calculation of $\mathrm{Proj}_{\delta F}(\delta\hat{f})$ is analyzed next.

Clearly, $\mathrm{Proj}_{\delta F}(\delta\hat{f})$ constitutes a strictly convex quadratic programming problem over a subset of the circulation subspace of the flow polytope. In detail, then, the following problem provides $\mathrm{Proj}_{\delta F}(\delta\hat{f})$:

$$\underset{z \in \Re^{|\mathcal{A}|}}{\text{minimize}}\ \tfrac{1}{2}\|z - \delta\hat{f}\|^2, \tag{17a}$$

subject to

$$A z_{pq} = 0, \qquad (p, q) \in \mathcal{C}, \tag{17b}$$

$$\sum_{(p,q)\in\mathcal{C}} z_{pq} - z = 0, \tag{17c}$$

$$z_{apq} \geq 0, \qquad a \in \mathcal{A}^0_{pq},\ (p, q) \in \mathcal{C}, \tag{17d}$$

$$z_{apq} = 0, \qquad a \in \mathcal{A}^>_{pq},\ (p, q) \in \mathcal{C},\ \text{with } f_a \neq \hat{f}_a, \tag{17e}$$

where, for each $(p, q) \in \mathcal{C}$, π^p_{pq} is the vector of multipliers for the constraints $Af_{pq} = d_{pq}$ in the definition of $\mathrm{Proj}_F(\hat{f})$, and $\mathcal{A}^0_{pq} := \{a = (i, j) \in \mathcal{A} \mid f_{apq} = 0 \text{ and } f_a - \hat{f}_a = \pi^p_{pqj} - \pi^p_{pqi}\}$, and $\mathcal{A}^>_{pq} := \{a = (i, j) \in \mathcal{A} \mid f_{apq} = 0 \text{ and } f_a - \hat{f}_a > \pi^p_{pqj} - \pi^p_{pqi}\}$ denote the set of links where the commodity flow is zero while, respectively, it lies on a shortest route, and it does not. (Note that $\hat{f}$, f, f_{pq}, $(p, q) \in \mathcal{C}$, and $\delta\hat{f}$ are all *given* $|\mathcal{A}|$-vectors and that the choice of disaggregate flow solution $(f_{pq})_{(p,q)\in\mathcal{C}}$ is immaterial to the definition of the feasible set of the problem (17) in terms of z, as established in Patriksson 2002 and Patriksson and Rockafellar 2002.)

5.2.2. The Case of [P-H]. In the case where [VIP-H] is the underlying traffic equilibrium model, we are instead interested in the calculation of $\text{Proj}_{\delta H}(\delta\hat{h})$, where $\delta H := \{\delta h \in T_H(h) \mid \delta h \perp (h - \hat{h})\}$. This projection separates over the different commodities $(p, q) \in \mathscr{C}$ to projection problems of the form $\text{Proj}_{\delta H_{pq}}(\delta\hat{h}_{pq})$ for the respective commodity flow polyhedron H_{pq}. To be precise, $\text{Proj}_{\delta H_{pq}}(\delta\hat{h}_{pq})$ is the unique solution to the problem to

$$\underset{z_{pq} \in \Re^{|\mathscr{R}_{pq}|}}{\text{minimize}} \ \tfrac{1}{2}\|z_{pq} - \delta\hat{h}_{pq}\|^2, \tag{18a}$$

subject to

$$\sum_{r \in \mathscr{R}_{pq}} z_{pqr} = 0, \tag{18b}$$

$$z_{pqr} \geq 0, \quad r \in \mathscr{R}^0_{pq}, \tag{18c}$$

$$z_{pqr} = 0, \quad r \in \mathscr{R}^>_{pq} \text{ with } h_{pqr} \neq \hat{h}_{pqr}, \tag{18d}$$

where π^p_{pq} is the Lagrange multiplier for the constraint $\sum_{r \in \mathscr{R}_{pq}} h_{pqr} = d_{pq}$ in the definition of $\text{Proj}_{H_{pq}}(\hat{h}_{pq})$, and $\mathscr{R}^0_{pq} := \{r \in \mathscr{R}_{pq} \mid h_{pqr} = 0 \text{ and } h_{pqr} - \hat{h}_{pqr} = \pi^p_{pq}\}$ and $\mathscr{R}^>_{pq} := \{r \in \mathscr{R}_{pq} \mid h_{pqr} = 0 \text{ and } h_{pqr} - \hat{h}_{pqr} > \pi^p_{pq}\}$ denote the set of routes where the flow is zero while, respectively, it is a shortest route, and it is not. (Note that $\hat{h}_{pq}$, h_{pq} and $\delta\hat{h}_{pq}$ all are *given* $|\mathscr{R}_{pq}|$-vectors.)

As we are not really focusing on merely calculating $\text{Proj}_{\delta F}(\delta\hat{f})$ (or, $\text{Proj}_{\delta H}(\delta\hat{h})$) for one fixed value of $\delta\hat{f}$ (or, $\delta\hat{h}$), we will not discuss their numerical computation by network optimization techniques, but refer to Qiu and Magnanti (1989), Patriksson (2002), and Patriksson and Rockafellar (2002) for numerical examples of related sensitivity problems.

We finally remark that $\text{Proj}_{\delta F}$ ($\text{Proj}_{\delta H}$) is piecewise linear in $\delta\hat{f}$ ($\delta\hat{h}$) (see also Rockafellar and Wets 1998, Proposition 12.30).

5.3. Computation of Descent Directions

5.3.1. The Case of [P-F]**.** To construct descent directions for Φ, we consider applying the techniques of Pang et al. for the constrained minimization of locally Lipschitz continuous functions. In our notation, given an iteration point $(\rho^\tau, \hat{f}^\tau)$, the search direction is found by solving the problem to

$$\text{minimize} \quad \Phi'(\rho^\tau, \hat{f}^\tau; \delta\rho, \delta\hat{f}) + (\delta\rho, \delta\hat{f})^T B^\tau (\delta\rho, \delta\hat{f}), \tag{19a}$$

$$\text{subject to} \quad \rho^\tau + \delta\rho \in P, \tag{19b}$$

$$\delta\hat{f} \in \Re^{|\mathscr{A}|}, \tag{19c}$$

where B^τ is a symmetric and positive definite matrix in $\Re^{(p+|\mathscr{A}|)\times(p+|\mathscr{A}|)}$. (Here, (19b) could be replaced with $\delta\rho \in T_P(\rho^\tau)$.) As we shall see, the problem (19) can be interpreted as an LCP constrained quadratic optimization problem corresponding to the minimization of a quadratically regularized linear approximation of the original problem [P-F].

