1 SIDE CONSTRAINED TRAFFIC EQUILIBRIUM MODELS—TRAFFIC MANAGEMENT THROUGH LINK TOLLS

Torbjörn Larsson[1] and Michael Patriksson[2]

[1]Department of Mathematics, Linköping University
S-581 83 Linköping, Sweden
tl@mai.liu.se

[2]Department of Mathematics, Chalmers University of Technology
S-412 96 Göteborg, Sweden
mipat@math.chalmers.se

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 restrictions (constraints) on the link flows. The tolls that achieve these goals are obtained by solving and extracting values of Lagrange multipliers from a generalization of the classical user equilibrium model which includes the link flow restrictions as side constraints. The set of toll prices obtained is not necessarily unique in the inelastic demand case; this fact enables the traffic manager to choose a toll scheme which satisfies exogenous constraints on the toll scheme implemented and which may optimize a secondary goal, such as with respect to the toll itself. For the elastic demand case it is however shown that the total toll revenue is unique; an interesting implication of this result is that if the traffic management goal is to obtain a system-optimal flow with a minimal toll collected, then the classical marginal cost pricing concept is optimal.

The overall model is derived as a restriction 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 transportation network. We give several examples of possible applications and extensions of the model, including the achievement of a system optimal flow and of commodity- and/or user class-specific traffic management goals, and the derivation of actions for making public transport more attractive. We propose a conceptual algorithm for solving it, present preliminary experience from its use, and an example that illustrates some fundamental properties of the toll optimization problem. It is hoped that the paper provokes continued research in both theoretical and applied directions.

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

1.1 INTRODUCTION

The criteria by which travellers choose their routes in a 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 through which the desired goals are fulfilled. The primary goals considered can be quite general, and are defined by a set of desired flows, described by a set of link flow restrictions (constraints). Travellers are assumed to choose their routes in accordance with Wardrop's principle of user equilibrium; this means that the tolls are to be levied on certain links 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 (measured in a common unit).

The means for deriving the proper link tolls are mathematical programming models that 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. This use of a mathematical model for the derivation of link tolls is in contrast to toll schemes that are based on trial and error (such as using heuristic adjustments of a vector of toll parameters in a standard user equilibrium model), where the resulting flows never can be anticipated a priori.[1] 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.

The side constrained traffic equilibrium model used to derive the link tolls is identical to the one analyzed in Larsson and Patriksson (1994, 1995b, 1997b), and hence its basic properties are known. In contrast to the use of side constraints within a decision model, as in this paper, however, they introduce side constraints as a means for improving a (descriptive or prescriptive) behavioural traffic equilibrium model, and so the characterizations and interpretations of its solutions are very different.

In this paper we analyze this model in particular with regards to the properties of the set of tolls that achieves the management goals, and the consequences of inconsistencies in the model. The latter topic is particularly interesting, 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 sample applications, and interesting topics for future research. Preliminary results have earlier been reported in Larsson and Patriksson (1997a).

We limit ourselves to the study of a *static* model,[2] and, for the initial part of the paper, to models with *fixed (inelastic) demands* and a *single class of users* having *additive route travel costs.*

To introduce the setting of the problem studied and most of the notation used, we discuss the Wardrop equilibrium conditions and traffic equilibrium models based on them in the next section. In Section 1.3, we introduce the general Stackelberg problem. Section 1.4 explains the basic properties of the restriction of the general 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 1.5. The section also contains a numerical example which illustrates some of the basic properties of the model, and contrasts in particular its properties in the elastic and inelastic demand cases. Finally, Section 1.6 discusses some interesting applications and extensions of the approach.

1.2 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 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 $c_{pqr} := c_{pqr}(h)$ the travel cost on the route given the vector h of route flows; with this notation, an equilibrium flow is defined by the conditions

$$h_{pqr} > 0 \implies c_{pqr} = \pi_{pq}, \qquad \forall r \in \mathcal{R}_{pq}, \tag{1.1a}$$

$$h_{pqr} = 0 \implies c_{pqr} \geq \pi_{pq}, \qquad \forall r \in \mathcal{R}_{pq}, \tag{1.1b}$$

where the value of $\pi_{pq} := \pi_{pq}(h)$ 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 restated as a *variational inequality problem*: the conditions (1.1) 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 ($\Gamma \in \Re^{|\mathcal{R}| \times |\mathcal{C}|}$ is the route–OD pair incidence matrix and $d \in \Re_{++}^{|\mathcal{C}|}$ is the vector of demands), $c : \Re_+^{|\mathcal{R}|} \mapsto \Re_{++}^{|\mathcal{R}|}$ 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

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

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 $F := \{\, f \in \Re^{|\mathcal{A}|} \mid \exists h \in H \text{ with } f = \Delta h \,\}$, where $\Delta \in \Re^{|\mathcal{A}| \times |\mathcal{R}|}$ is the link–route incidence matrix, denote the (bounded polyhedral) set of demand-feasible link flows,[3] 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 $c(h) = \Delta^{\mathrm{T}} t(f)$, for any pair $(h, f) \in H \times F$.)

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

1.3 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

$$g_k(f) \le 0, \qquad k \in \mathcal{K}, \tag{1.2}$$

where $\mathcal{K}$ is a finite index set (e.g., formed by subsets of $\mathcal{N}$, $\mathcal{A}$, $\mathcal{C}$, and $\mathcal{R}$), and $g_k : \Re_+^{|\mathcal{A}|} \mapsto \Re$, $k \in \mathcal{K}$, is continuously differentiable on $\Re_+^{|\mathcal{A}|}$. The link flows that satisfy (1.2) form a closed set in $\Re^{|\mathcal{A}|}$, which we denote by G. We can envisage several types of such restrictions; examples include upper limits of flows on certain links and on the sum of flows on links leading into specified areas, as well as flow capacity constraints described by environmental quantities (e.g., exhaust fumes and noise).

Example 1 (Capacity constraints) The most immediate example of a set of flow restrictions is that of upper bounds on some links' flows. In the framework of (1.2), such constraints are described by letting $\mathcal{K} := \underline{\mathcal{A}}$, and

$$g_a(f_a) := f_a - u_a, \qquad a \in \underline{\mathcal{A}}, \tag{1.3}$$

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

Remark 2 (Types of restrictions) 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, 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, tolls and/or additional queueing delays for private vehicles in order to favour public transport, and so on. We discuss these applications in more detail in Section 1.6, and note here only that the introduction of OD pair (or, commodity) specific restrictions demands a further property of the underlying equilibrium model: if the restrictions are to be fulfilled through the use of (commodity-specific) tolls, then the equilibrium model must be such that the commodity flows are uniquely determined (see Section 1.6.5). □

We let the vector $\rho \in \Re^{|\mathcal{A}|}$ 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, $\bar{t}(f, \rho)$, of the link travel cost function. There is a large freedom in interpretation of these actions, and several possible examples were given in the introduction.

These actions are restricted to some set $P \subseteq \Re^{|\mathcal{A}|}$, 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 (continuous) 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 greater detail.)

