6

Price-directive Traffic Management: Applications of Side Constrained Traffic Equilibrium Models

Torbjörn Larsson

Department of Mathematics, Linköping University

S-581 83 Linköping, Sweden

Michael Patriksson

Department of Mathematics, Chalmers University of Technology

S-412 96 Göteborg, Sweden

Abstract

We propose a systematic means for achieving a set of overall traffic management or planning goals with respect to the performance of the traffic network through the use of link tolls. The primary goals are defined by a set of link flow restrictions. The tolls that achieve these goals are obtained by solving a generalization of the classical user equilibrium model which includes a set of side constraints on the link flows. The set of toll prices obtained is not necessarily unique; this fact enables the traffic planner to choose a toll scheme which satisfies exogenous constraints and which may optimize a secondary goal, such as with respect to the toll itself. The overall model is derived as a special case of a mathematical program with equilibrium constraints (MPEC) describing a Stackelberg game involving the traffic manager and the users of the network. The model is shown to yield valuable

information also in the case where the management goals and exogenous toll constraints are inconsistent with each other or with the underlying network. We give several examples of possible applications of the model, including the achievement of a system optimal flow and the derivation of actions for making public transport more attractive, and propose a conceptual algorithm for solving it. The paper is hoped to provoke continued research in both theoretical and applied directions.

Keywords: Stackelberg game; link tolls; mathematical programs with equilibrium constraints; side constrained traffic equilibrium; Lagrange multipliers; toll optimization; network design; system optimal flows; subsidies

1 INTRODUCTION

The criteria by which travellers choose their routes in the traffic network are, to some degree, in conflict with society's goal of utilizing the network efficiently: a traveller can be expected to most often choose a route which minimizes a combination of travel time and expenses while society's goal often is to have low average travel times and little damage to the environment. The traffic system may be viewed as a non-cooperative Stackelberg game, in which a traffic manager, represented as the leader, changes the infrastructure so as to achieve some overall management goal with respect to the distribution of the traffic in the network. The travellers are then modelled as the followers; they react to the infrastructure changes by modifying their behaviour, for example by adjusting their route choices or travel modes. If the manager's infrastructure changes 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 traffic controls such as traffic lights and one-way traffic, to introduce tolls on some links, or some other network design measure, and to supply the travellers with information about alternative routes.

This work considers the use of a decentralized traffic control, namely the instrument of *link tolls*, to achieve a set of overall management or planning goals, and proposes a systematic means for deriving adequate toll prices. The primary goals considered can be quite general, and are defined by a set of desired restrictions on the link flows. Travellers are assumed to choose their routes in accordance with Wardrop's principle of user equilibrium; this means that the link tolls are to be set at a level such that the desired flow distribution is the result of an equilibration process where the travel costs equal the link tolls plus all other costs associated with the trips.

The means for deriving the proper link tolls are mathematical programming models which generalize the standard user equilibrium model by the introduction of *side constraints* describing the goals, and, in particular, the *Lagrange multipliers* associated with these constraints. The use of a mathematical model for the derivation of the link tolls is in contrast to heuristic toll schemes, based on trial and error, whose resulting traffic flows can never be anticipated. Further, the use of side constraints supports the process of identifying the proper goals to be formulated, and the definitions of the side constraints are quite flexible with respect to the goals that can be considered.

We analyze this model, in particular with regards to the properties of the set of tolls that will achieve the management goals, and the consequences of inconsistencies within the model. The latter is perhaps the most interesting part of the analysis, considering that in practice it is very likely that the management goals are unattainable, or that the tolls achieving them and those satisfying all exogenous toll constraints are incompatible. The work also supplies some example applications, and interesting topics for future research.

We limit ourselves to the study of a *static* model (an assumption which is essential), with *fixed demands* and a *single class of users* having *additive link travel costs* (assumptions none of which are essential).

To introduce the setting of the problem studied and most of the notation used, we discuss the Wardrop equilibrium conditions in the next section. In Section 3, we introduce the general Stackelberg problem. Section 4 explains the basic properties of the special case of model studied in the context of toll networks. A conceptual algorithm, as well as the implications of inconsistencies in the model, is discussed in Section 5. Finally, Section 6 discusses some interesting applications.

2 WARDROP EQUILIBRIUM

Let $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ be a strongly connected 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 for simplicity shall be assumed fixed) giving rise to a link traffic flow pattern.

Wardrop's user equilibrium principle states that for every origin–destination (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 in OD pair (p, q), by h_{pqr} the flow on route $r \in \mathcal{R}_{pq}$, and by $c_{pqr} = c_{pqr}(h)$ the travel cost on the route given the vector h of route flows; with this notation, an equilibrium flow is defined by the conditions

$$h_{pqr} > 0 \implies c_{pqr} = \pi_{pq}, \qquad \forall r \in \mathcal{R}_{pq}, \tag{2.1a}$$

$$h_{pqr} = 0 \implies c_{pqr} \geq \pi_{pq}, \qquad \forall r \in \mathcal{R}_{pq}, \tag{2.1b}$$

where the value of π_{pq} is the minimal (or, equilibrium) route cost in OD pair (p, q). An equilibrium state is reached precisely when no traveller can decrease his/her travel cost by shifting to another route.