As has already been established, the computation of $\Phi'(\rho, \hat{f}; \delta\rho, \delta\hat{f})$ separates into independent problems for $\delta\rho$ and $\delta\hat{f}$ and is furthermore linear and piecewise linear, respectively, in the respective arguments. From the standpoint of computational efficiency, this suggests choosing the matrix B^τ such that it is block-diagonal. On the other hand, the quality of the search direction suggests choosing the matrix so that a quasi-Newton-type method is produced, which would require it to contain second-order information about the function φ about the point $(\rho^\tau, \hat{f}^\tau)$. Until this conflict has been resolved by performing numerical tests, we henceforth assume that B^τ is block-diagonal, as discussed above. The problem (19) then separates into one problem for $\delta\rho$ that is strictly convex, quadratic, and linearly constrained and whose solution is straightforward in comparison with that of the part in $\delta\hat{f}$, and shall not be discussed further. The other separate problem is of the form

$$\underset{\delta\hat{f}}{\text{minimize}} \quad (\alpha^{1,\tau})^{\mathrm{T}} \text{Proj}_{\delta F}(\delta\hat{f}) + (\alpha^{2,\tau})^{\mathrm{T}} \delta\hat{f} + \delta\hat{f}^{\mathrm{T}} B^\tau_{\hat{f}} \delta\hat{f}, \tag{20}$$

where $\alpha^{i,\tau}$ are constant vectors in $\Re^{|\mathscr{A}|}$, $i = 1, 2$.

The difficulty of the problem (20) originates from the complementarity conditions arising from the inequalities (17d) that enter into the calculation of $\text{Proj}_{\delta F}(\delta\hat{f})$. To analyze this problem further, we study the characterization of the projection in (17), that is, the system

$$Az_{pq} = 0, \quad (p, q) \in \mathscr{C}, \tag{21a}$$

$$\sum_{(p,q) \in \mathscr{C}} z_{pq} - z = 0, \tag{21b}$$

$$z_{apq} - \delta\hat{f}_{apq} + (A^{\mathrm{T}}\pi_{pq})_a + v_a = 0,$$
$$a \in \mathcal{A}, (p, q) \in \mathcal{C}, \text{ with } f_{apq} > 0, \quad (21c)$$

$$0 \leq (z_{apq} - \delta\hat{f}_{apq} + (A^{\mathrm{T}}\pi_{pq})_a + v_a) \perp z_{apq} \geq 0,$$
$$a \in \mathcal{A}^0_{pq}, (p, q) \in \mathcal{C}, \quad (21d)$$

$$z_a = 0, \quad a \in \mathcal{A}^>_{pq}, (p, q) \in \mathcal{C}, \text{ with } f_a \neq \hat{f}_a, \quad (21e)$$

where π_{pq} and v are the Lagrange multipliers for the constraints (17b) and (17c), respectively.

Replacing the vector $\mathrm{Proj}_{\delta F}(\delta\hat{f})$ with the (unique) vector z, which solves this system, we obtain from (20) the LCP constrained quadratic program to

$$\underset{z,\,\delta\hat{f}}{\text{minimize}} \quad (\alpha^{1,\tau})^{\mathrm{T}} z + (\alpha^{2,\tau})^{\mathrm{T}}\delta\hat{f} + \delta\hat{f}^{\mathrm{T}} B^{\tau}_{\hat{f}} \delta\hat{f}, \quad (22a)$$

$$\text{subject to} \quad (21), \quad (22b)$$

$$\sum_{a\in\mathcal{A}} \delta\hat{f}_{pq} - \delta\hat{f} = 0. \quad (22c)$$

5.3.2. The Case of [P-H]**.** One main difference between the computations of descent directions for the two problems [P-F] and [P-H], is that in the latter the computations can be made to separate into computations for each commodity, by choosing the matrix B^{τ} to have a corresponding further block-diagonal structure with respect to $\mathcal{C}$. If this choice has been made, then the problem in $\delta\hat{h}$ separates into $|\mathcal{C}|$ problems of the form

$$\underset{\delta\hat{h}_{pq}}{\text{minimize}} \quad (\alpha^{1,\tau}_{pq})^{\mathrm{T}} \mathrm{Proj}_{\delta H_{pq}}(\delta\hat{h}_{pq}) + (\alpha^{2,\tau}_{pq})^{\mathrm{T}}\delta\hat{h}_{pq} + \delta\hat{h}^{\mathrm{T}}_{pq} B^{\tau}_{pq}\delta\hat{h}_{pq}, \quad (23)$$

where $\alpha^{i,\tau}_{pq}$ are constant vectors in $\Re^{|\mathcal{R}_{pq}|}$, $i = 1, 2$. To analyze this problem further, we develop the optimality conditions for the projection problem (18). The projection operation is characterized by a vector z_{pq} satisfying the conditions

$$\sum_{r\in\mathcal{R}_{pq}} z_{pqr} = 0, \quad (24a)$$

$$z_{pqr} - \delta\hat{h}_{pqr} + \lambda_{pq} = 0, \quad r \in \mathcal{R}_{pq} \text{ with } h^{\tau}_{pqr} > 0, \quad (24b)$$

$$0 \leq z_{pqr} - \delta\hat{h}_{pqr} + \lambda_{pq} \perp z_{pqr} \geq 0, \quad r \in \mathcal{R}^0_{pq}, \quad (24c)$$

$$z_{pqr} = 0, \quad r \in \mathcal{R}^>_{pq} \text{ with } h^{\tau}_{pqr} \neq \hat{h}^{\tau}_{pqr}, \quad (24d)$$

where λ_{pq} denotes the Lagrange multiplier for (18b). Replacing the vector $\mathrm{Proj}_{\delta H_{pq}}(\delta\hat{h}_{pq})$ with the (unique) vector z_{pq}, which solves this system, we obtain from (23) the LCP constrained quadratic program to

$$\underset{z_{pq},\,\delta\hat{h}_{pq}}{\text{minimize}} \quad (\alpha^{1,\tau}_{pq})^{\mathrm{T}} z_{pq} + (\alpha^{2,\tau}_{pq})^{\mathrm{T}}\delta\hat{h}_{pq} + \delta\hat{h}^{\mathrm{T}}_{pq} B^{\tau}_{pq}\delta\hat{h}_{pq}, \quad (25a)$$

$$\text{subject to} \quad (24). \quad (25b)$$

As (22), this is a nonconvex problem, but it is still possible to solve it efficiently given the realization of the algorithm. The reason why this problem is difficult lies in the presence of the complementarity constraints (24c). The number of such constraints is equal to the number of routes in the OD pair (p, q) for which the flow in h^{τ}_{pq} is degenerate, that is, belong to the set $\mathcal{R}^0_{pq}$. (This also tells us that when the projected flow is differentiable, that is, when the variables associated with the set $\mathcal{R}^0_{pq}$ can essentially be removed, then the descent direction is found through the solution of a system of nonlinear equations. The same is the case with the sets $\mathcal{A}^0_{pq}$ for the problem [P-F].) First, we note that in the course of the algorithm, not all the routes in $\mathcal{R}_{pq}$ will be known. We will be using a technique that has proved successful when solving traffic equilibrium problems (see, e.g., Larsson and Patriksson 1992, Patriksson 1994), wherein profitable routes in $\mathcal{R}_{pq}$ (that is, those for which $h^*_{pqr} > 0$ can be expected to hold) are generated algorithmically, through the solution of shortest route problems given temporarily fixed link costs. By also occasionally dropping previously generated routes that have received a zero flow during several consecutive iterations, the number of known routes that will give rise to the complementarity conditions (24c) will therefore be expected to be very low. Nevertheless, when such routes are present, we propose to deal with the situation when solving the problem (25) through a complete enumeration of the cases where the values of the corresponding variables z_{pqr} are zero or not. (The approach of enumerating the complementarity conditions in the LCP system was proposed in Pang et al. 1991.) Each of these restrictions of the problem (25) is a strictly convex, linearly constrained quadratic program.