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

[MPEC-TAP]

$$\text{minimize } \varphi(\rho, f) \tag{1.4a}$$

subject to

$$\rho \in P, \tag{1.4b}$$
$$f \in G, \tag{1.4c}$$

where

$$-\bar{t}(f, \rho) \in N_F(f). \tag{1.4d}$$

Remark 3 (Existence of solutions) Since the feasible set of [MPEC-TAP], $S := \{ (\rho, f) \in \Re^{2|\mathcal{A}|} \mid \rho \in P;\ f \in F(\rho) \cap G \}$, where $F(\rho)$ denotes the solution set to (1.4d) given $\rho \in P$, is closed by the assumptions, a solution to [MPEC-TAP] exists under the additional assumptions that S is nonempty (that is, there exists a $\rho \in P$ such that $F(\rho) \cap G \neq \emptyset$) and φ is *coercive* on S, that is, $\lim_{(\rho,f)\in S,\, \|(\rho,f)\|\to\infty} \varphi(\rho, f) = \infty$ (e.g., Luo et al., 1996, Section 1.4). □

Remark 4 (Uniqueness of user equilibrium flows) Should $\bar{t}(\cdot, \rho)$ be a strictly monotone mapping for every feasible choice of ρ, the lower-level (follower) problem (1.4d) has a unique solution, and as such f can be described as $f(\rho)$. Then, the function φ is in fact a function of ρ only, and [MPEC-TAP] reduces to a (non-convex) one-level problem. □

Mathematical programs with equilibrium constraints generalize bilevel optimization problems (a bilevel program is obtained whenever $\bar{t}(\cdot, \rho)$ is separable), and are in general non-convex. A comprehensive overview of the theory and algorithms for MPEC is found in Luo et al. (1996); for an overview of bilevel problems in transportation research, see Migdalas (1995).

Example 5 (Network design) A familiar form of the *equilibrium network design problem* is an instance of [MPEC-TAP]. Let ρ 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 $\bar{t}_a(f,\rho) := t_a(f_a/\rho_a)$. 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) plus the investment costs (that is, $\varphi(\rho,f) := \sum_{a\in\mathcal{A}}\{\bar{t}_a(f,\rho)f_a + \psi_a(\rho_a)\}$), while satisfying budget constraints on the investments made, $\rho \in P := \{\, \rho \in \Re_+^{|\mathcal{A}|} \mid \ell \leq \rho \leq u;\ \sum_{a\in\mathcal{A}}\rho_a \leq U \,\}$. Note then that in the general formulation of [MPEC-TAP], network performance may be measured also by constraints in the model. □

Example 6 (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 which, under user equilibrium conditions, optimizes some measure of the performance of the network, such as the total queueing delay. In this case, then, the variables ρ are the control parameters, for example the *green times* allocated to the signal controls, and the parameterized travel cost mapping $\bar{t}(\cdot,\rho)$ measures the travel times and the delays at intersections. (See Cantarella and Sforza, 1987, and Smith and Van Vuren, 1993, and references therein for examples of traffic control policies and mathematical models.) □

Remark 7 (Generalized Nash games) A Stackelberg game in which the leader imposes constraints on the followers' strategies is discussed by Harker (1991) in the context of generalized Nash games. He also derives the incentives (taxes or subsidies) that must be given to the players in order to make them obey the constraints, in the form of their dual variables. The type of constraints considered is, however, rather special, and there is no apparent relationship between his model and the one developed in this paper. □

In the following, we shall introduce assumptions on what the actions ρ represent which will enable us to solve the generally non-convex problem [MPEC-TAP] through the solution of two (potentially) convex problems.

1.4 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 travel costs can be described by $\bar{t}(f,\rho) := t(f) + \rho$, where ρ is the vector of toll fees.

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 making any restrictions in (1.2) saturated (that is, satisfied with equality). We also assume that the toll levied on a link can be decomposed into the contributions of all the different restrictions in (1.2), 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 also seems natural from the viewpoint of fairness.)

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

$$\rho_a := \sum_{k \in \mathcal{K}(f)} \beta_k \frac{\partial g_k(f)}{\partial f_a}, \qquad a \in \mathcal{A}, \tag{1.5}$$

where $\mathcal{K}(f)$ is the subset of $\mathcal{K}$ for which $g_k(f) = 0$ (that is, the set of saturated side constraints at f), and β_k, $k \in \mathcal{K}$, are non-negative constants. (In vector notation, then, $\rho := \nabla g(f)^{\mathrm{T}} \beta$ for a vector $\beta \in \Re_+^{|\mathcal{K}|}$ satisfying $\beta^{\mathrm{T}} g(f) = 0$.) This assumption, as we shall see, is equivalent to restricting ρ to be a *vector of Lagrange multipliers* for the side constraints (1.2), thus *removing its role as a set of independent variables.*

Example 8 (Capacity constraints) In the case of simple upper bounds (see Example 1), the expression (1.5) reduces to $\rho_a := \beta_a$, $a \in \underline{\mathcal{A}}$. In applications to queueing networks (see Larsson and Patriksson, 1995a and 1995b), ρ_a is interpreted as the equilibrium queueing delay on link a. The above complementarity condition then states that there is a queueing delay on the link (that is, $\rho_a > 0$) only if it is saturated (that is, if $f_a = u_a$). For more discussions on capacity constraints and queueing, see Remark 12. □

The effect of assumption (1.5) is more far-reaching than may 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 vectors f and ρ (through assigning ρ the role of being a Lagrange multiplier vector). Further, these single-level problems can, under some additional, but standard, assumptions, be formulated and solved as convex programs.

To avoid having to deal with 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*[4] (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 1.5.) The set $F \cap G$ is always compact, due to the compactness of F and the continuity of g_k, $k \in \mathcal{K}$.

With the assumptions (1.5) and the above on G in place, we observe that the requirements (1.4c)–(1.4d) are restricted to the primal–dual variational inequality problem of finding $(f, \rho) \in F \times \Re^{|\mathcal{A}|}$ satisfying[5]

$$-[t(f) + \rho] \in N_F(f), \tag{1.6a}$$
$$\rho \in N_G(f). \tag{1.6b}$$

Remark 9 (Properties of (1.6)) (a) The assumption (1.5) and the separability condition on $\bar{t}$ introduced above can be viewed as a restriction and approximation, respectively, introduced in the original MPEC problem (1.4) in order to reduce its complexity (the assumptions made on the toll scheme have the effect of making it possible to place the constraints (1.4c) at the lower level of [MPEC-TAP]; this is instrumental in making the problem tractable); whether this type of strategy can actually be used within an iterative scheme for solving [MPEC-TAP] is an interesting subject for further research.

(b) The formulation (1.6) is valid also when the management goals define non-convex constraints, through a proper re-definition of the normal cone mapping N_G (see, e.g., Clarke, 1983, Section 2.4).

(c) It is clear from the expression (1.6b) that the toll vector ρ is independent of the way in which the set G is described in terms of constraints. If, for example, the restrictions in (1.2) 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.[6] This assumes, of course, that the problem (1.6) is consistent; the effect of scaling in the inconsistent case is discussed in Example 16. □

In order for the vector ρ to be naturally associated with a toll, that is, for ρ to be non-negative, $N_G(f) \subseteq \Re_+^{|\mathcal{A}|}$ must hold at every solution f to (1.6). 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. (1.5)]. If the condition is not fulfilled, then some link tolls may be negative in a solution f, whence the "toll" should be interpreted 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 1.6.4.

Example 10 (Capacity constraints) The case of upper bounds on the link flows (cf. Examples 1 and 8) is one in which, clearly, tolls will be non-negative. □

We now observe that $N_{F\cap G}(f) = N_F(f) + N_G(f)$ holds everywhere (e.g., Rockafellar, 1981, Thm. 5C), and so solving (1.6) is equivalent to finding a solution f to[7]

[TAP-VIP-SC-F]
$$-t(f) \in N_{F\cap G}(f).$$

Because of the nonemptiness and convexity assumption on the set $F \cap G$, the problem [TAP-VIP-SC-F] is guaranteed to have a solution (e.g., Hartman and Stampacchia, 1966).