The Wardrop conditions can be stated as a *variational inequality problem*: the conditions (2.1) are equivalent to h satisfying

[TAP-VIP-H]

$$-c(h) \in N_H(h),$$

where $H = \{\, h \in \Re_+^{|\mathcal{R}|} \mid \Gamma^{\mathrm{T}} h = d \,\}$ is the set of demand-feasible route flows (Γ is the route–OD pair incidence matrix), $c : \Re_+^{|\mathcal{R}|} \mapsto \Re_{++}^{|\mathcal{R}|}$ is the vector of route travel cost functions, $|\mathcal{R}|$ is the total number of simple routes in the network, and $N_H(h)$ denotes the *normal cone* to H at h, that is, the set

$$N_H(h) = \begin{cases} \{\, z \in \Re^{|\mathcal{R}|} \mid z^{\mathrm{T}}(y - h) \leq 0, \quad \forall y \in H \,\}, & h \in H, \\ \emptyset, & h \notin H. \end{cases}$$

In the case where the travel cost of a route is the sum of the travel costs on the links defining it, or, in other words, the route costs are *additive*, then the Wardrop conditions can be described in terms of link flows. (This network description may be desirable, as the number of routes in a network can be extremely large.) Letting F denote the (bounded polyhedral) set of demand-feasible link flows, that is, the link flows f satisfying $f = \Delta h$ for some vector h in H where Δ is the link–route incidence matrix, the problem [TAP-VIP-H] can be equivalently written as

[TAP-VIP-F]

$$-t(f) \in N_F(f),$$

where $t : \Re_+^{|\mathcal{A}|} \mapsto \Re_{++}^{|\mathcal{A}|}$ is the vector of link travel cost functions. [The link and route costs are related by $c(h) = \Delta^{\mathrm{T}} t(f)$.] The flow polytope F can, equivalently by the positivity assumption on the link costs, be described by node conservation constraints.

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

3 A GENERAL STACKELBERG MODEL

We start by formulating a general Stackelberg problem; later, we shall introduce reasonable assumptions on the nature of the actions taken by the traffic manager that will allow us to reformulate it into a much simpler problem.

Consider a situation in which the manager of a traffic system wishes to achieve certain goals with respect to the performance of the system. The goals are formulated as a set of flow restrictions of the form

$$g_k(f) \leq 0, \qquad k \in \mathcal{K}, \tag{3.1}$$

where $\mathcal{K}$ is a finite index set, and $g_k : \Re^{|\mathcal{A}|} \mapsto \Re$ are C^1 functions. The link flows that satisfy (3.1) form a closed set in $\Re^{|\mathcal{A}|}$, which we denote by G.

We let the vector $p \in \Re^{|\mathcal{A}|}$ represent the actions taken by the manager; the actions are assumed to result only in an influence upon the travel costs, according to some parameterization, $\bar{t}(f,p)$, of the link travel cost. There is a large freedom in interpretation of these actions, and several possible examples were given in the introduction.

These actions are further assumed to be restricted to some set P, which may be determined by political, practical, environmental and economical constraints, and possibly other considerations as well. Among the possible actions and flows in G, the manager optimizes a function, $\varphi : P \times F \mapsto \Re$ of the actions taken and resulting link 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 are deliberately imprecise in the description of this general model, since it will soon be reduced to a simpler model which we shall deal with in more detail.)

Example 3.1 The most immediate example of a set of desired flow restrictions is that of upper bounds on some links' flows. In the framework of (3.1), such constraints are described by

$$g_a(f_a) := f_a - u_a \leq 0, \qquad a \in \underline{\mathcal{A}}, \tag{3.2}$$

where $\underline{\mathcal{A}} \subseteq \mathcal{A}$ and $u_a \geq 0$ is the upper bound on the flow on link a. □

Remark 3.2 The restrictions are given in terms of *link flows*. Since link flows are the likely ones to be observed and those that may be affected through control measures, this is a natural assumption. We may, however, easily extend the situation described to cover also the cases where the side constraints are given in *route flow* variables, and the restrictions distinguish between *different classes of travellers and/or vehicles*.

The latter extension could serve as a modelling tool for taking into account different *values-of-time* among the travellers; further, it can be used to derive proper tolls on links leading into a given residential area for tourists but not for residents there, and tolls and/or additional queueing delays for private vehicles in order to favour public transport. We discuss these applications in more detail in Section 6. □

This development results in a *mathematical program with equilibrium constraints*:

[MPEC]

$$\text{minimize } \varphi(p, f) \tag{3.3a}$$

subject to

$$p \in P, \tag{3.3b}$$
$$f \in G, \tag{3.3c}$$

where

$$-\bar{t}(f, p) \in N_F(f). \tag{3.3d}$$

Remark 3.3 Should $\bar{t}(\cdot, p)$ be a strictly monotone mapping for every feasible choice of p, the lower-level (follower) problem (3.3d) has a unique solution, and as such f can be described as $f(p)$. Then, the function φ is in fact a function of p only. □

Mathematical programs with equilibrium constraints generalize bilevel optimization problems, and are in general non-convex. [For an overview of bilevel problems in transportation research, see Migdalas (1995).]

Example 3.4 We consider a version of the *equilibrium network design problem*. Let p denote an investment in network capacity; the effect of an investment is that of a reduced travel time; $\bar{t}_a$ is often taken to be $\bar{t}_a(f, p) = t_a(f_a/p_a)$. An investment p_a is associated with an investment cost, $\psi_a(p_a)$. The goal is to minimize the total travel time (at a user equilibrium flow) plus the investment costs, that is, $\varphi(p, f) = \sum_{a \in \mathcal{A}} \{\bar{t}_a(f, p) f_a + \psi_a(p_a)\}$, while satisfying constraints on the investments made, $P = \{\, p \in \Re^{|\mathcal{A}|} \mid \ell \leq p \leq u;\ \sum_{a \in \mathcal{A}} p_a \leq U \,\}$. In [MPEC], network performance is measured also by constraints in the model. □

In the following, we shall introduce assumptions on what the actions p represent which reduce the problem [MPEC] to a sequence of two (potentially) well-behaved problems.