In the event that B^τ does not separate over commodities, the corresponding problem in z will be stated over the entire set of known routes and over all the restrictions (24), but will still be a quite manageable problem to solve, because the constraints are separable over the commodities.

We note, finally, that numerical tolerances may need to be introduced in the systems (21) and (24) to determine when a route is considered to be used (24b) and when two scalars are considered to be equal (21e) and (24b)–(24d).

6. The Algorithm and Its Convergence

6.1. Descent Properties

Before stating the algorithm, we establish a technical lemma that motivates the use of the search direction finding problem (19), as well as the step length rule proposed. We state it, as well as the complete algorithm, for the case of the equilibrium problem being [VIP-F], but remark that the derivation for the case of [VIP-H] is analogous.

Lemma 7 (Descent Property). *Let $(\rho, \hat{f})$ be feasible in* [P-F]. *Let the matrix B be symmetric and positive definite. Then, the problem* (19) *has a globally optimal solution with a nonpositive optimal value. This value is moreover zero if and only if $(\delta\rho, \delta\hat{f}) = (0, 0)$ is the only global solution, which in turn is true if and only if $(\rho, \hat{f})$ is stationary in* [P-F].

If $(\delta\rho, \delta\hat{f})$ is nonzero, then for any $\sigma \in (0, 1)$ there exists a scalar $\bar{\ell} > 0$ such that for every $\ell \in [0, \bar{\ell}]$,

$$\Phi(\rho + \ell\delta\rho, \hat{f} + \ell\delta\hat{f}) - \Phi(\rho, \hat{f}) \le -\frac{\sigma}{2}\ell(\delta\rho, \delta\hat{f})^{\mathrm{T}} B(\delta\rho, \delta\hat{f}) \tag{26}$$

holds.

Proof. As has been established previously, the function Φ is locally Lipschitz continuous on $P \times \Re^{|\mathcal{A}|}$. So, for all $(\delta\rho, \delta\hat{f})$ with $\rho + \delta\rho \in P$,

$$|\Phi'(\rho, \hat{f}; \delta\rho, \delta\hat{f})| \le M_\Phi \|(\delta\rho, \delta\hat{f})\| \tag{27}$$

holds (cf. Luo et al. 1996, Proposition 4.2.2.b), where $M_\Phi > 0$ is the modulus of Lipschitz continuity at $(\rho, \hat{f})$. Further, because B is positive definite, there exists an $m > 0$ such that

$$\Phi'(\rho, \hat{f}; \delta\rho, \delta\hat{f}) + \frac{1}{2}(\delta\rho, \delta\hat{f})^{\mathrm{T}} B(\delta\rho, \delta\hat{f}) \ge -M_\Phi \|(\delta\rho, \delta\hat{f})\| + \frac{m}{2}\|(\delta\rho, \delta\hat{f})\|^2.$$

Hence, if $\|(\delta\rho, \delta\hat{f})\|$ tends to infinity, then so does the objective of (19), which thus is weakly coercive. The feasible set of (19) being closed as well as nonempty, the problem therefore has a globally optimal solution. Because the objective value is zero at zero, it must further have a nonpositive optimal value.

Assume next that $(\delta\rho, \delta\hat{f})$ is nonzero. Then,

$$\Phi'(\rho, \hat{f}; \delta\rho, \delta\hat{f}) < -\frac{1}{2}(\delta\rho, \delta\hat{f})^{\mathrm{T}} B(\delta\rho, \delta\hat{f}). \tag{28}$$

Suppose that a positive $\bar{\ell}$, such that (26) is satisfied, does not exist. Then there must be a sequence $\Re_{++} \supset \{\ell_s\} \to 0$ such that for each s,

$$\Phi(\rho + \ell_s\delta\rho, \hat{f} + \ell_s\delta\hat{f}) - \Phi(\rho, \hat{f}) > -\frac{\sigma}{2}\ell_s(\delta\rho, \delta\hat{f})^{\mathrm{T}} B(\delta\rho, \delta\hat{f}).$$

Dividing the inequality by ℓ_s and letting $s \to \infty$ then yields that

$$\Phi'(\rho, \hat{f}; \delta\rho, \delta\hat{f}) \ge -\frac{\sigma}{2}(\delta\rho, \delta\hat{f})^{\mathrm{T}} B(\delta\rho, \delta\hat{f}).$$

But this contradicts (28), as $\sigma < 1$.

It remains to establish that $(\delta\rho, \delta\hat{f}) = (0, 0)$ is equivalent to $(\rho, \hat{f})$ being stationary. Assume first that $(\delta\rho, \delta\hat{f}) = (0, 0)$. Then, for all $(\delta\bar{\rho}, \delta\bar{f})$ and $\lambda > 0$ with $\rho + \lambda\delta\bar{\rho} \in P$,

$$0 \le \Phi'(\rho, \hat{f}; \lambda\delta\bar{\rho}, \lambda\delta\bar{f}) + \lambda^2(\delta\bar{\rho}, \delta\bar{f})^{\mathrm{T}} B(\delta\bar{\rho}, \delta\bar{f}).$$

Dividing the inequality by λ and letting it tend to zero then establishes that $(\rho, \hat{f})$ indeed is stationary. Conversely, if $(\rho, \hat{f}) \neq (0, 0)$, then the inequality (28) shows that $\Phi'(\rho, \hat{f}; \delta\rho, \delta\hat{f}) < 0$, which contradicts stationarity. □

6.2. The Algorithm

We are now ready to state the algorithm; it is given in Table 1.

Table 1 A Descent Algorithm

0. (*Initialization*): Choose an initial point $(\rho^0, \hat{f}^0) \in P \times \Re^{|\mathcal{A}|}$, let $\gamma, \sigma \in (0,1)$ be given, and let $\tau := 0$.

1. (*Search direction generation*): Let B^τ be a positive definite, symmetric matrix. Let $(\delta\rho^\tau, \delta\hat{f}^\tau)$ be a solution to (19). If the optimal value of (19) is zero, then terminate with $(\rho^\tau, \hat{f}^\tau)$ being a stationary point in [P-F]. Otherwise, continue.

2. (*Armijo line search*): Let i_τ be the smallest nonnegative integer i such that

$$\Phi(\rho^\tau + \gamma^i \delta\rho^\tau, \hat{f}^\tau + \gamma^i \delta\hat{f}^\tau) - \Phi(\rho^\tau, \hat{f}^\tau) \le -\frac{\sigma}{2}\gamma^i (\delta\rho^\tau, \delta\hat{f}^\tau)^{\mathrm{T}} B^\tau (\delta\rho^\tau, \delta\hat{f}^\tau). \tag{29}$$

The step length is $\ell_\tau = \gamma^{i_\tau}$. Let $(\rho^{\tau+1}, \hat{f}^{\tau+1}) := (\rho^\tau, \hat{f}^\tau) + \ell_\tau(\delta\rho^\tau, \delta\hat{f}^\tau)$.

3. (*Termination criterion and iteration*): If $(\rho^\tau, \hat{f}^\tau)$ is acceptable $\to$ Stop. Otherwise, go to Step 1 with $\tau := \tau + 1$.