Remark 11 (Behavioural versus non-behavioural equilibrium models) The side constrained traffic equilibrium problem [TAP-VIP-SC-F] has earlier been studied in detail in Larsson and Patriksson (1995b and 1997b) (in the optimization setting), and in Larsson and Patriksson (1994). These studies concern the introduction of side constraints as a means to improve the descriptive capabilities of the standard traffic equilibrium model, such as the effect of traffic control policies or to introduce additional information about the traffic flow situation at hand. As such, these models are *behavioural*, in the sense that the effects of the side constraints are assumed to be immediately transferable to the perception of travel costs among the trip-makers, for example as queueing delays; their solutions are also characterized and interpreted as flows satisfying the Wardrop equilibrium conditions in terms of generalized travel costs that include link queueing delays. In contrast, the present model is decision-based, and thus inherently *non-behavioural.* □

Remark 12 (Queueing models) **(a)** In applications of side constrained traffic equilibria to queueing networks, we have in Larsson and Patriksson (1994 and 1997b) 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 adjustment of the tolls made during a non-stationary (disequilibrium) state. (Essentially, as time passes, the toll contributed by a restriction in (1.2) 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.)

(b) The characterizations and interpretations of solutions to [TAP-VIP-SC-F] as equilibrium flows in terms of link travel costs and queueing delays were earlier formulated by Payne and Thompson (1975) and Miller et al. (1975) [see also Smith, 1987], albeit for networks with constant travel costs. See Larsson and Patriksson (1994, 1995a and 1997b) for detailed accounts of the historical development of queueing models, and characterizations of and numerical procedures for their solutions.

(c) 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. Examples of such models are given by Yang and Lam (1996) and Yang and Bell (1997). □

One cannot assume in general that the set of multipliers β_k for the flow restrictions (1.2) is a singleton, and therefore, while the link flow solution is unique whenever the travel cost mapping t is strictly monotone, the toll vector ρ which yields the desired equilibrium solution may not be unique. In the context of queueing networks, this can be considered a drawback, since the vector ρ is, in that case, interpreted as a queueing delay. So, in Larsson and Patriksson (1994 and 1997b), sufficient conditions were therefore derived under which this equilibrium link queueing delay is unique. In the context of traffic management through link tolls, however, *the non-uniqueness of the multipliers β_k is a positive feature of the model, since it facilitates the fulfilment of the traffic management goals while also satisfying exogenous restrictions (practical, political, environmental and economical constraints) on the toll scheme. Furthermore, one could choose a toll vector which simultaneously optimizes the secondary performance function φ.* The non-uniqueness of the tolls is, seemingly, a property with many interesting consequences; we are, however, not aware of any previous observations of this fact, besides those for the system optimal problem (see Section 1.6.3), and of no model in which exogenous constraints on the tolls are considered. In the next section, we shall elaborate on this further, and study means for solving the toll optimization model.

1.5 TOLL OPTIMIZATION

In this section, we study the upper-level problem of [MPEC-TAP] in terms of the toll vector ρ. 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^*. (This assumption is crucial in order to ensure that the flow f^* can be obtained through the use of link tolls; see further Section 1.6.5 for discussions on this issue in relation to commodity-specific traffic management goals.)

The set of toll vectors which achieves this solution while satisfying the requirement (1.5) is

$$T := \{\, \rho \in \Re^{|\mathcal{A}|} \mid -[t(f^*) + \rho] \in N_F(f^*); \quad \rho \in N_G(f^*) \,\}, \tag{1.7}$$

that is, the set of solutions to (1.6) with $f = f^*$. This set 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 (e.g., Gould and Tolle, 1971). We are then in a position to describe the conditions for f^* to solve [TAP-VIP-SC-F] in terms of its generalized Karush–Kuhn–Tucker conditions, thus providing a representation of T in terms of linear constraints that will be much more useful in practice than its implicit description in (1.7).

Using linear programming duality on (1.6a) [identifying f with f^*], the set T can be described as the (polyhedral) set of solutions in ρ of the following linear system in ρ, β and π:

$$\Delta^{\mathrm{T}}[t(f^*) + \rho] - \Gamma\pi \geq 0, \tag{1.8a}$$
$$[t(f^*) + \rho]^{\mathrm{T}} f^* - d^{\mathrm{T}}\pi = 0, \tag{1.8b}$$
$$\rho - \nabla g(f^*)^{\mathrm{T}}\beta = 0, \tag{1.8c}$$
$$\beta^{\mathrm{T}} g(f^*) = 0, \tag{1.8d}$$
$$\beta \geq 0; \tag{1.8e}$$

we note that (1.8a)–(1.8b) is a dual description of the normal cone inclusion (1.6a), representing dual feasibility and strong duality, respectively, and (1.8c)–(1.8e) is an explicit description of the corresponding inclusion (1.6b).[8] The constraints (1.8a) render this system intractable in practice, since it contains as many constraints as there are routes in the network (cf. the dimensions of Δ 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 satisfying (1.8c)–(1.8e), and should be comparatively very small. The algorithm to be outlined below is a route (column) generation procedure that generates these routes as needed.

1.5.1 The toll optimization problem

The toll optimization problem is

[TOP]

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

The problem has an optimal solution whenever $P \cap T$ *is non-empty,* and $\varphi : \Re^{|\mathcal{A}|} \mapsto \Re$ *is coercive* on this set (cf. Remark 3), by Weierstrass' Theorem. (The latter holds in particular if P is a bounded set.) By the polyhedral property of T, the problem is convex whenever P *is convex and* φ *is convex.* (It is further a linear program if P is polyhedral and φ is a linear function.) The algorithm to be presented relies on the *convexity of the problem, which hence is assumed to hold throughout.*[9]

Remark 13 (The inverse shortest routes problem) We note that the problem [TOP] is an extension to the *inverse shortest routes* problem, in which one is given a set of routes and wishes to recover the link costs that will result in these as shortest routes while also minimizing some merit function; see Burton and Toint (1992) for a description of the problem and an application in transportation planning. □

Example 14 (Instances) Several interesting instances of the problem [TOP] are conceivable. The set P may be defined by lower and upper bounds on the toll fees, and by a maximal number of tolled links. We may wish to levy the least total toll in the network, whence we choose $\varphi(\rho) := \sum_{a \in \mathcal{A}} f_a^* \rho_a$, or precisely the opposite, thus maximizing this function. 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; for example, the problem of minimizing the number of tolled links can be formulated as a mixed integer problem. □

Remark 15 (Unboundedness of the toll set) A numerical investigation that we recently have conducted on small and medium-size instances of the link flow capacitated side constrained version of [TAP-VIP-SC-F] (cf. Example 1) has shown that the toll set T is very likely to be unbounded; in consequence, it has always been possible to make the toll revenue as large as desired; see Section 1.5.3 for a numerical example with such a property. (It is, however, easy to construct trivial instances where the multiplier vector β^* is unique, so it is *not* always true that T is unbounded.) The restriction of the set of feasible tolls to the intersection of T with a bounded set P seems, therefore, to be crucial for obtaining a useful result.