4 A TOLL-BASED MODEL

We restrict our attention to considering an action of particular interest: on each link (at least potentially), a fixed toll is levied; thus, the influence on the travellers can be described by $\bar{t}(f, p) = t(f) + p$, where p is the toll fee.

Introducing some assumptions on this toll scheme, we first argue that it is natural to assume that link tolls are introduced only on those links where they are needed for achieving the goals; in other words, tolls should not be levied on links that do not contribute to the saturation of any restriction in (3.1). We also assume that the toll levied on a link can be decomposed into the contributions of all the different restrictions in (3.1), in a proportional manner according to their contribution to the saturation of the different restrictions; their contribution is measured in terms of derivatives of the constraint functions. (This decomposition principle makes it possible to investigate the economics of each flow restriction, and facilitate the economical evaluation of the toll scheme; the proportionality principle seems natural from the viewpoint of fairness.)

The implication of these conditions is that the link toll is distributed according to

$$p_a = \sum_{k \in \mathcal{K}(f)} \beta_k \frac{\partial g_k(f)}{\partial f_a}, \qquad a \in \mathcal{A}, \tag{4.1}$$

where $\mathcal{K}(f)$ is the subset of $\mathcal{K}$ for which $g_k(f) = 0$, and β_k, $k \in \mathcal{K}$, are nonnegative constants. (In vector notation, then, $p = \nabla g(f)\beta$, with $\beta^{\mathrm{T}} g(f) = 0$ and $\beta \geq 0$.)

Example 4.1 In the case of simple upper bounds (see Example 3.1), the expression (4.1) reduces to $p_a = \beta_a$, $a \in \underline{\mathcal{A}}$. In applications to queueing networks (see Larsson and Patriksson, 1995a and 1995b), p_a is interpreted as the equilibrium queueing delay on link a. The above complementarity condition then states that there is a queueing delay ($p_a > 0$) only if the link is saturated ($f_a = u_a$). □

The effect of assumption (4.1) is more far-reaching than what might appear at first sight. The main effect is that of reducing the two-level problem into a sequence of two single-level ones; the reason is that what appears to be a mere definition is actually a complementarity condition, linking the two variable groups; in fact, (4.1) has the effect of reducing p to be a vector of Lagrange multipliers for the side constraints (3.1). Further, these single-level problems can, under some additional, but standard, assumptions, be formulated as convex programs.

To avoid having to deal with a lot of mathematical details without reaching a much deeper insight or generality, we will henceforth assume that G *is a convex set which further satisfies a constraint qualification* (e.g., Bazaraa *et al.*, 1993, Chapter 5) and, until further notice, that $F \cap G$ *is non-empty.* (The case of inconsistency is covered in Section 5.) (In the case where all functions g_k are affine, the former requirements are fulfilled automatically.) It follows immediately from the assumptions that $F \cap G$ is non-empty, compact and convex.

With these assumptions in place, we observe that the requirements (3.3c)–(3.3d) simplify to the primal–dual variational inequality problem of finding $(f, p) \in F \times \Re^{|\mathcal{A}|}$ satisfying[1]

$$-[t(f) + p] \in N_F(f), \tag{4.2a}$$

$$p \in N_G(f). \tag{4.2b}$$

Remark 4.2 **(a)** The assumptions made on the toll scheme have the effect of making it possible to place the constraints (3.3c) at the lower level of [MPEC]; this is instrumental in making the problem tractable.

(b) The formulation (4.2) is valid also when the management goals define non-convex constraints, through the proper re-definition of the normal cone mapping N_G.

(c) It is clear from the expression (4.2b) that the toll vector p is completely independent of the way in which the set G is described in terms of constraints. If, for example, the restrictions in (3.1) are given in different units or are scaled differently, a rescaling into the same units will not be necessary. Also, clearly, any presence of redundancy in the description of the set G will have no effect whatsoever on the resulting toll vector.[2] □

[1] The relation (4.2a) is equivalent to (3.3d), while (4.2b) is a restriction of (3.3c); the latter is easily verified by observing that, by convex analysis, the relation (4.2b) is equivalent to f being in the face of G exposed by the vector p, which is a subset of G.

[2] This fact is highlighted when describing the variational inequality in terms of normal cones; it was one of the main motivations for choosing this form of presentation. Another is that the traditional way of describing a variational inequality is not useful in the case of non-convex constraints.

In order for the vector p to be naturally associated with a toll, that is, p is non-negative, $N_G(f) \subseteq \Re^{|\mathcal{A}|}_+$ must hold at every solution f to [TAP-VIP-SC-F]. A sufficient condition for this to hold is that $\frac{\partial g_k(f)}{\partial f_a} \geq 0$ for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$ [cf. (4.1)]. If the condition is not fulfilled, then some link tolls can be negative, whence the "toll" should be regarded as a subsidy or a network improvement. (A natural condition is, however, that $t(f) + N_G(f) \subseteq \Re^{|\mathcal{A}|}_+$ holds.) (Applications of negative tolls are described in Section 6.4.)