We remark that the Armijo line search is only one among a large variety of step length rules that may be employed in the scheme of Table 1. For example, the algorithm of Pang et al. (1991) employs a *nonmonotone line search*, first analyzed in Grippo et al. (1991), whose mechanism allows for unit steps to often be taken to speed up convergence.

6.3. Convergence Conditions

When establishing the convergence of the algorithm, we must first introduce of all an assumption on the choice of the matrices B^τ such that they are bounded and uniformly positive definite:

$$\exists\, m, M > 0 : m\|d\|^2 \le d^{\mathrm{T}} B^\tau d \le M\|d\|^2, \qquad d \in \Re^{p+|\mathcal{A}|}. \tag{30}$$

The property of weak coercivity of φ, which was introduced in Corollary 6, ensures that the sequence $\{(\rho^\tau, \hat{f}^\tau)\}$ is bounded. However, we note that the often-used condition for $(\rho^0, \hat{f}^0) \in P \times \Re^{|\mathcal{A}|}$,

$$\text{the set } \{(\rho, \hat{f}) \in P \times \Re^{|\mathcal{A}|} \mid \Phi(\rho, \hat{f}) \le \Phi(\rho^0, \hat{f}^0)\} \text{ is bounded,} \tag{31}$$

is implied by weak coercivity and is enough to guarantee both the existence of a globally optimal solution to the problem [P-F] and the boundedness of the sequence $\{(\rho^\tau, \hat{f}^\tau)\}$.

Finally, when using the analysis of Pang et al. (1991) we need to assume that *the function* Φ *is differentiable at the limit point*. This is a restrictive assumption in general, and we will now explain how we try to enforce it through our realization of the algorithm.

6.3.1. The Case of [P-F]**.** We first observe that a nondifferentiability of Φ at a limit point $(\rho^\infty, \hat{f}^\infty)$ would be caused by a nondifferentiability of the projection operation Proj_F at $\hat{f}^\infty$. Although Proj_F is, as remarked before, piecewise linear, it may have kinks where the active constraints change. The condition that Φ is differentiable at $(\rho^\infty, \hat{f}^\infty)$ is equivalent to the condition that the variables δf_{apq} associated with the set $\mathcal{A}^0$ would be zero in the characterization (21) for any choice of perturbations in (ρ, β), and therefore also zero in the solution to the direction-finding problem (22). This condition is satisfied in particular if

$$f^\infty \text{ is strictly complementary with respect to the cost } f^\infty - \hat{f}^\infty. \tag{32}$$

Although it is a restrictive assumption, given f^∞, the condition (32) can be checked, by solving an entropy maximization problem over the vectors f_{pq}; cf. Akamatsu (1997), Patriksson (2002).

6.3.2. The Case of [P-H]**.** In the case of the problem [P-H], differentiability of Φ at a limit point $(\rho^\infty, \hat{h}^\infty)$ is equivalent to the condition that the variables associated with the set $\mathcal{R}^0_{pq}$ essentially can be removed because the corresponding values δh_{pqr} in the sensitivity problem (24) would be zero for any choice of $(\delta\rho, \delta\beta)$, and therefore also in the optimal solution to (25). This condition is satisfied in particular if

$$h^\infty \text{ is strictly complementary with respect to the cost } h^\infty - \hat{h}^\infty. \tag{33}$$

In this case, then, the concern is which routes r in the sets $\mathcal{R}_{pq}$, $(p, q) \in \mathcal{C}$, will have a positive flow at $h^\infty := \mathrm{Proj}_H(\hat{h}^\infty)$, and whether we will be able to identify them finitely. We propose to deal with the situation as follows. As remarked before, not all routes will ever be known. Instead, we will solve the shortest route problems at regular intervals with link costs based on the current variable values and include the shortest routes in the set of variables. This way, only subsets $\widehat{\mathcal{R}}^\tau_{pq} \subset \mathcal{R}_{pq}$ of the routes will be known at any given iteration τ. In the course of the algorithm, some of these routes will also receive a zero flow. Such routes will be identified and removed from the corresponding sets $\widehat{\mathcal{R}}^{\tau}_{pq}$. (There is neither a guarantee that a generated route will always retain a positive flow, nor that a discarded route will not be regenerated.) If we keep only routes in the subsets $\widehat{\mathcal{R}}^{\tau}_{pq}$ having positive flow, then we will actually be visiting differentiable points of the corresponding restrictions of the problem [P-F] to the subsets $\widehat{\mathcal{R}}^{\tau}_{pq}$. For these restrictions, our algorithm will act as a scaled gradient projection algorithm. Interestingly, it is known that the active constraints at the limit point of a projection algorithm used to minimize a differentiable function over a polyhedral set will be identified after a finite number of iterations (e.g., the surveys in Patriksson (1998) and Bertsekas (1999)). The following assumption that we make is therefore not so far-fetched: We assume that the route generation and deletion process is such that after a finite number of iterations, the subsets $\widehat{\mathcal{R}}^{\tau}_{pq}$ stay constant and no route in these subsets has a zero flow at h^{∞}. (This way, we are guaranteed to obtain a stationary point to the restriction of the original problem [P-H] to the sets of routes that are retained in the limit.) In other words,

$$\exists\, \bar{\tau}: \widehat{\mathcal{R}}^{\tau}_{pq} = \widehat{\mathcal{R}}_{pq} \subset \mathcal{R}_{pq},\ \tau \geq \bar{\tau}, \quad \text{and}$$
$$h^{\infty}_{pqr} > 0, \quad r \in \widehat{\mathcal{R}}_{pq}, \quad (p,q) \in \mathcal{C}. \tag{34}$$

6.4. Convergence Theorem

We now state and prove the main result of this paper.

Theorem 8 (Convergence of the Algorithm). *Consider the problem* [P-F]. *Suppose that the assumptions* (30) *and* (31) *hold. Let* $\{(\rho^{\tau}, \hat{f}^{\tau})\}$ *be the sequence of iterates produced by the algorithm of Table* 1. *Then, the sequence* $\{(\rho^{\tau}, \hat{f}^{\tau})\}$ *is bounded. Let* $(\rho^{\infty}, \hat{f}^{\infty})$ *denote any of its accumulation points. If* $\hat{f}^{\infty}$ *satisfies* (32), *then* $(\rho^{\infty}, \hat{f}^{\infty})$ *is a stationary point for the problem* [P-F]. *Further, if the sequence* $\{\ell_{\tau}\}$ *of step lengths is bounded away from zero, then each limit point of the sequence* $\{(\rho^{\tau}, \hat{f}^{\tau})\}$ *is stationary in* [P-F] *even without the assumption* (32).

Proof. The sequence $\{(\rho^{\tau}, \hat{f}^{\tau})\}$ satisfies

$$\Phi(\rho^{\tau+1}, \hat{f}^{\tau+1}) \leq \Phi(\rho^{\tau}, \hat{f}^{\tau}) - \frac{\sigma}{2}\ell_{\tau}(\delta\rho^{\tau}, \delta\hat{f}^{\tau})^{\mathrm{T}} B^{\tau}(\delta\rho^{\tau}, \delta\hat{f}^{\tau})$$
$$< \Phi(\rho^{\tau}, \hat{f}^{\tau}),$$

so the sequence $\{\Phi(\rho^{\tau}, \hat{f}^{\tau})\}$ is strictly decreasing. By the assumption (31), Φ is lower bounded on $P \times \Re^{|\mathcal{A}|}$, whence $\{\Phi(\rho^{\tau}, \hat{f}^{\tau})\}$ converges. Then, by the above inequality, also $\{\ell_{\tau}(\delta\rho^{\tau}, \delta\hat{f}^{\tau})^{\mathrm{T}} B^{\tau}(\delta\rho^{\tau}, \delta\hat{f}^{\tau})\} \to 0$ holds. That the sequence $\{(\rho^{\tau}, \hat{f}^{\tau})\}$ is bounded also follows from the assumption (31).