We note that this remark is made with reference to the inelastic demand case only; the total toll revenue is unique in the elastic demand case (see Section 1.5.4). □

1.5.2 An algorithm for the toll optimization problem

1.5.2.1 Solving the problem [TAP-VIP-SC-F]. We consider solving [TAP-VIP-SC-F] using a nonlinear 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 (1.5), a toll vector $\rho \in T$) 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 is found in Section 1.5.2.3.) We refer the reader to Hearn and Ribera (1980), Patriksson (1994, Sec. 4.6) 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) if we wish to utilize the representation (1.8) of the set T in the solution of [TOP], then we must choose an algorithm that generates the necessary routes explicitly; 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).[10]

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. (1.6a)] all be least-cost routes given the fixed link travel cost vector $t(f^*) + \rho$, where $\rho := \nabla g(f^*)^{\mathrm{T}}\beta$ and β is the vector of multipliers for (1.2) which is obtained in the limit of the multiplier scheme.

1.5.2.2 A regularity condition. Larsson and Patriksson (1994) introduce a regularity condition on the solution to [TAP-VIP-SC-F], such that the set of least-cost routes is invariant over the set of multipliers β for the side constraints, and hence over the set $\rho \in T$ (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 toll set T explicitly.

If this is not the case, then the least-cost routes are not the same for every value 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 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. Indeed, our numerical experience is that the number of routes that needs to be generated after termination of the augmented Lagrangean/DSD algorithm for solving [TAP-VIP-SC-F] is very limited (see Remark 19).

1.5.2.3 The case of primal inconsistency. If the traffic manager has set goals that are unattainable, then the result will be that $F \cap G = \emptyset$. 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* $\hat{\rho} = \nabla g(\hat{f})^{\mathrm{T}} \hat{\beta}$ *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 (1.2) 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.*

Example 16 (Goal programming approach) 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 9.c), the resulting toll vector will change when introducing different multiplicative 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 (compare 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.*

1.5.2.4 Solving [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 $\rho \in T$ are given. The algorithm also supplies subsets $\overline{\mathcal{R}}_{pq}$

of $\mathcal{R}_{pq}$, $(p,q) \in \mathcal{C}$, corresponding to routes that are used in the route flow solution to [TAP-VIP-SC-F] generated. These subsets of routes correspond to a subset of the constraints in (1.8a), and we thus have at hand an explicit description of an outer approximation, $\overline{T}$, of T (that is, $\overline{T} \supseteq T$); if we denote by $\overline{\Delta}$ and $\overline{\Gamma}$ the sub-matrices of Δ and Γ corresponding to the subsets of the routes generated, the set $\overline{T}$ is the set of solutions in ρ to the restriction of the system (1.8) where $\overline{\Delta}$ and $\overline{\Gamma}$ replace Δ and Γ, respectively, in (1.8a).

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 ρ 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 (that is, belongs to P) and optimal (that is, 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 ρ as a starting point, we search for a point in the intersection of P and $\overline{T}$. 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); whenever P is a polyhedron or a bounded set, it is a well-defined problem in the sense that the minimum distance between the two sets is attained. 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 *soft* constraints.[11] We note that in the case where P is described by linear constraints or when $P = \Re^{|\mathcal{A}|}$ (that is, no toll constraints are imposed), 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 the sets P and $\overline{T}$ are disjoint. 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 ρ 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 obtainable 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 belongs to $T \setminus P$.) We can, however, expect that by using the toll vector obtained, at least a subset of the goals described by (1.2) 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 $P \cap \overline{T} \neq \emptyset$, and then the output is a toll vector $\overline{\rho}$ belonging to $P \cap \overline{T}$. The reader is, however, warned against interpreting this as a confirmation that the problem [TOP] is consistent. That is investigated in the next step.

The toll vector $\overline{\rho}$, 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 $\overline{\rho} \in T$ actually holds, by checking whether any constraints in (1.8a) are violated. With the link costs $t(f^*) + \overline{\rho}$ at hand, the most violated, if any, constraint in (1.8a) for an OD pair (p,q) is obtained by finding the shortest route, $r \in \mathcal{R}_{pq}$ say; the constraints (1.8a) for the OD pair are satisfied if and only if $\sum_{a\in\mathcal{A}} \delta_{pqra}[t_a(f^*) + \overline{\rho}_a] \geq \overline{\pi}_{pq}$; 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 $\overline{\rho} \in T$ at hand, or to the generation of constraints (that is, routes) in some OD pairs. In the latter case, we augment the sets $\overline{\mathcal{R}}_{pq}$ accordingly, and thus obtain a better outer approximation of T; the Phase-I problem is then re-solved. 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 $\overline{\rho} \in P \cap T$.

Having already discussed the case of dual inconsistency, we henceforth assume that *the Phase-I procedure was terminated with a toll vector in* $P \cap T$,[12] 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 relaxation of [TOP], where T is replaced by $\overline{T}$.[13] We denote this problem by $[\overline{\text{TOP}}]$. Let $(\overline{\rho}, \overline{\beta}, \overline{\pi})$ be a solution to this problem.[14] We next investigate if this solution solves [TOP], by checking whether any constraints in (1.8a) 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 $(\overline{\rho}, \overline{\beta}, \overline{\pi})$ being a *proven optimal solution to* [TOP], or in the generation of constraints (routes) in some OD pairs. In the latter case, we augment the sets $\overline{\mathcal{R}}_{pq}$ accordingly, and thus obtain a better outer approximation of T; $(\overline{\rho}, \overline{\beta}, \overline{\pi})$ is used as an (infeasible) starting solution when solving the improved approximation $[\overline{\text{TOP}}]$ of [TOP]. Obviously, since the number of routes is finite, this procedure is finite, and its output is a feasible and optimal toll vector, ρ^*, to [TOP].[15]

Remark 17 (Unbounded solution to $[\overline{\text{TOP}}]$) Unless φ is coercive on $P \cap \overline{T}$ (cf. Remark 3), it is not guaranteed that $[\overline{\text{TOP}}]$ has a solution (as assumed above). Assume then that this problem has an unbounded solution.[16] The feasibility test corresponding to the one above for the case of a finite optimal solution then is as follows. The unbounded solution is reached via a ray $\nu := (\nu^\rho, \nu^\beta, \nu^\pi)$, whose component $\nu^\rho \in \Re_+^{|\mathcal{A}|}$ describes a feasible direction in $P \cap \overline{T}$ (in the case where [TOP] is a linear program, this ray may be identified from the last simplex tableau). Construct a subnetwork $(\mathcal{G}, \overline{\mathcal{A}})$ with $\overline{\mathcal{A}} := \{\, a \in \mathcal{A} \mid \nu_a^\rho = 0 \,\}$, that is, a subnetwork in which only links with tolls remaining finite along the ray ν^ρ are included. For each OD pair (p, q) we then investigate if there exists a route from p to q in this network; if there is, then it is a route that clearly was not included in the known subset of the routes in $\mathcal{R}_{pq}$ (since its cost is finite, it would otherwise have violated the restriction of the Wardrop conditions (1.8a) to this subset).

In practice one finds such a route, if one exists, by solving the problem of finding the shortest route between p and q, where a link a in $\overline{\mathcal{A}}$ is assigned a cost equal to the sum of $t_a(f^*)$ and the link toll $\overline{\rho}_a$ at the origin of the ray ν, while the links in $\mathcal{A} \backslash \overline{\mathcal{A}}$ are assigned costs sufficiently large to prevent them from belonging to a shortest route; thus, in some sense, we search for a most violated constraint in (1.8a) also in this (unbounded) case. Performing this test for each OD pair will either result in termination, with the conclusion that [TOP] has an unbounded solution, or in the generation of some new routes, and a continuation of the algorithm with an improved approximation $\overline{T}$ of T, using, for example, the vector $(\overline{\rho}, \overline{\beta}, \overline{\pi})$ corresponding to the origin of the ray ν as the (infeasible) starting point. □

We summarize the algorithm stated above in Table 1.1.