Example 4.3 The case of upper bounds on the link flows (cf. Examples 3.1 and 4.1) is one in which, clearly, tolls will be non-negative. □

We now observe that, by convex analysis, $N_{F\cap G}(f) = N_F(f) + N_G(f)$ holds everywhere, and so solving the system (3.3c)–(3.3d) reduces to finding a solution f to

[TAP-VIP-SC-F]

$$-t(f) \in N_{F\cap G}(f).$$

The above assumptions ensure the existence of a solution to [TAP-VIP-SC-F] (e.g., Hartmann and Stampacchia, 1966). This problem was studied in detail in Patriksson (1994) and Larsson and Patriksson (1995b and 1996) (in the optimization setting), and in Larsson and Patriksson (1994), in the context of descriptive models of traffic equilibria including *queueing effects*; the above development shows that the model is also intimately associated with *link tolls*.

Remark 4.4 In applications of side constrained traffic equilibria to queueing networks, we have in Larsson and Patriksson (1994 and 1996) derived natural conditions on the queueing dynamics that ensure the existence of an equilibrium in both link flows and queues. These conditions, and the results obtained, can be repeated verbatim (with the word 'queue' replaced by the word 'toll' throughout) for toll networks, thus obtaining an existence result for an equilibrium in both link flows and tolls, given natural conditions on the properties of the tolls during a non-stationary (disequilibrium) state. (Essentially, as time passes the toll contributed by a restriction in (3.1) increases to infinity if the flow violates the restriction, decreases to zero if the restriction is strictly fulfilled, and stays constant if the restriction is saturated exactly.)

The descriptive and prescriptive modelling approaches can of course be combined into a single bilevel model, in which, then, the lower level describes a queue equilibrium model with toll parameters (see, for example, Yang and Bell, 1997). □

One cannot assume in general that the set of multipliers β_k for the flow restrictions (3.1) is a singleton, and therefore, while the link flow solution is unique whenever the travel cost mapping t is strictly monotone, the toll vector p which yields the desired equilibrium solution is not. In the context of queueing networks, this can be considered a drawback, since the vector p is, in that case, interpreted as a queueing delay. So, in Larsson and Patriksson (1994 and 1996), sufficient conditions were therefore derived under which this equilibrium link queueing delay is unique. In the context of this work, however, the non-uniqueness of the multipliers β can be considered a positive feature of the model, since it facilitates satisfying multi-objective strategies (e.g., those having multiple practical, political, environmental and economical constraints on the toll scheme). Similarly, one could choose a toll vector which simultaneously optimizes the function φ. In the next section, we shall elaborate on this further, and study means for solving this optimization model.

5 TOLL OPTIMIZATION

In this section, we study the upper-level problem of [MPEC] in terms of the toll vector p. We then introduce the further assumption that t *is strictly monotone on* F, and denote the unique link flow solution to [TAP-VIP-SC-F] by f^*. The set of toll vectors which achieves this solution while satisfying the requirement (4.1) is given by the set of solutions to (4.2) with $f = f^*$. This set, which we will denote by T, is clearly a non-empty polyhedral set, since F is described by linear constraints and G satisfies a constraint qualification, so the respective normal cones at f^* are polyhedral.

5.1 The toll optimization problem

The toll optimization problem is

[TOP]

$$\underset{p \in P \cap T}{\text{minimize}} \ \varphi(p).$$

This problem is convex, whenever *the set* P *of feasible toll schemes is convex, and the function* $\varphi : \Re^{|\mathcal{A}|} \mapsto \Re$ *is convex.* (It is further a linear program if P is polyhedral and φ is a linear function.) To be well-defined, a further requirement on the problem is that $P \cap T$ *is non-empty.* The convexity and non-emptiness conditions are assumed to hold for the present. We shall further assume throughout that P *is bounded*, and that *every vector* $p \in P$ *satisfies* $\sum_{a \in \mathcal{A}_C} [t_a(f^*) + p_a] \geq 0$ *for every compatible cycle* $\mathcal{A}_C$ *in the network*, that is, there is no feasible toll vector such that a traveller can benefit by travelling a full cycle. These natural conditions ensure the existence of an optimal solution to [TOP] (under the above conditions), and that the algorithm proposed for solving it is well-defined.

Example 5.1 Several optimization models of the form [TOP] are conceivable. We may wish to levy the least total toll in the network, whence we choose $\varphi(p) = \sum_{a \in \mathcal{A}} p_a f_a^*$, or precisely the opposite; the function φ may also include a measure of the cost of implementing toll stations on the different links, which we naturally would like to minimize. □

5.2 An algorithm for the toll optimization problem

Solving the problem [TAP-VIP-SC-F]. We consider solving [TAP-VIP-SC-F] using a non-linear pricing scheme which simultaneously generates f^* and a vector β of multipliers for the side constraints. One such example is the *(proximal) method of multipliers* of Rockafellar (1978). Reasons for choosing such a scheme are, at least, threefold: (1) in the optimization setting, the corresponding *augmented Lagrangean* dualization schemes have been found to work very well in numerical experiments (Hearn and Ribera, 1980; Larsson and Patriksson, 1995a); (2) a vector β (and hence, through relation (4.1), a toll vector $p \in T$) is obtained through a convergent sequence of dual vectors; and (3) even in the case of inconsistency (that is, $P \cap T = \emptyset$), a toll vector with interesting properties is produced. (More on the last point below.) We refer the reader to Hearn and Ribera (1980), Patriksson (1994) and Larsson and Patriksson (1994 and 1995a) for further details on the properties of this class of schemes in the context of traffic problems.

