Side Constrained Traffic Equilibrium Models— Analysis, Computation and Applications

Torbjörn Larsson
Division of Optimization
Department of Mathematics
Linköping Institute of Technology
S-581 83 Linköping
Sweden

and

Michael Patriksson
Department of Mathematics
Chalmers University of Technology
S-412 96 Göteborg
Sweden

December 8, 1997

Abstract

We consider the introduction of side constraints for refining a descriptive or prescriptive traffic equilibrium assignment model, and analyze a general such a model. Side constraints can be introduced for several diverse reasons; we consider three basic ones. First, they can be used to describe the effects of a traffic control policy. Second, they can be used to improve an existing traffic equilibrium model for a given application by introducing, through them, further information about the traffic flow situation at hand. As such, these two strategies complement the refinement strategy based on the use of non-separable, and typically asymmetric, travel cost functions. Third, they can be used to describe flow restrictions that a central authority wishes to impose upon the users of the network.

We study a general convexly side constrained traffic equilibrium assignment model, and establish several results pertaining to the above described areas of application. First, for the case of prescriptive side constraints that are associated with queueing effects, for example those describing signal controls, we establish a characterization of the solutions to the model through a Wardrop user equilibrium principle in terms of generalized travel costs and an equilibrium queueing delay result; in traffic networks with queueing the solutions may therefore be characterized as Wardrop equilibria in terms of well-defined and natural travel costs. Second, we show that the side constrained problem is equivalent to an equilibrium model with travel cost functions properly adjusted to take into account the information introduced through the side constraints. Third, we show that the introduction of side constraints can be used as a means to derive the link tolls that should be levied in order to achieve a set of traffic management goals.

The introduction of side constraints makes the problem computationally more demanding, but this drawback can to some extent be overcome through the use of dualization approaches, which we also briefly discuss.

Keywords: Traffic Assignment, User Equilibrium, Side Constraints, Generalized Wardrop Conditions, Distributed Queues, Queue Equilibrium, Queue Dynamics, Lagrangean Duality.

1 Introduction

The mathematical traffic planning tools known as *traffic assignment* models are used to describe, predict or prescribe a traffic flow pattern in a road network where there is some (fixed or elastic) travel demand and, usually, where because of traffic congestion the link travel times increase when the load of traffic becomes heavier. Depending of the characteristics of the real-world traffic situation modelled and the purpose of the model, they may also include a variety of other model components. (See, e.g., Sheffi, 1985; Nagurney, 1993; and Patriksson, 1994, for overviews of traffic assignment models.)

The traffic flow pattern is presupposed to comply with a *performance criterion*, which, typically, involves a measure of the disutility, for example, cost, of the total traffic flow in the urban area. (The cost of a travel is highly correlated to its duration in time, and we therefore use these two terms synonymously.) The performance criteria most commonly employed are the two optimality principles of Wardrop (1952). The first one is based on the assumption of rational traveller behaviour in the respect that each user of the congested traffic network seeks to minimize his/her own travel time, and it is therefore also known as the *user optimum*, or *equilibrium*, principle. Wardrop's second performance criterion is the minimization of the average travel time (or, equivalently, the total travel time), and it is therefore referred to as the *system optimum* principle.

1.1 Preliminaries

Let $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ denote a strongly connected transportation network, with $\mathcal{N}$ and $\mathcal{A}$ being the set 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 fixed positive travel demands d_{pq} which give rise to a link traffic flow pattern when distributed through the network. Further, for each link $a \in \mathcal{A}$ there is a positive and strictly increasing travel cost function $t_a : \Re_+^{|\mathcal{A}|} \mapsto \Re_{++}$.