Let $\{(\delta\rho^{\tau}, \delta\hat{f}^{\tau})\}$ be the sequence of solutions to (19) generated by the algorithm. We next establish that this sequence is bounded. By the assumption (30), for each τ we have that

$$|\Phi'(\rho^{\tau}, \hat{f}^{\tau}; \delta\rho^{\tau}, \delta\hat{f}^{\tau})| \geq \frac{1}{2}(\delta\rho^{\tau}, \delta\hat{f}^{\tau})^{\mathrm{T}} B^{\tau}(\delta\rho^{\tau}, \delta\hat{f}^{\tau})$$
$$\geq \frac{m}{2}\|(\delta\rho^{\tau}, \delta\hat{f}^{\tau})\|^2.$$

Moreover, $|\Phi'(\rho^{\tau}, \hat{f}^{\tau}; \delta\rho^{\tau}, \delta\hat{f}^{\tau})| \leq M_{\Phi}\|(\delta\rho^{\tau}, \delta\hat{f}^{\tau})\|$ holds by (27), so $\|(\delta\rho^{\tau}, \delta\hat{f}^{\tau})\| \leq \frac{2M_{\Phi}}{m}$. This implies that both the sequences $\{(\delta\rho^{\tau}, \delta\hat{f}^{\tau})\}$ and $\{\Phi'(\rho^{\tau}, \hat{f}^{\tau}; \delta\rho^{\tau}, \delta\hat{f}^{\tau})\}$ are bounded.

We next turn to a subsequence $\mathcal{T}$ for which we assume that $\{\rho^{\tau}, \hat{f}^{\tau}\}_{\tau\in\mathcal{T}} \to (\rho^{\infty}, \hat{f}^{\infty})$, $\{\delta\rho^{\tau}, \delta\hat{f}^{\tau}\}_{\tau\in\mathcal{T}} \to (\delta\rho^{\infty}, \delta\hat{f}^{\infty})$, and $\{B^{\tau}\}_{\tau\in\mathcal{T}} \to B^{\infty}$. By the assumption (30), B^{∞} is symmetric and positive definite. We may further assume that $\{\Phi'(\rho^{\tau}, \hat{f}^{\tau}; \delta\rho^{\tau}, \delta\hat{f}^{\tau})\}_{\tau\in\mathcal{T}}$ has a limit.

Suppose now that the sequence $\{\ell_{\tau}\}$ of step lengths does not tend to zero, that is, $\liminf_{\tau\to\infty} \ell_{\tau} > 0$. It then immediately follows that $\{(\delta\rho^{\tau}, \delta\hat{f}^{\tau})^{\mathrm{T}} B^{\tau}(\delta\rho^{\tau}, \delta\hat{f}^{\tau})\} \to 0$, whence $(\delta\rho^{\infty}, \delta\hat{f}^{\infty}) = (0,0)$ must hold by the positive definiteness of B^{∞}. Now, because $(\delta\rho^{\tau}, \delta\hat{f}^{\tau})$ is optimal in (19),

$$\Phi'(\rho^{\tau}, \hat{f}^{\tau}; \delta\rho^{\tau}, \delta\hat{f}^{\tau}) + \frac{1}{2}(\delta\rho^{\tau}, \delta\hat{f}^{\tau})^{\mathrm{T}} B^{\tau}(\delta\rho^{\tau}, \delta\hat{f}^{\tau})$$
$$\leq \Phi'(\rho^{\tau}, \hat{f}^{\tau}; \delta\rho, \delta\hat{f}) + \frac{1}{2}(\delta\rho, \delta\hat{f})^{\mathrm{T}} B^{\tau}(\delta\rho, \delta\hat{f})$$

holds for all $(\delta\rho, \delta\hat{f})$ for which $\rho^{\tau} + \delta\rho \in P$. Taking limits on both sides of the inequality, we first note that $\{(\delta\rho^{\tau}, \delta\hat{f}^{\tau})\}_{\tau\in\mathcal{T}} \to (0,0)$, so the left-hand side tends to zero. Further,

$$\limsup_{\tau\in\mathcal{T}} \Phi'(\rho^{\tau}, \hat{f}^{\tau}; \delta\rho, \delta\hat{f}) \leq \Phi'(\rho^{\infty}, \hat{f}^{\infty}; \delta\rho, \delta\hat{f}),$$

so we find that

$$0 \leq \Phi'(\rho^{\infty}, \hat{f}^{\infty}; \delta\rho, \delta\hat{f}) + \frac{1}{2}(\delta\rho, \delta\hat{f})^{\mathrm{T}} B^{\infty}(\delta\rho, \delta\hat{f})$$

holds for all $(\delta\rho, \delta\hat{f})$ for which $\rho^\tau + \delta\rho \in P$. As in the proof of Lemma 7, we may conclude that $(\rho^\infty, \hat{f}^\infty)$ is stationary.

Assume now instead that $\liminf_{\tau\to\infty} \ell_\tau = 0$. Then it must be that

$$\Phi(\rho^\tau + (\ell_\tau/\gamma)\delta\rho^\tau, \hat{f}^\tau + (\ell_\tau/\gamma)\delta\hat{f}^\tau) - \Phi(\rho^\tau, \hat{f}^\tau) > -\frac{\sigma}{2}(\ell_\tau/\gamma)(\delta\rho^\tau, \delta\hat{f}^\tau)^{\mathrm{T}} B^\tau (\delta\rho^\tau, \delta\hat{f}^\tau)$$

holds for all sufficiently large τ. Dividing the inequality by ℓ_τ/γ and taking the limit on both sides in $\mathcal{T}$ yields

$$\Phi'(\rho^\infty, \hat{f}^\infty; \delta\rho^\infty, \delta\hat{f}^\infty) \geq -\frac{\sigma}{2}(\delta\rho^\infty, \delta\hat{f}^\infty)^{\mathrm{T}} B^\infty (\delta\rho^\infty, \delta\hat{f}^\infty). \tag{35}$$

On the other hand, from (28) it follows that

$$\lim_{\tau\in\mathcal{T}} \Phi'(\rho^\tau, \hat{f}^\tau; \delta\rho^\tau, \delta\hat{f}^\tau) \leq -\frac{1}{2}(\delta\rho^\infty, \delta\hat{f}^\infty)^{\mathrm{T}} B^\infty (\delta\rho^\infty, \delta\hat{f}^\infty). \tag{36}$$

By the assumption (32), Φ is differentiable at $(\rho^\infty, \hat{f}^\infty)$. Together with the fact that $\sigma < 1$ holds, the combination of the inequalities (35) and (36) implies that $(\delta\rho^\infty, \delta\hat{f}^\infty) = (0, 0)$ must hold. We conclude that also under this circumstance, $(\rho^\infty, \hat{f}^\infty)$ is stationary in [P-*F*]. □

For the case of the problem [P-*H*], the only difference in the analysis is that the two assumptions (30) and (31) need to be translated into route flow space, and (32) needs to be replaced by (33). Further, if we replace the assumption (33) with (34), then $\hat{h}^\infty$ will be of a dimension corresponding to the number of routes kept after iteration $\bar{\tau}$, and it will then be stationary for the restriction of the problem [P-*H*] to that corresponding subset of routes.