Within this scheme a sequence of subproblems of the form [TAP-VIP-F] are constructed and solved; in these problems, the mapping t is augmented by primal–dual terms associated with the side constraints. When selecting an efficient algorithm for their solution, we note

these important points: (1) the algorithm should have efficient primal reoptimization capabilities, to utilize that successive subproblems differ only in the travel cost mapping; this suggests using an efficient primal algorithm; and (2) when solving the toll problem [TOP], we must have an explicit description of the set T of optimal tolls, or a good approximation thereof; this suggests the use of column generation/cutting plane approaches.

The algorithm chosen is the *disaggregate simplicial decomposition* (DSD) method, which has been successfully applied in previous works on optimization versions of [TAP-VIP-F] (Larsson and Patriksson, 1992) and [TAP-VIP-SC-F] (Larsson and Patriksson, 1995a).[3]

We refer the reader to those references and Patriksson (1994, Sections 4.2.3, 4.3.5 and 5.3.5) for details on this method, but mention here that, at termination, the algorithm stores those routes in each OD pair that have a positive flow at some optimal route flow solution. In the application to [TAP-VIP-SC-F], then, the DSD algorithm will provide, for each OD pair (p,q), the routes in $\mathcal{R}_{pq}$ which are utilized in *one* route flow solution h to [TAP-VIP-SC-F]; these routes will also [cf. (4.2a)] all be least-cost routes given the fixed link travel cost vector $t(f^*)+p$, where $p=\nabla g(f^*)\beta$ and β is the vector of multipliers for (3.1) which is obtained in the limit of the multiplier scheme.

A regularity condition. Using linear programming duality on (4.2a), the set T can be described as the set of solutions in p of the following linear system in p, β and π:

$$\begin{aligned}
\Delta^{\mathrm{T}}[t(f^*)+p]-\Gamma\pi &\geq 0, && \text{(5.1a)}\\
[t(f^*)+p]^{\mathrm{T}}f^*-d^{\mathrm{T}}\pi &= 0, && \text{(5.1b)}\\
p-\nabla g(f^*)\beta &= 0, && \text{(5.1c)}\\
\beta^{\mathrm{T}}g(f^*) &= 0, && \text{(5.1d)}\\
\beta &\geq 0; && \text{(5.1e)}
\end{aligned}$$

(5.1a)–(5.1b) is a dual description of the normal cone inclusion of (4.2a), and (5.1c)–(5.1e) is an explicit description of the corresponding inclusion of (4.2b). The constraints (5.1a) render this system intractable in practice, since it contains as many constraints as there are routes in the network (cf. the dimensions of Δ^{T} and Γ); the constraints actually needed, however, are defined by the subset corresponding to the union of the routes being at minimum cost for *some* multiplier vector β, and should be comparatively very small.

Larsson and Patriksson (1994) introduce a regularity condition on the solution to [TAP-VIP-SC-F], so that the set of least-cost routes is invariant over the set of multipliers β for the side constraints, and hence over the set $p\in T$ (cf. ibid, Thm. 2.6). The condition says that, for each such vector, the Wardrop conditions are satisfied with *strict complementarity*, that is, that all routes with the minimal cost are actually used. In this case, using the DSD algorithm for the traffic equilibrium subproblems in the multiplier scheme will yield all the information necessary to describe the (interesting subset of the) toll set T explicitly.

If this is not the case, then the least-cost routes are not the same for all values of the multiplier vector. Although, as remarked earlier, the description of the toll set T requires the union of the least-cost routes over *all* values of the multiplier vector β in the solution to [TAP-VIP-SC-F], the set of routes provided by the DSD algorithm will most probably

[3] One complication with the DSD algorithm, as with any pricing scheme which includes shortest-route calculations, is that for some non-optimal multiplier values, negative cycles may appear. In such cases, one may either derive and incorporate a dual constraint in the scheme, or, preferably, continue the algorithm, ignoring the route generation stage.

be a very good approximation of the total number of least-cost routes needed (it is, for example, quite conceivable that regularity holds at least for some OD pairs), and that is the main motivation for using the DSD algorithm in this context. Our limited numerical experience, which is described in Larsson and Patriksson (1998), also supports this belief.

The case of primal inconsistency. If the traffic manager has set unattainable goals, then the result will be $F \cap G = \emptyset$. Under this circumstance, the algorithm will still provide the manager with useful information. In fact, the algorithm will provide *the smallest possible adjustment of the manager's aspiration levels so that the goals become achievable*, and also provide *the toll vector* $\hat{p} = \nabla g(\hat{f})\hat{\beta}$ *that will achieve these slightly less ambitious aims.*

More formally, in the optimization setting, we may apply Theorem 4.1 of Golshtein and Tretyakov (1996, Sec. 3.4) to conclude that (a very small adjustment of) the algorithm stated above will yield a primal–dual pair $(\hat{f}, \hat{\beta})$ which solves a *perturbation* of [TAP-VIP-SC-F], in which the original restrictions (3.1) have been replaced by

$$g_k(f) \leq \varepsilon_k, \qquad k \in \mathcal{K},$$

the vector of ε_k being that of *smallest Euclidean norm* which makes the problem consistent.[4]

Example 5.2 A traffic manager can utilize this scheme to evaluate several scenarios, possibly involving goals which are deliberately chosen to be unattainable. In contrast to the consistent case [cf. Remark 4.2.c], the resulting toll vector is sensitive to the introduction of scaling factors on the goals. Should the manager be uncertain if the goals are attainable, he/she can fulfill the most important goals while sacrificing others, through the use of different scaling factors on the goals according to their respective importance (cf. the use of aspiration levels and reference goals in multi-criteria optimization). □

Having observed that the method detects primal inconsistency, and also in this case yields information that will be of use to the traffic manager, we move on to consider the problem [TOP] under the assumption that *the problem* [TAP-VIP-SC-F] *is consistent.*

An algorithm for [TOP]. The above discussion can be concluded with the remark that, at termination of the multiplier algorithm, the link flow solution f^* to [TAP-VIP-SC-F] and a toll vector $p \in T$ are given. The subset of routes generated within the DSD algorithm corresponds to a subset of the constraints in (5.1a). With this as the starting solution and information about the toll set T, a conceptual algorithm for [TOP] proceeds as follows.