The user equilibrium principle of Wardrop states that for each origin–destination (OD) pair $(p, q) \in \mathcal{C}$, the routes utilized have equal and minimal travel costs, so that no traveller can decrease his/her travel cost by shifting to another route in the OD pair. Denoting by $\mathcal{R}_{pq}$ the set of (simple) routes in OD pair (p, q), by h_{pqr} the flow on route $r \in \mathcal{R}_{pq}$, by h the vector of route flows, and by $c_{pqr} = c_{pqr}(h)$ the travel cost on a route, an equilibrium flow is defined by the conditions a

$$h_{pqr} > 0 \implies c_{pqr} = \pi_{pq}, \qquad \forall r \in \mathcal{R}_{pq}, \tag{1a}$$

$$h_{pqr} = 0 \implies c_{pqr} \geq \pi_{pq}, \qquad \forall r \in \mathcal{R}_{pq}, \tag{1b}$$

where π_{pq} is the equilibrium (that is, least) travel cost in OD pair (p, q). A user equilibrium state can be interpreted as a Nash equilibrium in a non-cooperative game among the OD pairs; in this game, each OD pair observes the traffic flows that result from the decisions made by the other OD pairs, and then distributes its travel demand among the OD routes in such a way that the routes utilized in the pair are among the least costly ones, at the costs obtained when the OD travel demand has been distributed.

The notion of an equilibrium should be thought of as a steady-state evolving after a transient (disequilibrium) phase in which the travellers successively adjust their route-choices, in order to minimize travel costs under the prevailing traffic conditions, until a situation with stable route travel costs and route flows has been reached. (See, e.g., Friesz *et al.*, 1994, on the issue of an adjustment process leading to a Wardrop equilibrium.)

In cases where the travel costs functions are *separable*, that is, where the travel costs can be expressed as $t_a(f) = t_a(f_a)$, the Wardrop conditions (1) are conform with the first-order optimality conditions for the convex network optimization problem (e.g., Beckmann *et al.*, 1956, and Dafermos, 1972)

[TAP]

$$\text{minimize } T(f) := \sum_{a \in \mathcal{A}} \int_0^{f_a} t_a(s)ds, \tag{2a}$$

subject to

$$\sum_{r \in \mathcal{R}_{pq}} h_{pqr} = d_{pq}, \qquad \forall (p,q) \in \mathcal{C}, \tag{2b}$$

$$h_{pqr} \geq 0, \qquad \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C}, \tag{2c}$$

$$\sum_{(p,q) \in \mathcal{C}} \sum_{r \in \mathcal{R}_{pq}} \delta_{pqra} h_{pqr} = f_a, \qquad \forall a \in \mathcal{A}, \tag{2d}$$

where

$$\delta_{pqra} := \begin{cases} 1, & \text{if route } r \in \mathcal{R}_{pq} \text{ uses link } a, \\ 0, & \text{otherwise,} \end{cases} \qquad \forall a \in \mathcal{A},\ \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C},$$

is the link-route incidence matrix, and f_a denotes the total flow on link a. Below, we will also use the notations $H := \{ h \in \Re^{|\mathcal{R}|} \mid h \text{ satisfies (2b)–(2c)} \}$ and $F := \{ f \in \Re^{|\mathcal{A}|} \mid f \text{ satisfies (2b)–(2d)} \}$. The reader should note that the Wardrop equilibrium principle is intimately associated with the inherent Cartesian product structure of the feasible set of [TAP], that is, the independence of the flows in the different OD pairs, a fact that is exposed if the definitional constraints (2d) are used to eliminate the link flow variables from the problem.

The equilibrium assignment model [TAP] is well-known and frequently applied in transportation analysis; its popularity is to a large extent explained by its simplicity and nice interpretations, which makes it easy to access for practitioners. Several methods have been developed for its effective and efficient solution. (See, e.g., Patriksson, 1994, for a thorough review of solution methods.) Worth noting is that all efficient algorithms for [TAP] exploit its Cartesian product structure (e.g., Larsson and Patriksson, 1992).

The *validity* of the model [TAP] (i.e., its ability to describe real-world traffic flows accurately enough with respect to the model's purpose) rests on the following presumptions. First, since the Wardrop conditions postulate a steady-state situation, also the model [TAP] is valid for this situation only. Second, it requires the knowledge about sufficiently accurate estimates of the data of the model's components (e.g., functional form and parameters of the link travel cost functions), which, because of the steady-state assumption, must also be presumed to be stable.