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

Table 1.1 The toll optimization algorithm

---

0. (*Initialization*): Solve [TAP-VIP-SC-F] using the augmented Lagrangean/DSD algorithm. If $F \cap G = \emptyset$, then terminate ([TAP-VIP-SC-F] is inconsistent). Otherwise, output is f^*, $\rho \in T$, $\overline{\mathcal{R}}_{pq} \subseteq \mathcal{R}_{pq}$, $(p,q) \in \mathcal{C}$, and sub-matrices $\overline{\Delta}$ and $\overline{\Gamma}$, defining $\overline{T}$.

1. (*Phase-I*): Search for $\rho \in P \cap \overline{T}$. If $P \cap \overline{T} = \emptyset$, then terminate ([TOP] is inconsistent).

2. (*Feasibility test*): For each $(p,q) \in \mathcal{C}$, solve a shortest route problem given link costs $t(f^*) + \rho$. If the shortest route in any OD pair (p,q) does not belong to $\overline{\mathcal{R}}_{pq}$, then augment it with the new route; update the set $\overline{T}$, and go to Step 1.

3. (*Phase-II*): Solve $[\overline{\text{TOP}}]$. If $[\overline{\text{TOP}}]$ has an unbounded solution, then output is a ray ν^ρ; set M to a very large value. Otherwise, output is an optimal solution $\overline{\rho}$; set $M := 0$.

4. (*Feasibility test*): For each $(p,q) \in \mathcal{C}$, solve a shortest route problem given link costs $t(f^*) + \overline{\rho} + M\nu^\rho$. If the shortest route in any OD pair (p,q) does not belong to $\overline{\mathcal{R}}_{pq}$, then augment it with the new route; update the set $\overline{T}$, and go to Step 3.

5. (*Termination*): If $[\overline{\text{TOP}}]$ has an unbounded solution, then so does [TOP], otherwise $\overline{\rho}$ is an optimal solution to [TOP].

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.

1.5.3 A numerical example

Through a small numerical example we illustrate some of the features of the model discussed above.

We consider a directed graph consisting of 5 nodes and 6 links, with cost functions given in Table 1.2; there are two OD pairs, $(1,5)$ and $(2,5)$, with a demand of 2 and 3, respectively.

The solution to this instance of [TAP-VIP-F] (that is, in the absence of the upper bounds) is $f \doteq (2.00, 2.09, 0.91, 1.27, 2.82, 2.18)^{\mathrm{T}}$, while the solution to [TAP-VIP-SC-F] is $f^* = (2,2,1,2,2,3)^{\mathrm{T}}$. Applying the augmented Lagrangean algorithm of Larsson and Patriksson (1995a) outlined in Section 1.5.2.1 to this problem produces the multiplier vector $\beta \doteq (0, 0.219, 0.219, 0, 8.0, 0)^{\mathrm{T}}$.

Table 1.2 Network data

Link	$t_a(f_a)$	u_a
1: $(1,3)$	$t_1(f_1)=f_1$	2
2: $(2,3)$	$t_2(f_2)=4f_2$	2
3: $(2,4)$	$t_3(f_3)=12f_3$	1
4: $(3,4)$	$t_4(f_4)=2f_4$	2
5: $(3,5)$	$t_5(f_5)=4f_5$	2
6: $(4,5)$	$t_6(f_6)=4f_6$	4

Observing that, in this case (cf. Example 8), $\rho=\beta$ holds, the set T of link tolls satisfying (1.8) is the set of solutions in ρ to

$$
\begin{array}{cccccccccc}
\rho_1 & & & +\rho_4 & & +\rho_6 & -\pi_{15} & & \geq & -18, \\
\rho_1 & & & & +\rho_5 & & -\pi_{15} & & \geq & -10, \\
 & \rho_2 & & +\rho_4 & & +\rho_6 & & -\pi_{25} & \geq & -24, \\
 & \rho_2 & & & +\rho_5 & & & -\pi_{25} & \geq & -16, \\
 & & \rho_3 & & & +\rho_6 & & -\pi_{25} & \geq & -24, \\
2\rho_1 & +2\rho_2 & +\rho_3 & +2\rho_4 & +2\rho_5 & +3\rho_6 & -2\pi_{15} & -3\pi_{25} & = & -92, \\
\rho_1, & \rho_2, & \rho_3, & \rho_4, & \rho_5, & \rho_6, & & & \geq & 0, \\
 & & & & & \rho_6 & & & = & 0.
\end{array}
$$

To illustrate that the toll set T may not be a singleton, we consider solving $\min_{\rho\in T}\varphi(\rho) := (f^*)^{\mathrm{T}}\rho$, that is, we identify the toll schedule yielding the *smallest* revenue. (Note that $P=\Re^{|\mathcal{A}|}$ in this case.) The solution to this problem is $\rho^*=(0,0,0,0,8,0)^{\mathrm{T}}$, with $\varphi(\rho^*)=16$ and $\pi^*=(18,24)^{\mathrm{T}}$. Compared to this solution, the toll vector obtained by the augmented Lagrangean algorithm yields the larger toll revenue $\varphi(\beta)\doteq 16.657$; perhaps more important for practical purposes is however the observation that the latter toll is distributed over three links while the former toll scheme requires only one toll link.

To further illustrate the nature of the toll set T, we also solve $\max_{\rho\in T}\varphi(\rho) := (f^*)^{\mathrm{T}}\rho$, that is, we identify the toll schedule with the *largest* revenue. This maximization problem has an unbounded optimal solution [take $\rho=(0,\nu,\nu,0,8,0)^{\mathrm{T}}$, where $\nu\rightarrow+\infty$]. Clearly, then, not only is the set T a non-singleton set, it is in fact unbounded, and φ is unbounded above on it.

Remark 18 (Non-unique equilibrium link queueing delays) It has often been claimed that in capacitated user equilibrium models, the equilibrium link queueing delay is unique. Viewing the above example as a queueing model clearly demonstrates that this claim is not always valid; in the light of this result, the output from a queueing model must be used with the greatest of caution. □

Remark 19 (Numerical experience) We have solved seven instances of [TOP], constructed from small and medium-scale instances of capacitated inelastic user equilibrium models, and provide here some experience from these experiments.

In all but one case, the set T is unbounded, and the total revenue is also unbounded above on it. This phenomenon occurs in situations similar to that of the above example, namely when it is possible to identify a subset of the saturated links having the property that, within each OD pair, the number of those links that are used by the routes is the same; when adding the same constant to the toll fee on all those links, the changes in the costs are identical on all routes within the OD pair, and hence Wardrop's conditions are fulfilled for an arbitrary, but equal, change in the tolls on those links. Clearly, the occurrence of this phenomenon depends on the network topology as well as on the level of congestion; *our hypothesis is that in highly congested large-scale networks the set T is a singleton.*

With respect to the problem of minimizing the total revenue, we note that the toll supplied by the augmented Lagrangean algorithm is always very close to the optimal one, which further motivates its use. The number of routes generated at termination of this scheme is $(11, 30, 50)$, signifying the lowest, median and maximum number taken over the six smallest instances, respectively, while the corresponding figures for the routes generated within the algorithm for [TOP] are $(0, 4, 14)$. This shows that the solutions to [TAP-VIP-SC-F] are not regular in the sense of Section 1.5.2.2, but the degeneracy present is mild.