We begin by noting that although the toll vector p is in T, it is not necessarily in the set P of feasible tolls. The fact that we do not know the set T explicitly further complicates the search for a toll vector which is simultaneously feasible (i.e., belongs to P) and optimal (i.e., belongs to T), should such a vector exist. The starting point in the solution of [TOP] is a *Phase-I* method, which attempts to find such a vector.

With the vector p as a starting point, we search for a point in the intersection of P and the set of tolls described by the subset of constraints in (5.1) currently known. (We shall use the notation $\overline{T}$ for this set.) This is an example of the problem of finding a point in the intersection of a finite number of convex sets, and a variety of methods exist for its solution (e.g., Bauschke and Borwein, 1996). Preferably, we should always stay in the set P, since it is to be considered to be described by *hard* constraints, while the set T is described by

[4] The perturbation ε found is actually the projection of the origin onto the effective domain of the perturbation function.

soft constraints.[5] We note that in the case where P is described by linear constraints, the Phase-I problem is an ordinary linear Phase-I problem, as solved within a simplex method.

Two outcomes are possible from this procedure. In the *first* case, the procedure reports that there is a positive distance between the sets P and $\overline{T}$. Since $T \subseteq \overline{T}$ holds, it must therefore be the case that the problem [TOP] is inconsistent, that is, $P \cap T = \emptyset$. (We refer to this property as *dual inconsistency.*)

The Phase-I method described above will, at termination, then have at hand a vector p which belongs to the set P (and is hence feasible), but which does not belong to the set T. This implies that *the link flow f^* solving* [TAP-VIP-SC-F] *is not possible to obtain through the use of a link toll scheme which is feasible with respect to the set P.* (The toll vector obtained through the solution of [TAP-VIP-SC-F], for example, then clearly is infeasible.) We can, however, expect that by using the toll vector obtained, at least a subset of the goals described by (3.1) will be achieved, and hence, *the toll vector obtained will achieve, by way of the Phase-I procedure, the highest aspiration levels possible within the limits set by the toll constraints.*

The second outcome possible is that the distance is reported to be zero, and then the output is a toll vector belonging to $P \cap \overline{T}$. The reader is, however, warned against interpreting this as a confirmation that the problem [TOP] is consistent.

This toll vector, $\overline{p}$, together with the vectors $\overline{\beta}$ and $\overline{\pi}$, obtained from the solution to the above Phase-I problem, is next used to investigate if $\overline{p} \in T$ actually holds, by checking whether any constraints in (5.1a) are violated. With the link costs $t(f^*) + \overline{p}$ at hand, the most violated, if any, constraint in (5.1a) for an OD pair (p,q) is obtained by finding the shortest route, $r \in \mathcal{R}_{pq}$ say; the constraints (5.1a) for the OD pair are satisfied if and only if $\sum_{a \in \mathcal{A}} \delta_{pqra}[t_a(f^*) + \overline{p}_a] \geq \overline{\pi}_{pq}$; otherwise, the most violated constraint has been identified, and the corresponding route is added to the subset of $\mathcal{R}_{pq}$ currently known. Performing this test for each OD pair will either result in termination of the Phase-I procedure, with a toll vector $\overline{p} \in T$ at hand, or to the generation of constraints (routes) in some OD pairs. In the latter case, a better outer approximation $\overline{T}$ of T has been found, and the Phase-I problem is resolved. This repeated procedure will terminate finitely, since the number of routes is finite. At termination, the procedure either reports that $P \cap T = \emptyset$, or provides a vector $\overline{p}$ in this set.

Having already discussed the case of dual inconsistency, we henceforth assume that *the Phase-I procedure was terminated with a toll vector in $P \cap T$*,[6] and hence *the problem* [TOP] *is consistent.* We next turn to the Phase-II part of the algorithm for [TOP].

In the second phase, we solve the modification of [TOP], where T is replaced by $\overline{T}$.[7] Let $(\overline{p}, \overline{\beta}, \overline{\pi})$ be a solution to this problem.[8] We next investigate if this solution solves [TOP], by checking whether any constraints in (5.1a) are violated. This is done in precisely the same manner as described above in the Phase-I procedure. Performing this test for each OD pair will either result in termination, with $(\overline{p}, \overline{\beta}, \overline{\pi})$ being a *proven optimal solution to* [TOP], or to the generation of constraints (routes) in some OD pairs. In the latter case, a better outer approximation of T has been found, and $(\overline{p}, \overline{\beta}, \overline{\pi})$ is used as an infeasible starting solution when solving the improved approximation of [TOP]. Obviously, since the

[5] We can envisage the set P to be described by much simpler constraints than those describing T, and hence such a procedure should be realistic.

[6] This solution also provides an upper bound on the optimal value of [TOP].

[7] We do not discuss particular methods for the solution of these restrictions in this paper; the adequate method will be a function of the properties of the side constraints and the toll restrictions.