Any model of a real-world traffic equilibrium problem is, of course, approximate, since the data used in the model are estimated quantities. Further, it is normally the average values of naturally varying parameters that are estimated, and, furthermore, the values of these parameters may change in an unpredictable manner because of unforeseen changes in the traffic system described by the model (e.g., accidents, the weather conditions, or the proportion of different types of vehicles in the traffic flow). Hence, both the knowledge about the data of the model and their stability may be questionable. To ensure the validity of a model it is therefore natural to restrict its use to some traffic situations only, or to let some of its data depend of the situation under consideration. (An example of this is the introduction of time-slices to capture variations in the real-world traffic system, especially the variations in travel demands and travel time characteristics.)

1.2 Motivation

The invalidity of a traffic equilibrium model of the form [TAP] may also be due to its structural limitations, that is, its inherent simplicity which makes it inapplicable to more complex traffic problems (e.g., Sender and Netter, 1970); examples of such limitations are that the travel demand is presumed to be independent of the travel times and that there is no discrimination between different types of vehicles.

An illustrative example of a deficiency of the model and its possible consequences is provided by Hearn (1980), who comments on its property of allowing every road to carry arbitrarily large volumes of traffic. He states that this property causes that "the predicted flow on some links will be far lower or far greater than the traffic engineer knows they should be *if all assumptions of the model are correct*." Further, as a result of this, "the model predictions are ignored, or, more often, the user will perturb the components of the model (trip table, volume delay formulas, etc.) in an attempt to bring the model output more in line with the anticipated results."

In order to avoid such heuristic tampering with components of the model available, traffic planners must be supplied with analysis tools whose underlying traffic models are sufficiently general, reliable and accurate; much research has therefore been devoted to the task of refining the basic traffic equilibrium assignment model [TAP] in various respects. An overwhelming portion of this research has dealt with extensions of the equilibrium model's travel cost functions.

Flow relationships such as interactions between the flows on intersecting links or between vehicles of different types are captured by introducing non-separable, and typically also asymmetric, travel cost functions (e.g., Akcelik, 1988, and Toint and Wynter, 1996, respectively). However, due to the non-integrability of such functions, the problem of finding a solution to the Wardrop conditions can then not be formulated as an optimization model of the form [TAP]. Instead, these conditions are in this more general case stated as the finite-dimensional *variational inequality* problem of finding an $f^* \in F$ such that

[VIP]

$$c(f^*)^{\mathrm{T}}(f - f^*) \geq 0, \qquad \forall f \in F,$$

where $c : \Re_+^{|\mathcal{A}|} \mapsto \Re_{++}^{|\mathcal{A}|}$ is the travel cost mapping. Of course, since the variational inequality problem is merely a restatement of Wardrop's user equilibrium principle, the basic assignment model [TAP] and this class of more general assignment models both comply with the same fundamental behavioural principle. Finite-dimensional variational inequality models arise also in several other areas of operations research and applied mathematics, and have been extensively studied, mainly from a theoretical and algorithmical point of view; see Harker and Pang (1990) for a recent survey of general theory, algorithms, and areas of applications of this class of mathematical models, and, for example, Nagurney (1993) and Patriksson (1994) for overviews of this field in the context of traffic assignment.

While improving the basic model's ability to accurately describe, reproduce, and predict a real-world traffic situation, the utilization of non-separable, and possibly asymmetric, travel time functions is, however, not a natural and adequate means for handling supplementary traffic flow restrictions such as for example those imposed by a traffic control policy. (See Yang and Yagar, 1994, for an example of a traffic control policy which gives rise to link flow capacity constraints. Ferrari, 1995, presents an example in which a capacity constraint incorporates flow from two conflicting traffic streams through a junction.)

The natural—but so far little studied—approach for describing and capturing such traffic flow restrictions is to introduce *side constraints*. Under the presumption that the traffic flow restriction to be modelled have well-defined physical meanings, the resulting side constraints will possess immediate interpretations, and it may thus be relatively easy for the traffic engineer to identify a suitable set of side constraints (that is, their functional forms and the proper values of their parameters), as compared to the task of making proper estimates of the values of the parameters in travel cost functions. (For example, in the situation described by Hearn, 1980, a proper refinement of an assignment model may simply be the introduction of the link capacity constraints corresponding to the engineer's anticipation of reasonable levels of traffic flow.)