6.5. Comparison with Previous Work

Some comments on the convergence result are in order. Most sensitivity-based heuristics for solving bilevel optimization problems in transportation planning (e.g., Friesz et al. 1990) assume that the implicit function [in our case, $(\rho, \beta) \mapsto \varphi(\rho, \beta, f(\rho, \beta))$] is differentiable everywhere. In contrast, the method of Pang et al. (1991) applied to the minimization of this function assumes its differentiability only at the limit point. The assumptions that we make in our convergence result (a strictly complementary limit point) is slightly stronger than necessary, but checking for differentiability is not trivial. (Further characterizations that may be computationally viable to check are given in Patriksson 2002.) Note also that our objective function is Φ, which does not involve the lower-level equilibrium problem, and ultimately does not require, in contrast to the other developments mentioned above, that the travel cost function is strictly or strongly monotone. Further, also in the implicit approach one uses LCP constrained quadratic programs to search for improving directions (see, e.g., Pang et al. 1991, Luo et al. 1996, Outrata et al. 1998).

The main differences between the two approaches seem therefore to be (1) the present one does not require the numerical calculation of Wardrop equilibria in each iteration and (2) the strict (or, strong) monotonicity requirements have been eliminated. These improvements, seemingly, have something more than a theoretical value because they show that numerical approaches can be devised for a larger class of bilevel problems in traffic management and control, and the calculations can, also seemingly, be made simpler at the same time. The price to be paid for these improvements is the introduction of the differentiability assumptions (32) and (33) for the projection operation at the limit point, which we however hope to be able to relax.

7. Extensions and Further Research

We discuss in brief two extensions of the traffic equilibrium models [VIP-*F*] and [VIP-*H*] in the context of the traffic management models [MPEC-*F*] and [MPEC-*H*].

7.1. Elastic Demands

Consider an extension of the model [VIP-*H*] in which the demand is not fixed, but it is given by a function of the cheapest-route costs. Specifically, we assume that the demands are given by functions $g_{pq} : \Re^{|\mathcal{C}|} \mapsto \Re_+$ that are nonnegative, upper bounded, and continuous on $\Re^{|\mathcal{C}|}$ for each $(p, q) \in \mathcal{C}$.

The elastic demand extension of [VIP-H] then is

$$-[\bar{c}(h)-\Gamma\pi,\Gamma^{\mathrm{T}}h-g(\pi)]\in N_{\Re_+^{|\mathcal{R}|}}(h)\times N_{\Re_+^{|\mathcal{C}|}}(\pi). \qquad \text{[EVIP-}H\text{]}$$

Under the additional assumption that $-g$ is strictly monotone on $\Re^{|\mathcal{C}|}$, $-g$ is also maximal Rockafellar and Wets (1998, Example 12.7), and the problem can further be written as

$$[-\bar{c}(h),g^{-1}(d)]\in N_{H_d}(h,d), \qquad \text{[EVIP-}H_d\text{]}$$

where

$$H_d:=\left\{(h,d)\in\Re^{|\mathcal{R}|+|\mathcal{C}|}\mid \exists\, h\in\Re_+^{|\mathcal{R}|} \text{ with } \Gamma^{\mathrm{T}}h=d\right\}.$$

For overviews of these models, see Nagurney (1993) and Patriksson (1994).

The corresponding extension of [MPEC-H] becomes

[MPEC-H_d]

$$\text{minimize} \quad \varphi(\rho,\beta,h,d) \tag{37a}$$

$$\text{subject to} \quad \rho\in P, \tag{37b}$$

$$[-c(\rho,h)-\beta,g^{-1}(d)]\in N_{H_d}(h,d). \tag{37c}$$

The analogous Minty parameterization that leads to [P-F] here leads to the problem to

[P-H_d]

$$\text{minimize} \quad \Phi(\rho,\hat{h},\hat{d}):=\varphi(\rho,\Psi(\rho,\hat{h},\hat{d}),\Upsilon(\hat{h},\hat{d})), \tag{38a}$$

$$\text{subject to} \quad \rho\in P, \tag{38b}$$

$$\hat{h}\in\Re^{|\mathcal{R}|}, \tag{38c}$$

$$\hat{d}\in\Re^{|\mathcal{C}|}, \tag{38d}$$

where

$$\Psi(\rho,\hat{h},\hat{d}):=\mathrm{Proj}^h_{H_d}(\hat{h},\hat{d})-\hat{h}-c(\rho,\mathrm{Proj}^h_{H_d}(\hat{h},\hat{d})), \tag{38e}$$

$$\Upsilon(\hat{h},\hat{d})):=\hat{d}-g^{-1}(\mathrm{Proj}^d_{H_d}(\hat{h},\hat{d})), \tag{38f}$$

and $(\mathrm{Proj}^h_{H_d},\mathrm{Proj}^d_{H_d})=\mathrm{Proj}_{H_d}$ is the representation of the projection operator in the two components h and d. We remark that this problem has properties similar to [P-H], and the algorithm given extends readily to [P-H_d].

7.2. Queueing Models

The presence of signal controls in the network are sometimes modelled by means of capacity constraints. In such a circumstance, we would introduce in the equilibrium model a set of side constraints of the form

$$s_\kappa(h)\le 0, \qquad \kappa\in\mathcal{K}, \tag{39}$$

where $\mathcal{K}$ is a finite index set (for example, formed by subsets of $\mathcal{N}$, $\mathcal{A}$, $\mathcal{C}$, and $\mathcal{R}$) and $s_\kappa:\Re_+^{|\mathcal{R}|}\mapsto\Re$, $\kappa\in\mathcal{K}$ is continuously differentiable on $\Re_+^{|\mathcal{R}|}$. The flows that satisfy (39) form a closed set in $\Re^{|\mathcal{R}|}$, which we denote by G.

Note that the constraints are given in terms of route flows without any loss of generality, because a constraint on the total link flows, can be written as $s_\kappa(\Lambda h)\le 0$ for example.

Example 9 (Capacity Constraints and Signal Control). The most immediate example of a set of flow restrictions is that of upper bounds on some links' flows. In the framework of (39), such constraints are described by letting $\mathcal{K}:=\overline{\mathcal{A}}$ and

$$s_a(f_a):=f_a-u_a, \qquad a\in\overline{\mathcal{A}}, \tag{40}$$

where $\overline{\mathcal{A}}\subseteq\mathcal{A}$ and $u_a\ge 0$ are the upper bound on the flow on link a. In the context of traffic signals, the constant u_a may be regarded as the upper bound on the traffic that can pass link a during the green-time period. See Miller et al. (1975), Smith and Van Vuren (1993), Ferrari (1997), and Larsson and Patriksson (1994, 1995, and 1999).