[8] Since φ is optimized over the intersection of P and an outer approximation of T, $\varphi(f^*, \overline{p})$ is a lower bound on the optimal value of [TOP].

number of routes is finite, this procedure is finite, and its output is a feasible and optimal toll vector, p^*, to [TOP].[9]

The algorithm stated here is by no means the only one possible; an alternative constraint generation approach would be the result of using an aggregate simplicial decomposition (SD) algorithm for the [TAP-VIP-F] type subproblems in place of the DSD algorithm. The route generation process is then replaced by the generation of all-or-nothing solutions. Convergence of the disaggregate algorithm given should, however, be much faster, since the latter algorithm corresponds to aggregating constraints generated in the former.

6 APPLICATIONS

6.1 Capacity constraints

In the case of capacity constraints (see Example 3.1), the toll p_a equals the Lagrange multiplier β_a associated with the corresponding constraint in (3.2); see Example 4.1. Algorithms for the generation of f^* solving this special case of [TAP-VIP-SC-F] appear in Hearn and Ribera (1980) and Larsson and Patriksson (1995a). Both of these are multiplier methods which simultaneously provide a multiplier vector β; the latter utilizes the DSD algorithm for the subproblems, and is therefore an instance of the algorithm for [TAP-VIP-SC-F] described above.

Ferrari (1995) considers an elastic demand model with *environmental* capacity constraints, and establishes the existence of an equilibrium which satisfies the capacity constraints in the presence of proper additional costs on links that would otherwise produce unacceptable environmental damage.

6.2 Obtaining a given flow solution

An interesting special case of a goal described by a set of side constraints is that of achieving an *a priori* given flow solution f^*. We may consider several instances of such solutions; for example, f^* may be a calculated system optimal flow or be based on an observed flow. The model [MPEC] applies to such models, with the set of side constraints described by

$$g_a(f_a) := f_a - f_a^* = 0, \qquad a \in \mathcal{A}. \tag{6.1}$$

Tracing the above analysis, we find that the normal cone mapping N_G is non-empty exactly at f^*, where it equals $\Re^{|\mathcal{A}|}$. Hence, the requirement (4.2b) poses no further restrictions on p. This implies, in turn, that the solution of [TOP] is simplified substantially.

Example 6.1 Let f^* be the solution to [TAP-VIP-SC-F]. Using model (6.1) will result in a larger freedom in choosing the tolls compared to that given by the toll set T; it will then be possible to levy a toll on a link which does not contribute to the saturation of any flow restrictions. □

If the given flow f^* does not belong to F, then it is not possible to achieve with any toll vector. In this case, the use of the multiplier scheme outlined in Section 5.2 will provide the toll which *achieves the demand-feasible flow which is closest in Euclidean norm to* f^*. So, again we observe that the traffic management scheme given in this paper provides valuable information also in the case where the goals are unattainable.

[9] Convergence of this algorithm can also be monitored by the bounds on the optimal value of [TOP] generated within the algorithm, and which are remarked upon in previous footnotes.

6.3 System optimal flows

Obtaining a system optimal flow by means of link tolls is a classic topic, with contributions tracing back to Pigou (1920). The traditional approach is *marginal cost pricing*, in which the toll levied equals the marginal cost minus user cost. Dafermos and Sparrow (1971) observe that this toll is not the only one possible, and that it may have certain drawbacks. (For example, a toll will be levied on every link.) They proceed to describe the set of toll vectors which will yield the desired link flow solution, and discuss possible secondary goals to direct the choice of toll vector.

Alternative tolls for achieving a system optimal flow have much later been considered also by Bergendorff *et al.* (1995 and 1997).

Bergendorff (1995) describes a particular case of capacitated model, wherein the upper bounds are the system optimal link flows. Based on this model, she develops a toll optimization problem similar to the one in this paper, and presents a method, working in the space of commodity link flows and node prices, for solving linear instances of [TOP].

Bergendorff *et al.* (1997) present characterizations of the set of toll vectors additional to those of Dafermos, under various assumptions on the travel costs.

It is to be noted that the results reached in these references are, in essence, special cases of the results obtained in this paper, obtained by choosing side constraints of the form (6.1) with f^* being a system optimal solution.

By the generality of the traffic management scheme described in this paper, we can consider more general considerations than the achievement of system optimal flows; for example, we can introduce restrictions of the form (6.1) only on a subnetwork, thus achieving an approximate system optimal flow on a subset of the links, while the rest of the network is governed by the user optimality principle; the resulting model is reminiscent of the combined equilibrium models used in the modelling of vehicle guidance systems.

6.4 Network improvement, subsidies and the redistribution of resources

Consider a situation in which the traffic manager would like to achieve a given link flow solution f^*. The absence of any conditions on the toll vector p other than (4.2a) makes it possible that some link tolls may be negative. An interesting *network design model* emanates from this fact. Indeed, while interpreting a positive value p_a as a toll levied on link a, a negative value of p_a may be interpreted as the *travel time improvement* that is necessary in order to make link a more attractive to some travellers; this information may be used, for example, in the calculation of necessary investments in network capacity. Note that the network design problem stated here has the reverse viewpoint to that of the traditional one.

Example 6.2 Side constrained traffic equilibrium models can be used to derive actions that favour public transport. Denoting the OD flow for private (public) transport by d^{pr} (d^{pu}), side constraints describing a favourable situation for public transport could, for example, be constraints of the form

$$\gamma_{pq} d^{\mathrm{pr}}_{pq} - d^{\mathrm{pu}}_{pq} \leq 0, \qquad (p,q) \in \mathcal{C},$$

where $\gamma_{pq} \geq 0$ is the least ratio of public transport flows over that of private transport in OD pair (p,q); other possibilities include constraints describing maximal travel times for public transport vehicles.