The introduction of asymmetric travel cost functions and side constraints constitutes two principally different strategies for refining or extending the basic traffic equilibrium model, and they can be used separately or in conjunction.

1.3 Background

Network optimization models with side constraints arise naturally and frequently in many areas of applications, such as, for example, transportation and distribution management, task assignment and scheduling, and project planning. Further, non-network optimization models may occasionally be reformulated into side constrained network models (e.g., through the introduction of auxiliary variables) or be identified as being side constrained network models, possibly also with extra, non-network, variables (in this context, one sometimes use the term *embedded* networks). Furthermore, many optimization models can be regarded to be network models to which a set of complicating side constraints has been added (e.g., the travelling salesman problem). In general, side constraints in network models describe limitations on the availability of scarce resources (e.g., transportation or production capacities, storage limitations, investment capital available, etc.) which are shared by several activities (link flows), or restrictions of a logical or technological nature (e.g., consistency between successive time periods, or conversions between different types of flow commodities in a production process). Selected references on side constrained network models are Weigel and Cremeans (1972), Shepardson and Marsten (1980), and Choi *et al.* (1988); see also Section 16.5 of Ahuja *et al.* (1993).

Worth noting is that side constraints in network models are sometimes *hard,* that is they need to be fulfilled exactly (e.g., definitional and technological constraints), and they are sometimes *soft,* that is they may in practice be slightly violated (e.g., some types of resource constraints); the distinction between hard and soft side constraints is relevant also in our application.

Although the use of side constraints seems to be a quite natural and practical means for refining a traffic assignment model, this approach has received very little attention. A main reason is that, as a result of the addition of the side constraints, the Cartesian product structure of the feasible set of the basic model is lost, thus obtaining a computationally more demanding model. Further, the solutions to the resulting models can no longer be given characterizations as Wardrop equilibria in the classical sense.

However, the special case of link capacity side constrained traffic assignment models has been studied rather thoroughly. Link capacities have been introduced as a means for modelling congestion effects (see Charnes and Cooper, 1961, and Jorgensen, 1963, for early examples of this) and then represent the so-called *saturation* link flows. When a link is saturated, congestion effects result in queueing and any excess flow will accumulate in the queue; in an equilibrium state, the saturated links may therefore carry stationary queues (e.g., Smith, 1987). Link capacities also arise naturally when links are signal-controlled (e.g., ibid. and Yang and Yagar, 1994).

For the link capacitated model, it is known that solutions can be characterized as Wardrop equilibria in terms of well-defined generalized route (or, link) travel costs (Jorgensen, 1963, Hearn, 1980, and Inouye, 1987); a similar characterization has recently been made for the *route* capacitated model, see Maugeri (1994). Further, the Lagrange multipliers for the capacity constraints can be given interesting interpretations. First, they are the link tolls that the travellers are willing to pay for being allowed to use the links (Jorgensen, 1963), and, second, they may be interpreted as the delays in steady-state link queues (Payne and Thompson, 1975, and Miller *et al.*, 1975); the link queueing interpretation also provides a *queue equilibrium* characterization of solutions to the model (ibid.). In Larsson and Patriksson (1995), we review these theoretical results for the capacitated model and show that it can be efficiently dealt with computationally. We will in this work generalize the theoretical findings for the capacitated model to the general side constrained model.