An important computational issue is the fact that the construction of the set T requires f^* to be known exactly. Since this is not possible in general, the solution of [TOP] must be supplied with some tolerance with respect to the system (1.8), which *always is inconsistent* when f^* is replaced by an only near-optimal solution to [TAP-VIP-SC-F]; our remedy has been to allow tolls also on links that are close to being saturated. □

1.5.4 The elastic demand case

Consider an extension of the model (1.6), in which the demand is not fixed, but 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 non-negative, upper bounded and continuous for each $(p, q) \in \mathcal{C}$, and such that $-g$ is strictly monotone on $\Re_+^{|\mathcal{C}|}$.*

The elastic demand extension of (1.6) is

$$(-[t(f) + \rho], g^{-1}(d)) \in N_{F_d}(f, d), \tag{1.9a}$$

$$\rho \in N_G(f), \tag{1.9b}$$

where $F_d := \{ (f, d) \in \Re_+^{|\mathcal{A}|+|\mathcal{C}|} \mid \exists h \in \Re_+^{|\mathcal{R}|} \text{ with } \Gamma^{\mathrm{T}} h = d; \quad f = \Delta h \}$; the solution to this side constrained elastic demand traffic equilibrium problem is unique in the demands and link flows, whose values we denote by (d^*, f^*), and the set of link tolls achieving them, which we shall denote by T_d, is the (polyhedral) set of solutions in ρ to the linear system

$$\Delta^{\mathrm{T}}[t(f^*) + \rho] - \Gamma\pi \geq 0, \tag{1.10a}$$

$$[t(f^*) + \rho]^{\mathrm{T}} f^* - (d^*)^{\mathrm{T}}\pi = 0, \tag{1.10b}$$

$$-g^{-1}(d^*) + \pi \geq 0, \tag{1.10c}$$

$$[-g^{-1}(d^*) + \pi]^{\mathrm{T}} d^* = 0, \tag{1.10d}$$

$$\rho - \nabla g(f^*)^{\mathrm{T}}\beta = 0, \tag{1.10e}$$

$$\beta^{\mathrm{T}} g(f^*) = 0, \tag{1.10f}$$

$$\beta \geq 0, \tag{1.10g}$$

which is derived in a manner similar to that for (1.8).

Remark 20 (Unique toll revenues) Since the demand for transportation is influenced by the total travel cost, the choice of link tolls for achieving the solution

(d^*, f^*) to the side constrained user equilibrium problem is very limited. For example, combining (1.10b) and (1.10d) leads to the conclusion that *the total toll revenue is unique*. Further, by (1.10a) and (1.10c), the toll fee for a traveller on any route is lower bounded, and π_{pq} (the least total travel cost) is uniquely determined for all OD pairs (p, q) with $d^*_{pq} > 0$. There is hence a very limited variability of the link tolls in the context of elastic demand models, in contrast to the fixed demand case in which there is no option *not* to travel; the only freedom-of-choice essentially amounts to transferring a toll fee between links on a route, and then only between links that in principle could be collapsed to one link. □

1.6 APPLICATIONS AND RELATIONS TO PREVIOUS WORK

1.6.1 Capacity constraints

1.6.1.1 Models. In the case of capacity constraints (see Example 1), the toll ρ_a equals the Lagrange multiplier β_a associated with the corresponding constraint in (1.3); see Example 8. Hearn (1980) uses these multipliers to define a perturbation of the travel time formulas so that an equivalent, un-capacitated assignment model can be solved. Characterizations of solutions to capacitated traffic assignment models have been presented also in the case where the capacity side constraints are the result of the presence of traffic controls; then, the multipliers β are the equilibrium queueing delays at the signalized junctions. (See Larsson and Patriksson, 1995a, for an overview of capacitated models and their properties.) The use of the multipliers as tolls is implicit in the work of Beckmann and Golob (1974), albeit for a model governed by the principle of system optimality.

Ferrari (1995, 1997) considers an elastic demand model with *environmental* capacity side 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.

It is to be noted that the above models define special cases of that developed here. Further, in none of these works is it observed that the tolls that should be levied to achieve the goals are not necessarily unique; consequently, the interesting applications that are brought forward by this fact are not recognized.

1.6.1.2 Methods. 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 the link flow f^* and 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.

1.6.2 Obtaining a particular 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-TAP] applies to such models, with the set of side constraints described by

$$g_a(f_a) := f_a - f^*_a = 0, \qquad a \in \mathcal{A}. \tag{1.11}$$

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 (1.6b) [or, (1.9b)] poses no further restrictions on ρ, and the set of tolls achieving f^* is thus, by definition,

$$\tilde{T} := \{ \rho \in \Re^{|\mathcal{A}|} \mid -[t(f^*) + \rho] \in N_F(f^*) \},$$
$$\tilde{T}_d := \{ \rho \in \Re^{|\mathcal{A}|} \mid (-[t(f^*) + \rho], g^{-1}(d^*)) \in N_{F_d}(f^*, d^*) \},$$

in the inelastic and elastic demand cases, respectively.

Clearly, $\tilde{T} \supseteq T$ and $\tilde{T}_d \supseteq T_d$. Also, the solution of [TOP] is simplified substantially, since $\tilde{T}$ is described by the simpler system (1.8a)–(1.8b), and $\tilde{T}_d$ by (1.10a)–(1.10d). Moreover, in contrast to the sets T and T_d (cf. the discussion in Remark 19 on the inconsistency of the system (1.8) when f^* is replaced by a near-optimal solution to [TAP-VIP-SC-F]), these sets will be nonempty as long as f^* and (f^*, d^*), respectively, are compatible with the network constraints (cf. the relation to the inverse shortest routes problem discussed in Remark 13).

Example 21 (A larger toll set) Let f^* be the solution to [TAP-VIP-SC-F]. Using model (1.11) will result in a larger freedom in choosing the tolls compared to that given by the toll set T; it will, for example, be possible to levy a toll on a link which does not contribute to the saturation of any flow restrictions. With reference to Remark 15, we note that it is in this case easy to show that, in contrast to the set T, $\tilde{T}$ *is always unbounded,*[17] *and the toll collected can be made arbitrarily large.* Since negative tolls are allowed in this case, the link flow solution f^* can always be achieved also through subsidies to the travellers; the function $\rho \mapsto (f^*)^{\mathrm{T}}\rho$ is actually unbounded below (as well as above) on $\tilde{T}$. This fact is illustrated by viewing the example of Section 1.5.3. Note, however, from the discussions in Section 1.5.4, that *the total link toll collected will be constant in the elastic demand case even when allowing negative tolls and tolls on unsaturated links.* □

Remark 22 (Well-posedness) It is clear that, under the given assumptions, the mathematical model for achieving management goals that is described by [MPEC-TAP] is not guaranteed to be well-posed in the fixed demand case, since any demand-feasible link flow desired can be obtained through the use of arbitrarily large toll fees or subsidies. An important step towards reaching a well-defined traffic management model is the future study of natural conditions that can be imposed on the tolls (that is, the set P). □

If the given flow f^* does not belong to F (or F_d, together with any consistent demand vector d^*), then it is not possible to achieve it using any toll vector. In this case, the use of the multiplier scheme outlined in Section 1.5.2 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.