The Lagrange multipliers associated with these constraints in the resulting side constrained *two-mode* traffic equilibrium model are then used to derive (1) an additional

queueing delay for private vehicles, or (2) a subsidy to the public transportation sector. If, further, toll constraints of the form $P = \{ p \in \Re^{|\mathcal{C}|} \mid \sum_{(p,q)\in\mathcal{C}} [p^{\rm pr}_{pq} + p^{\rm pu}_{pq}] = 0 \}$ are introduced in the model, then the subsidy takes the form of a reallocation of resources from the private to the public sector; the subsidy is realized, for example, through tolls levied on private vehicles only, through the design (or, improvement) of designated lanes for the public transport vehicles, or through lower ticket prices. □

6.5 Extensions and further research

In Larsson and Patriksson (1998), the present model is extended in various directions; it is, in particular, shown how elastic demands will affect the toll optimization problem, and how commodity-specific traffic management goals can be fulfilled through the use of consistent, commodity-specific, toll schedules; a particularly interesting model in the latter case is the stohastic user equilibrium model. The paper also discusses some additional properties of the sets of tolls for the fixed and elastic demand problems, and present limited numerical experience with the fixed demand model on some small and medium scale problems.

The price-directive traffic management schemes described here may possibly be applied in other areas, where society through prices (taxes or subsidies) may influence markets. Examples of areas of application include protective duties in international trades, and environmentally derived taxes and subsidies.

This paper provides the basis for several interesting paths of research, both in directions of theoretical research and applications, several of which are mentioned.

ACKNOWLEDGEMENTS

The authors were sponsored by the Swedish Transport and Communications Research Board (Grants KFB 95-118-63 and 96-206-33). The authors thank Pia Bergendorff for inspiring work, and Drs. Jan T. Lundgren and Laura Wynter for constructive remarks; the second author thanks Prof. R. Tyrrell Rockafellar for valuable, continuing discussions.

REFERENCES

Bauschke, H. H. and Borwein, J. M. (1996). On projection algorithms for solving convex feasibility problems. *SIAM Review*, **38** (to appear).

Bazaraa, M. S., Sherali, H. D., and Shetty C. M. (1993). *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, New York, NY, second ed.

Bergendorff, P. (1995). The bounded flow approach to congestion pricing. Master's thesis, Department of Industrial & Systems Engineering, University of Florida, Gainesville, FL.

Bergendorff, P., Hearn, D. W., and Ramana, M. V. (1997). Congestion toll pricing of traffic networks. In: *Network Optimization*, Proceedings of the Network Optimization Conference, University of Florida, Gainesville, FL, February 12–14, 1996 (P. Pardalos, D. W. Hearn, and W. W. Hager, eds.), pp. 51–71. Lecture Notes in Economics and Mathematical Systems, vol. 450, Springer-Verlag, Berlin.

Dafermos, S. and Sparrow, F. T. (1971). Optimal resource allocation and toll patterns in user-optimised transport networks. *Journal of Transportation Economy and Policy*, **5**, 184–200.

Ferrari, P. (1995). Road pricing and network equilibrium. *Transportation Research*, **29B**, 357–372.

Golshtein, E. G. and Tretyakov, N. V. (1996). *Modified Lagrangians & Monotone Maps in Optimization*. John Wiley & Sons, New York, NY.

Hartman, P. and Stampacchia, G. (1966). On some non-linear elliptic differential-functional equations. *Acta Mathematica*, **115**, 271–310.

Hearn, D. W. and Ribera, J. (1980). Bounded flow equilibrium problems by penalty methods. In: *Proceedings of the 1980 IEEE International Conference on Circuits and Computers*, pp. 162–166.

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

Larsson, T. and Patriksson, M. (1994). Equilibrium characterizations of solutions to side constrained asymmetric traffic assignment models. *Le Matematiche*, **49**, 249–280.

Larsson, T. and Patriksson, M. (1995a). An augmented Lagrangean dual algorithm for link capacity side constrained traffic assignment problems. *Transportation Research*, **29B**, 433–455.

Larsson, T. and Patriksson, M. (1995b). On side constrained models of traffic equilibria. In: *Variational Inequalities and Network Equilibrium Problems*, Proceedings of the 19th Course of the International School of Mathematics "G. Stampacchia," Erice, Italy, June 19–25, 1994 (F. Giannessi and A. Maugeri, eds.), pp. 169–178. Plenum Press, New York, NY.

Larsson, T. and Patriksson, M. (1996). Side constrained traffic equilibrium models—analysis, computation and applications. Revised for *Transportation Research, B*.

Larsson, T. and Patriksson, M. (1998). Side constrained traffic equilibrium models—traffic management through link tolls. In: *Equilibrium and Advanced Transportation Modelling* (P. Marcotte and S. Nguyen, eds.). Kluwer Academic Publishers, New York, NY (to appear).

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

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

Patriksson, M. (1994). *The Traffic Assignment Problem: Models and Methods*. Topics in Transportation Series, VSP, Utrecht, The Netherlands.

Pigou, A. C. (1920). *The Economics of Welfare*. Macmillan & Co, London.

Rockafellar, R. T. (1978). Monotone operators and augmented Lagrangian methods in nonlinear programming. In: *Nonlinear Programming 3*, Proceedings of the Nonlinear Programming Symposium 3, University of Wisconsin–Madison, Madison, WI, July 11–13, 1977 (O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds.), pp. 1–25. Academic Press, New York, NY.

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