1.4 Other uses of side constraints

There may of course be other reasons for considering the use of side constraints in assignment models, besides the one discussed above, and we here give three other examples. First, side constraints may be used as a practical means for improving the quality of an available (but maybe not well calibrated) assignment model by directly incorporating into the model additional information about the actual real-world traffic situation. The side constraints may then for example describe the requirement that observed flows on some links or observed travel times in some origin–destination pairs should be reproduced (at least approximately) in the calculated solution. Second, side constraints arise quite naturally in time-sliced traffic assignment as a means to capture dynamic effects, that is, to describe the coupling between the assignment problems in successive time-slices. (Whether these constraints are to be considered as side constraints depends on the modelling and solution strategies employed; one may describe such a problem such that the constraints correspond to node balancing constraints in the definition of an expanded, multi-copy, network, whence they would not be considered as side constraints, or they may be considered as coupling side constraints in a system which is otherwise decoupled over the time periods. This freedom-of-choice in viewing a set of side constraints may of course also be present in other applications.) Third, as shall be shown, side constraints may be used for calculating proper link tolls to be imposed on the travellers in order to limit some volumes of traffic to levels that are acceptable. In this case, a side constraint may for example model the maximal volume of traffic which can be allowed to enter the central area of a city; the solution of the side constrained model will then provide the toll which need to be introduced in order to reach this goal without imposing any centralized traffic control.

The examples of different types of side constraints discussed so far show that they may differ significantly with respect to their purposes and properties. However, we distinguish two principally different types of side constraints in traffic assignment models: *prescriptive* (hard) and *descriptive* (soft). (Cf. the distinction between physical and environmental capacity constraints made by Ferrari, 1995.)

The prescriptive side constraints, which are imposed upon the users of the traffic system and typically arise from traffic management and control policies (e.g., speed limit regulations and traffic signals with fixed cycle times), are known exactly and can (or, may) never be violated. A simple special case of prescriptive constraints are link capacities which are used to model saturation flows in a queueing network. Among the prescriptive side constraints we also include those constructed with the purpose of calculating tolls to be imposed on the travellers in order to reach some traffic management goal (cf. the discussion above).

Traditionally, side constraints are used in *decision* models, and are as such *prescriptive* (cf. the discussion made in Section 1.3). In contrast, transportation planning models, such as those of traffic equilibrium, are founded on *behavioural* principles. The validity of using *prescriptive* side constraints in traffic models therefore rests on the assumption that their effects are transferable to the perception of travel costs among the trip-makers, for example as queueing delays or link tolls. As shall be shown, prescriptive constraints may be binding at a steady-state flow and cause stationary link queues to appear (cf. Section 3).

Descriptive side constraints may be introduced as a means to refine the model by including additional (maybe approximate) traffic flow restrictions into the model (e.g., joint capacities in roundabouts), as a (rough) means for modelling congestion effects, or in order to incorporate into the model some a priori knowledge of the equilibrium traffic flow (e.g., observed flows on some links). An interesting usage of descriptive side constraints is in the derivation of an adjusted, tentative travel cost function which more correctly reflects the travel cost perception and/or provides a more reasonable output from an equilibrium model (cf. the discussion made by Hearn, 1980); see Theorem 2.2. Clearly, because of the nature of the descriptive constraints they do not need to become satisfied exactly, and this fact may also be exploited in solution procedures (cf. Section 5).

As in all applications of mathematical modelling of real-life situations, it is necessary to bear in mind that some caution is of course needed also when side constraints are utilized. For example, side constraints constructed for some certain purpose and for some given conditions in a traffic system may very well be invalid if the prevailing presumptions, such as the traffic conditions (e.g., the proportions of different classes of vehicles) change; however, this is also the case for other model components (e.g., the travel cost functions).

1.5 Scope

Our main goal is to make a thorough theoretical analysis of the side constrained model, in order to reach insight into the properties of this modelling methodology and provide a solid basis for its exploitation, rather than to contribute to the practical aspects of the use of side constraints in traffic models (even though we briefly discuss some different usages). We focus on equilibrium characterizations of solutions to side constrained models, and on the relationship between the introduction of side constraints and changes in the travel cost functions. Further, we suggest a very natural solution strategy for side constrained models and discuss various interpretations of the theory developed and the solution strategy suggested.