When viewing the traffic management problem as a hierarchical problem, in the example discussed above, the side constraints are part of the equilibrium problem and are therefore *lower-level* constraints. For example, the equilibrium problem [VIP-H] is then extended to include the additional side constraints (39), creating the inclusion

$$-\bar{c}(h)\in N_{H\cap G}(h). \qquad \text{[SCVIP-}H\text{]}$$

The corresponding changes in [P-H] and in the algorithm lie only in the additional constraints' appearance in the projection formulae, where H is everywhere to be replaced by $H\cap G$. (A similar effect is obtained if pure link-flow-side constraints are added to the model [VIP-F].) A main additional complication that this extension causes is a more complex construction of descent directions, as the set G influences the form that the projection $\text{Proj}_{\delta(F\cap G)}(\delta\hat{f})$ (respectively, $\text{Proj}_{\delta(H\cap G)}(\delta\hat{h})$) takes. Given that the form of G allows descent directions to still be calculated (for example, the formula (15) also applies to the set $F\cap G$ as long as G is polyhedral), the algorithm would work as it now stands, and the only issue remaining to be made on its convergence would be on the differentiability of the operator $\text{Proj}_{F\cap G}$ (respectively, $\text{Proj}_{H\cap G}$) at the limit point, which again would rely on the form of G.

However, side constraints of the form (39) can also be imposed upon the travellers by means of some decentralized control measure, such as a tax. In such a case, the side constraints would not be part of the equilibrium problem but instead would be a set of constraints on the equilibrium flows (perhaps also as joint constraints with the control variables) and as such would be placed as *upper-level* constraints. The presence of lower-level variables in upper-level constraints could however in general complicate the problem immensely because the feasible set becomes much more complex (see, e.g., Luo et al. 1996). Therefore, in the MPEC models [MPEC-F] and [MPEC-H], such side constraints would be assumed to be represented as smooth penalties in the upper-level objective function φ.

7.3. Further Research

Interesting further research topics fall into several categories, some of which go far beyond the present application. Most urgent is perhaps the (already mentioned) reduction of the assumption of differentiability of the projection operator at the limit point of the sequence of iterates produced by the algorithm. The problem [P-F] constructed by the use of the Minty parameterization can of course be solved by other algorithms, some of which may very well produce stationary points to the problem [MPEC-F] under milder conditions. For example, *bundle algorithms* can be devised along the lines described in Schramm and Zowe (1992), Kiwiel (1996), Outrata et al. (1998), and Mäkelä et al. (1999) for the minimization of locally Lipschitzian and upper semidifferentiable functions—the latter condition of which, for a locally Lipschitzian function $\omega : \Re^p \mapsto \Re$, requires in addition that for any $x, d \in \Re^p$ and sequences $\{g^t\}$ and $\{\ell_t\}$ with $g^t \in \partial\omega(x+\ell_t d)$ and $\{\ell_t\} \downarrow 0$ that $\limsup_{t\to\infty} (g^t)^{\mathrm{T}} d \geq \liminf_{t\to\infty} [\omega(x+\ell_t d) - \omega(x)]/\ell_t$, cf. Bihain (1984)—based on the generation of arbitrary subgradients of the function Φ (see Outrata and Zowe 1995, Dempe and Vogel 1997, and Outrata et al. 1998 for their computations). Bundle algorithms can certainly be expected to be viable for the present problem, although upper semidifferentiability does not hold everywhere; in any case, global upper semidifferentiability is, perhaps arguably, milder than differentiability at the limit point. We note that the subgradient formula given in Patriksson (2002) amounts to solving a sequence of affine traffic equilibrium models, thus avoiding, at least partly, the combinatorial nature of the present algorithm.

From a wider perspective, the Minty parameterization is likely to also be beneficial for the solution of other MPEC problems, where the type of parameterization utilized in the problem [MPEC-F] makes sense, as well as for special cases of MPEC, such as mathematical programs with complementarity constraints (MPCC, see, e.g., Scheel and Scholtes 2000). Another interesting technical question surrounding the utilization of the Minty parameterization concerns the associated optimality conditions of first and second order, which may have interesting properties and consequences in their own right for the equivalent MPEC problem.

References

Akamatsu, T. 1997. Decomposition of path choice entropy in general transport networks. *Transportation Sci.* **31** 349–362.

Bard, J. F. 1998. *Practical Bilevel Optimization: Algorithms and Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

——, J. E. Falk. 1982. An explicit solution to the multi-level programming problem. *Comput. Oper. Res.* **9** 77–100.

Bergendorff, P., D. W. Hearn, M. V. Ramana. 1996. Congestion toll pricing of traffic networks. P. Pardalos, D. W. Hearn, and W. W. Hager, eds. *Network Optimization*. *Proc. Network Optim. Conf.* February 12–14, University of Florida, Gainesville, FL. *Lecture Notes in Economics and Mathematical Systems*, vol. 450, Springer-Verlag, Berlin, Germany, 51–71.

Bertsekas, D. P. 1999. *Nonlinear Programming*, 2nd ed. Athena Scientific, Bellmont, MA.

Bihain, A. 1984. Optimization of upper semidifferentiable functions. *J. Optim. Theory and Appl.* **44** 545–568.

Bonnans, J. F., A. Shapiro. 2000. *Perturbation Analysis of Optimization Problems*. Springer-Verlag, New York.

Cantarella, G. E., A. Sforza. 1986. Methods for equilibrium network traffic signal setting. A. R. Odoni, L. Bianco, and G. Szegö, eds. *Flow Control of Congested Networks*. *Proc. NATO Advanced Res. Workshop on Flow Control of Congested Networks*, October 12–18, Capri, Italy. *NATO ASI Series F: Computer and Systems Sciences*, vol. 38, Springer-Verlag, Berlin, Germany, 69–89.

Cho, H.-J., T. E. Smith, T. L. Friesz. 2000. A reduction method for local sensitivity analyses of network equilibrium arc flows. *Transportation Res.* **34B** 31–51.

Dafermos, S., F. T. Sparrow. 1971. Optimal resource allocation and toll patterns in user-optimised transport networks. *J. Transportation Econom. Policy* **5** 184–200.

Davis, G. A. 1994. Exact local solution of the continuous network design problem via stochastic user equilibrium assignment. *Transportation Res.* **28B** 61–75.

Dempe, S. 1997. On the leader's dilemma and a new idea for attacking bilevel programming problems. Technical report, Fachbereich Mathematik, Technische Universität Chemnitz, Chemnitz, Germany.

——. 2000. A bundle algorithm applied to bilevel programming problems with non-unique lower level solutions. *Comput. Optim. Appl.* **15** 145–166.

——, H. Schmidt. 1996. On an algorithm solving two-level programming problems with nonunique lower level solutions. *Comput. Optim. Appl.* **6** 227–249.

——, S. Vogel. 1997. The subdifferential of the optimal solution in parametric optimization. Technical report, Fachbereich Mathematik, Technische Universität Chemnitz, Chemnitz, Germany.

Dontchev, A. L., R. T. Rockafellar. 2000. Ample parameterization of variational inclusions. Preprint, Department of Mathematics, University of Washington, Seattle, WA.

Ferrari, P. 1997. Capacity constraints in urban transport networks. *Transportation Res.* **31B** 291–301.

Ferris, M. C., D. Ralph. 1995. Projected gradient methods for nonlinear complementarity problems via normal maps. D. Du, L. Qi, and R. Womersley, eds. *Recent Advances in Nonsmooth Optimization*. World Scientific Publishers, River Edge, NJ, 57–87.