Example 23 (Goal targets) In the model of Nagurney and Ramanujam (1996) the flow f^* is a *goal target*; deviations from it are penalized through the use of penalty functions which are used to derive taxes or subsidies on the links depending on whether a link flow is above or below the target value. Only in the case where f^* is consistent with F can the two models produce the same result; the output from our model in the case of inconsistency was discussed above; the output from their model depends heavily on the choice of penalty functions, guidelines for which are not provided. □

1.6.3 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 that has a strictly increasing travel cost function.) 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 and colleagues. Bergendorff (1995) considers a link capacitated model, in the presence of elastic demands and/or multiple modes of transportation. Following the work of Hearn (1980) and Larsson and Patriksson (1994) she derives equilibrium characterizations of its solutions, and studies especially the case where the upper bound u equals the system optimal solution. For this problem, a toll optimization problem is defined over the set of tolls achieving the system optimal solution; in particular, it is shown that a toll vector can be derived such that the total toll fee is never higher than that collected by using traditional marginal cost pricing. 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 of this paper, obtained by choosing side constraints of the form (1.11) with f^* being a system optimal solution.

Remark 24 (Marginal cost pricing is optimal) As noted by Bergendorff (1995) and following the discussion in Example 21 it is in the fixed demand case often possible to find toll schemes that achieve a system optimal solution while collecting a smaller toll fee than the marginal cost pricing scheme does. It is then interesting to note, by the discussion made in Section 1.5.4, that in the elastic demand case, the total toll collected is uniquely determined by the flow obtained. Therefore, if a system-optimality-inducing toll scheme is sought such that the total toll collected is minimized, then, *in the case of elastic demands, marginal cost pricing is optimal.* It goes without saying that the optimal solution is a zero flow, whenever the corresponding marginal cost pricing scheme is feasible with respect to P. □

Through the generality of the traffic management scheme described in this paper, we can consider more general goals than the achievement of system optimal flows; for example, we can introduce restrictions of the form (1.11) 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 (e.g., Harker, 1988) used in the modelling and simulation of vehicle guidance systems.

1.6.4 Network improvement, subsidies and the redistribution of resources

Example 25 (A network design model) 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 ρ other than (1.6a) 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 of ρ_a as a toll levied on link a, a negative value of ρ_a may be interpreted as the *travel time improvement* (or, subsidy) 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 a viewpoint that is the reverse to that of the traditional one, whose first priority is to stay within the given budget. □

Example 26 (Subsidies to public transport) 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

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

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 $P := \{ \rho \in \Re^{|\mathcal{C}|} \mid \sum_{(p,q)\in\mathcal{C}} [\rho_{pq}^{\mathrm{pr}} + \rho_{pq}^{\mathrm{pu}}] = 0 \}$ 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 or through the design (or, improvement) of designated lanes for the public transport vehicles or through lower ticket prices. □

1.6.5 Consistent toll schedules

Wardrop's user equilibrium principle states that costs are equal among the routes utilized in an OD pair, but it does not contain a mechanism by which their respective flows can be uniquely determined. (In technical terms, equilibrium commodity flows are not unique because the link cost mapping t is not strictly monotone in the commodity flow variables.) This inherent limitation of the Wardrop user equilibrium principle has two negative implications, which we shall shortly describe below; possible remedies to this limitation are then discussed.

1.6.5.1 The use of toll revenues. There are several reasons why one might wish to obtain disaggregated information about the traffic that passes a particular toll station, rather than just the total volume. Disaggregated traffic information could distinguish between (1) local, regional and long-distance traffic; (2) different vehicle modes such as cars, trucks and buses; (3) resident, commuter and commercial traffic; and (4) drivers with different values-of-time. A fundamental question raised in studies of toll schemes is: *how should the toll revenues be used?* As pointed out by Small (1992), "the people who benefit from congestion relief and revenue uses do not necessarily coincide with those who pay the fees or who suffer inconvenience in order to avoid them." Having the type of information mentioned above facilitates an analysis of the pricing policy, for example with the objective of determining which categories of travellers that are affected by the toll, and who among those that need to be compensated or who pay less than their contribution to congestion or environmental effects warrants. (See also Rossi et al., 1989, for a discussion on consistent impact-fee assessment in urban networks.) Based on such an analysis, a current revenue distribution could be re-evaluated. We then note that the Wardrop principle (in the case of additive cost structures) does not allow for these categories of traffic to be distinguished in the equilibrium solution, only their aggregates.

1.6.5.2 Commodity-specific management goals. Suppose that some management goals are described in terms of individual commodity flows. For example, we may consider restrictions of the form $g_{pq}(h_{pq}) \leq 0$, where h_{pq} is the vector of route flows in commodity (p,q). (See Remark 2 for further examples of possible restrictions over individual flows.) If we are interested in obtaining flows satisfying these restrictions through the use of tolls, then the tolls would have to be imposed on the individual commodities, that is, it would be insufficient to introduce a toll based on the total link flow only. (This is quite evident from the equilibrium characterization of such a side constrained traffic equilibrium model.) Having solved the side constrained model, we obtain a unique link flow solution, and a consistent commodity flow which satisfies the side constraints. The commodity-specific toll which is automatically derived will, however, not necessarily influence the flow to satisfy the side constraints, even though the unique link flow solution will be obtained. The reason for this apparently counter-intuitive result is that while the link flow solution is unique, its decomposition into commodity flows is not, and among these commodity flows there are some that do not satisfy all the side constraints.[18] (In technical terms, the tolled problem is non-strictly convex in the commodity flows, a property which in the context of Lagrangean duality bears the term *non-coordinability*.)

1.6.5.3 Remedies. There are basically two remedies to this problem.

The first, and most natural, remedy is the introduction of the principle of a *stochastic* user equilibrium, by which travellers are allowed to have different perceptions of the actual travel costs. (This has the additional advantage of being a more general model of route-choice.) In most stochastic models proposed (see, e.g., Patriksson, 1994, Sec. 2.8.1) the *stochastic equilibrium route flows are unique*. In the context of the discussion on the use of toll revenues, then, the implication of this property is that the different categories of traffic that passes the toll station could be distinguished; in the example concerning commodity-specific traffic management goals, this property has the immediate implication that tolls (or, taxes) that will achieve the traffic management goals could be determined. The results presented in this paper are easily transfered from the deterministic case considered to that of a framework of stochastic equilibrium models, with Wardrop's user equilibrium principle replaced by the principle of stochastic user equilibrium.

The second, and less general, remedy is to introduce assumptions in the original equilibrium model which makes the commodity flows unique; for example, one may introduce non-additive travel costs such that the route costs are strictly monotone.

1.6.6 Extensions and further research

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 in the text; a current research endeavour is the development of models of traffic management through link tolls in the situation where *demand is stochastic*, as well as applications to *combined* traffic models.

Acknowledgments

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. Laura Wynter and Jan T. Lundgren and the two referees for constructive remarks; the second author thanks Prof. R. Tyrrell Rockafellar for valuable, continuing discussions. Kinga Ulman supplied the numerical results of Section 1.5.3.

Notes

1. Such a heuristic tampering with the standard model is also discussed by Hearn (1980), in that case in order to produce equilibrium flows within a priori given upper and lower bounds. His conclusion, as ours, is that the remedy to the problem of non-anticipativity is to use traffic models that are general and accurate enough to reproduce the anticipated flows automatically.