We consider a general convexly side constrained extension of the basic equilibrium assignment model [TAP] and investigate its optimality conditions, which may be interpreted as a generalization of Wardrop's equilibrium principle (1) in the sense that an equilibrium holds in terms of generalized travel costs. The queue equilibrium result of Miller *et al.* (1975) for link capacitated problems is extended to the case of general convex side constraints, and we thereby obtain a characterization of solutions to the side constrained traffic assignment problem as Wardrop equilibria in terms of travel costs and link queueing delays, that is, in terms of the natural costs to be minimized by the individual travellers in a network with queueing; this result asserts that solutions to side constrained traffic equilibrium assignment models comply with the basic assumption of rational traveller behaviour. It is also shown that if the values of the Lagrange multipliers for the side constraints are at hand, then one may alternatively and equivalently solve an ordinary traffic equilibrium assignment problem with well-defined adjusted travel cost functions. Any convergent algorithm for finding these multiplier values may be viewed as a systematic way of calibrating the proper travel cost functions, as opposed to the heuristic tampering described in the example of Hearn (1980).

All the results presented hold also when the travel cost functions are non-separable (under a strict monotonicity assumption) and, further, most of them can be extended to the model [VIP] subject to side constraints. (Some of those results are given in Larsson and Patriksson, 1994.) In order to monitor the fundamental results in a plain context, we have chosen to here study the side constrained extension of the basic model [TAP]. We briefly mention more complex models that have natural side constrained extensions in Section 6.

2 A side constrained assignment model

For the sake of simplicity of the presentation, we presume that the side constraints involve the link flows only. This presumption is also quite natural since it is usually only the link flows that can be observed and, for example, affected through control actions. Further, only inequality side constraints are considered. However, neither of these two restrictions cause any serious loss of generality, and all the results that are derived can easily be generalized to the cases of affine equality side constraints and side constraints that involve also commodity link or route flows, however only under additional assumptions on the traffic model (cf. the discussion made in Section 6).

We thus suppose that the supplementary traffic flow restrictions introduced in order to extend or improve the equilibrium assignment model [TAP] are described by the side constraints

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

where the functions $g_k : \Re_+^{|\mathcal{A}|} \mapsto \Re$, $k \in \mathcal{K}$, are *convex and continuously differentiable.* The index set $\mathcal{K}$ may include, for example, subsets of the indices of the sets of links, nodes, or OD pairs. The vector with components $g_k(\cdot)$, $k \in \mathcal{K}$, is denoted $g(\cdot)$.

The side constrained traffic equilibrium problem is then given by

[TAP-SC]

$$\text{minimize } T(f) := \sum_{a \in \mathcal{A}} \int_0^{f_a} t_a(s)ds, \tag{3a}$$

subject to

$$\sum_{r \in \mathcal{R}_{pq}} h_{pqr} = d_{pq}, \qquad \forall (p,q) \in \mathcal{C}, \tag{3b}$$

$$h_{pqr} \geq 0, \qquad \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C}, \tag{3c}$$

$$\sum_{(p,q) \in \mathcal{C}} \sum_{r \in \mathcal{R}_{pq}} \delta_{pqra} h_{pqr} = f_a, \qquad \forall a \in \mathcal{A}, \tag{3d}$$

$$g_k(f) \leq 0, \qquad \forall k \in \mathcal{K}. \tag{3e}$$

We presume that the feasible set of [TAP-SC] is *non-empty.* In case some function g_k is nonlinear, we also presume that a *constraint qualification* (e.g., Bazaraa *et al.*, 1993, Chapter 5) holds. The convexity of the problem then ensures that the optimal solutions are characterized by the first-order optimality conditions for [TAP-SC]; the optimal link flow solution and the set of optimal route flow solutions is denoted f^* and H^*, respectively. (The link flow solution is uniquely determined, even though there are, in general, alternative optimal route flow solutions.) The *Lagrange multipliers* associated with the constraints (3b) and (3e), respectively, are denoted by $\pi \in \Re^{|\mathcal{C}|}$ and $\beta \in \Re^{|\mathcal{K}|}$.

2.1 Generalized Wardrop equilibrium

In our analysis of [TAP-SC], we begin by showing that its solutions are Wardrop equilibrium flows in terms of well-defined *generalized route travel costs.*

Definition 2.1 (Generalized route travel cost) *Let (h^*, f^*) be a solution to the problem* [TAP-SC] *and let β^* be a vector of Lagrange multipliers for the side constraints (3e). Then, the generalized route travel costs are defined as*

$$\overline{c}_{pqr} := c_{pqr}(h^*) + \sum_{k \in \mathcal{K}} \beta_k^* \left(\sum_{a \in \mathcal{A}} \delta_{pqra} \frac{\partial g_k(f^*)}{\partial f_a} \right), \qquad \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C}. \tag{4}$$

We next state our first main result.