Fisk, C. 1980. Some developments in equilibrium traffic assignment. *Transportation Res.* **14B** 243–255.

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

Grippo, L., F. Lampariello, S. Lucidi. 1991. A class of nonmonotone stabilization methods in unconstrained optimization. *Numerische Mathematik* **59** 779–805.

Heydecker, B. G. 1983. Some consequences of detailed junction modeling in road traffic assignment. *Transportation Sci.* **17** 263–281.

Kiwiel, K. C. 1996. Restricted step and Levenberg–Marquardt techniques in proximal bundle methods for nonconvex nondifferentiable optimization. *SIAM J. Optim.* **6** 227–249.

Larsson, T., M. Patriksson. 1992. Simplicial decomposition with disaggregated representation for the traffic assignment problem. *Transportation Sci.* **26** 4–17.

——, ——. 1994. Equilibrium characterizations of solutions to side constrained asymmetric traffic assignment models. *Le Matematiche* **49** 249–280.

——, ——. 1995. An augmented Lagrangean dual algorithm for link capacity side constrained traffic assignment problems. *Transportation Res.* **29B** 433–455.

——, ——. 1998. Side constrained traffic equilibrium models—Traffic management through link tolls. P. Marcotte and S. Nguyen, eds. *Equilibrium and Advanced Transportation Modelling*. Kluwer Academic Publishers, New York, 125–151.

——, ——. 1999. Side constrained traffic equilibrium models: Analysis, computation and applications. *Transportation Res.* **33B** 233–264.

LeBlanc, L. J., M. Abdulaal. 1979. Continuous equilibrium network design models. *Transportation Res.* **13B** 19–32.

Loridan, P., J. Morgan. 1992. Least-norm regularization for weak two-level optimization problems. V. Barbu, J. F. Bonnans, and D. Tiba, eds. *Optimization, Optimal Control and Partial Differential Equations*. Papers from the *First Franco-Romanian Conf.* September 7–11, Iaşi. *International Series of Numerical Mathematics*, vol. 107, Birkhäuser, Basel, 307–318.

——, ——. 1996. Weak via strong Stackelberg problems: New results. *J. Global Optim.* **8** 263–287.

Luo, Z.-Q., J.-S. Pang, D. Ralph. 1996. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, U.K.

Mäkelä, M. M., M. Miettinen, L. Lukšan, J. Vlček. 1999. Comparing nonsmooth nonconvex bundle methods in solving hemivariational inequalities. *J. Global Optim.* **14** 117–135.

Marcotte, P. 1986. Network design problem with congestion effects: A case of bilevel programming. *Math. Programming* **34** 142–162.

Migdalas, A. 1995. Bilevel programming in traffic planning: Models, methods and challenge. *J. Global Optim.* **7** 381–405.

Miller, S. D., H. J. Payne, W. A. Thompson. 1975. An algorithm for traffic assignment on capacity constrained transportation networks with queues. April 2–4, *Johns Hopkins Conf. on Inform. Sci. Systems*, The Johns Hopkins University, Baltimore, MD.

Minty, G. J. 1967. A theorem on monotone sets in Hilbert spaces. *J. Math. Analysis and Appl.* **16** 434–439.

Nagurney, A. 1993. *Network Economics: A Variational Inequality Approach*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Outrata, J. V. 1997. On a special class of mathematical programs with equilibrium constraints. P. Gritzmann, R. Horst, E. Sachs, and R. Tichatschke, eds. *Recent Advances in Optimization*. *Proc. 8th French-German Conf. Optim.*, July 21–26, Trier. *Lecture Notes in Economics and Mathematical Systems*, vol. 452, Springer-Verlag, Berlin, Germany, 246–260.

——, J. Zowe. 1995. A numerical approach to optimization problems with variational inequality constraints. *Math. Programming* **68** 105–130.

——, M. Kočvara, J. Zowe. 1998. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Pang, J.-S., L. Qi. 1993. Nonsmooth equations: Motivations and algorithms. *SIAM J. Optim.* **3** 443–465.

——, D. Ralph. 1996. Piecewise smoothness, local invertibility, and parametric analysis of normal maps. *Math. Oper. Res.* **21** 401–426.

——, S.-P. Han, N. Rangaraj. 1991. Minimization of locally Lipschitz functions. *SIAM J. Optim.* **1** 57–82.

Patriksson, M. 1994. *The Traffic Assignment Problem—Models and Methods. Topics in Transportation*, VSP BV, Utrecht, The Netherlands.

——. 1998. *Nonlinear Programming and Variational Inequalities: A Unified Approach*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

——. 2002. Sensitivity analysis of traffic equilibria. Under revision for *Transportation Sci.*

——, R. T. Rockafellar. 2002. Sensitivity analysis of variational inequalities over aggregated polyhedra, with application to traffic equilibria. Forthcoming, *Transportation Sci.*

Qi, L. 1993. Convergence analysis of some algorithms for solving nonsmooth equations. *Math. Oper. Res.* **18** 227–244.

——, J. Sun. 1993. A nonsmooth version of Newton's method. *Math. Programming* **58** 353–368.

Qiu, Y., T. L. Magnanti. 1989. Sensitivity analysis for variational inequalities defined on polyhedral sets. *Math. Oper. Res.* **14** 410–432.

Robinson, S. M. 1980. Strongly regular generalized equations. *Math. Oper. Res.* **5** 43–62.

——. 1985. Implicit B-differentiability in generalized equations. Technical summary report no. 2854, Mathematics Research Center, University of Wisconsin at Madison, Madison, WI.

——. 1992. Normal maps induced by linear transformations. *Math. Oper. Res.* **17** 691–714.

Rockafellar, R. T., R. J-B. Wets. 1998. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften, vol. 317, Springer-Verlag, Berlin, Germany.

Scheel, H., S. Scholtes. 2000. Mathematical programs with complementarity constraints: Stationarity, optimality and sensitivity. *Math. Oper. Res.* **25** 1–22.

Schramm, H., J. Zowe. 1992. A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. *SIAM J. Optim.* **2** 121–152.

Smith, M. J., T. Van Vuren. 1993. Traffic equilibrium with responsive traffic control. *Transportation Sci.* **27** 118–132.

Tobin, R. L., T. L. Friesz. 1988. Sensitivity analysis for equilibrium network flow. *Transportation Sci.* **22** 242–250.

Toint, Ph., L. Wynter. 1996. Asymmetric multiclass traffic assignment: A coherent formulation. J.-B. Lesort, ed. *Transportation and Traffic Theory. Proc. 13th Internat. Symposium on Transportation and Traffic Theory*, July 24–26, Lyon, France. Pergamon Press, Oxford, U.K., 237–260.

Wardrop, J. G. 1952. Some theoretical aspects of road traffic research. *Proc. Inst. Civil Engineers, Part II*, 325–378.

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

Yen, N. D. 1995. Lipschitz continuity of solutions of variational inequalities with a parametric polyhedral constraint. *Math. Oper. Res.* **20** 695–708.

Zhang, R. 1994. Problems of hierarchical optimization in finite dimension. *SIAM J. Optim.* **4** 521–536.

Received: June 2000; revision received: March 2001; accepted: March 2001.