2. A dynamic extension of the model could be of interest for the study of time-varying network conditions and corresponding management actions.

3. 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 set is larger than F because it contains cyclic flows, but due to the positivity assumption on t no equilibrium flow will utilize any cyclic flow, so this alternative representation is, in that sense, equivalent.

4. In the case where all functions g_k are affine, G is polyhedral, and thus satisfies these requirements automatically.

5. The relation (1.6a) is a restatement of (1.4d), while (1.6b) is a restriction of (1.4c); the latter is easily verified by observing that the relation (1.6b) is equivalent to f being in the face of G exposed by the vector ρ, that is, $f \in \arg\max_{y \in G} \rho^{\mathrm{T}} y$; this set clearly is a subset of G.

6. These facts are 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.

7. Although ρ does not appear explicitly in [TAP-VIP-SC-F] it is, through the equivalence with (1.6), present *implicitly* as a multiplier (or, normal) vector for the set G.

8. The description (1.8) of the set T used hinges on the fact that under primal and dual feasibility, complementarity and strong duality are equivalent; we choose to express strong duality in favour of complementarity because the latter would involve the non-unique route flow solution to [TAP-VIP-SC-F].

9. A natural requirement that will ensure that the algorithm proposed for the solution of [TOP] is well-defined is that *every vector* $\rho \in P$ *satisfies* $\sum_{a \in \mathcal{A}_C} [t_a(f^*) + \rho_a] \geq 0$ *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. A sufficient condition for this to hold is that $t(f^*) + N_G(f^*) \subseteq \Re_+^{|\mathcal{A}|}$, which, in turn, is implied by the assumption that $\frac{\partial g_k(f^*)}{\partial f_a} \geq 0$ for all $a \in \mathcal{A}$ and $k \in \mathcal{K}$.

10. 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, temporarily ignoring the route generation stage.

11. 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.

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

13. We do not discuss particular methods for the solution of $[\overline{\text{TOP}}]$ in this paper; the adequate method will be a function of the properties of the side constraints and the toll restrictions.

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

15. 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 Notes 12 and 14.

16. We ignore the case of $\inf_{\rho \in P \cap \overline{T}} \varphi(\rho)$ being finite but un-attained, since most forms of φ and P that seem natural (linear or quadratic, and polyhedral, respectively) preclude such a situation.

17. $\tilde{T}$ contains the line through ρ in the direction of $t(f^*) + \rho$, for any valid toll vector ρ.

18. The only way in which the situation can be dealt with (within the framework of additive cost structures) in the Wardrop equilibrium model is that *the toll schedule is not fixed*, but is a strictly monotone function of the commodity flows, that is, a *nonlinear price schedule* (Jennergren, 1972).

References

Bauschke, H. H. and Borwein, J. M. (1996). "On projection algorithms for solving convex feasibility problems," SIAM Review 38, 367–426.

Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (1993). *Nonlinear Programming: Theory and Algorithms*, 2nd ed. New York, NY: John Wiley & Sons.

Beckmann, M. J. and Golob, T. F. (1974). "Traveler decisions and traffic flows: a behavioral theory of network equilibrium," In: D.J. Buckley (ed.), *Transportation and Traffic Theory, Proceedings of the 6th International Symposium on Transportation and Traffic Theory*, Sydney, Australia, August 26–28, 1974 (pp. 453–482). New York, NY: Elsevier.

Bergendorff, P. (1995). "The bounded flow approach to congestion pricing," Master's thesis, Division of Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden.

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

Burton, D. and Toint, Ph. L. (1992). "On an instance of the inverse shortest paths problem," *Mathematical Programming*, 53, 45–61.

Cantarella, G. E. and Sforza, A. (1987). "Methods for equilibrium network traffic signal setting," In: A. R. Odoni, L. Bianco and G. Szegö (eds.), *Flow Control of Congested Networks*, Proceedings of the NATO Advanced Research Workshop on Flow Control of Congested Networks, Capri, Italy, October 12–18, 1986 (pp. 69–89). NATO ASI Series F: Computer and Systems Sciences, vol. 38. Berlin: Springer-Verlag.

Clarke, F. H. (1983). *Optimization and Nonsmooth Analysis.* New York, NY: John Wiley & Sons.

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.

Ferrari, P. (1997). "Capacity constraints in urban transport networks," *Transportation Research*, 31B, 291–301.

Gould, F. J. and Tolle, J. W. (1971). "A necessary and sufficient qualification for constrained optimization," SIAM Journal on Applied Mathematics, 20, 164–172.

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

Harker, P. T. (1988). "Multiple equilibrium behaviors on networks," *Transportation Science*, 22, 39–46.

Harker, P. T. (1991). "Generalized Nash games and quasi-variational inequalities," *European Journal of Operational Research*, 54, 81–94.

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

Hearn, D. W. (1980). "Bounding flows in traffic assignment models," Research Report 80-4, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL.

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).

Jennergren, L. P. (1972). "Decentralization on the basis of price schedules in linear decomposable resource-allocation problems," *Journal of Financial and Quantitative Analysis*, 7, 1407–1417.

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: F. Giannessi and A. Maugeri (eds.), *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* (pp. 169–178). New York, NY: Plenum Press.

Larsson, T. and Patriksson, M. (1997a). "Price-directive traffic management—An approach utilizing side constrained traffic equilibrium models," *Rendiconti del Circolo Matematico di Palermo, Serie II*, Suppl. 48, 147–170.

Larsson, T. and Patriksson, M. (1997b). "Side constrained traffic equilibrium models—analysis, computation and applications," Revised for *Transportation Research, B.*

Luo, Z.-Q. Pang, J.-S. and Ralph, D. (1996), *Mathematical Programs with Equilibrium Constraints.* Cambridge, UK: Cambridge University Press.

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

Miller, S. D., Payne, H. J., and Thompson, W. A. (1975). "An algorithm for traffic assignment on capacity constrained transportation networks with queues," Paper presented at the *Johns Hopkins Conference on Information Sciences and Systems*, The Johns Hopkins University, Baltimore, MD, April 2–4, 1975.

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

Nagurney, A. and Ramanujam, P. (1996). "Transportation network policy modeling with goal targets and generalized penalty functions," *Transportation Science*, 30, 3–13.

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

Payne, H. J., and Thompson, W. A. (1975). "Traffic assignment on transportation networks with capacity constraints and queueing," Paper presented at the *47th National ORSA Meeting/TIMS 1975 North-American Meeting*, Chicago, IL, April 30–May 2, 1975.

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

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

Rockafellar, R. T. (1981). *The Theory of Subgradients and its Applications to Problems of Optimization: Convex and Nonconvex Functions*. Berlin: Heldermann Verlag.

Rossi, T. F., McNeil, S., and Hendrickson, C. (1989). "Entropy model for consistent impact-fee assessment," *Journal of Urban Planning and Development*, 115, 51–63.

Small, K. A. (1992). "Using the revenues from congestion pricing," *Transportation*, 19, 359–381.

Smith, M. J. (1987). "Traffic control and traffic assignment in a signal-controlled network with queueing," In: N. Gartner and A. G. Wilson (eds.), *Proceedings of the Tenth International Symposium on Transportation and Traffic Theory*, Massachusetts Institute of Technology, Boston, MA, 1987 (pp. 319–338). New York, NY: Elsevier.

Smith, M. J. and Van Vuren, T. (1993). "Traffic equilibrium with responsive traffic control," *Transportation Science*, 27, 118–132.

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

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