Theorem 2.1 (Solutions to [TAP-SC] are generalized Wardrop equilibria) *Suppose that* (h^*, f^*) *solves the problem* [TAP-SC] *and that the vectors* π^* *and* β^* *are Lagrange multipliers for the constraints (3b) and (3e), respectively. Let generalized route travel costs be given by the expression (4). Then,*

$$h^*_{pqr} > 0 \implies \bar{c}_{pqr} = \pi^*_{pq}, \qquad \forall r \in \mathcal{R}_{pq}, \tag{5a}$$
$$h^*_{pqr} = 0 \implies \bar{c}_{pqr} \geq \pi^*_{pq}, \qquad \forall r \in \mathcal{R}_{pq}, \tag{5b}$$

holds for all OD pairs $(p, q) \in \mathcal{C}$.

Proof. Since a constraint qualification is assumed to hold for [TAP-SC], the first-order necessary optimality conditions

$$h_{pqr}\left(\bar{c}_{pqr} - \pi_{pq}\right) = 0, \qquad \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C}, \tag{6a}$$
$$\bar{c}_{pqr} - \pi_{pq} \geq 0, \qquad \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C}, \tag{6b}$$
$$\sum_{r \in \mathcal{R}_{pq}} h_{pqr} = d_{pq}, \qquad \forall (p,q) \in \mathcal{C}, \tag{6c}$$
$$h_{pqr} \geq 0, \qquad \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C}, \tag{6d}$$
$$\sum_{(p,q) \in \mathcal{C}} \sum_{r \in \mathcal{R}_{pq}} \delta_{pqra} h_{pqr} = f_a, \qquad \forall a \in \mathcal{A}, \tag{6e}$$
$$\beta_k g_k(f) = 0, \qquad \forall k \in \mathcal{K}, \tag{6f}$$
$$g_k(f) \leq 0, \qquad \forall k \in \mathcal{K}, \tag{6g}$$
$$\beta_k \geq 0, \qquad \forall k \in \mathcal{K} \tag{6h}$$

are satisfied by the solution (h^*, f^*) and the multipliers π^* and β^*. Then, for any OD pair $(p, q) \in \mathcal{C}$, the condition (6b), together with (6a) and (6c), implies that the multiplier value π^*_{pq} is the minimal generalized route travel cost $\bar{c}_{pqr}$, and, further, condition (6a) implies that these costs are equal for all routes utilized in the OD pair. Hence, the optimality conditions (6a)–(6b) imply the generalized Wardrop conditions (5), and the theorem is proved. □

The converse conclusion is invalid since the complementarity conditions for the side constraints are not necessarily satisfied whenever the Wardrop-type conditions of the theorem hold; however, a partial converse (i.e., the conclusion that the converse result is true under an additional hypothesis) will be established in the next section.

The interpretation of the solution to (5) as a Wardrop equilibrium rests, as noted previously, on the presumption that the effects of the side constraints may be transferred to the trip-makers' perception of the travel costs, for example as delays in queues.

The Lagrange multiplier values π^* and β^* are the shadow prices for the travel demand constraints (3b) and the side constraints (3e), respectively, that is, the sensitivities of the objective of [TAP-SC] with respect to the right-hand sides of these constraints. [As stated in Theorem 2.1, the multiplier value π^*_{pq} provides the minimal generalized equilibrium route travel cost in OD pair $(p, q) \in \mathcal{C}$; this relation is the reason for using the same notation as in the Wardrop conditions (1).] In some applications, for example, in traffic management through link tolls, it may actually be the multiplier values β^* that are primarily sought for, rather than the equilibrium link flows (see end of section). In Section 3 we consider a queueing network and establish a close relationship between the values of the Lagrange multipliers β and the link queueing delays.

Expressing the route travel costs as

$$c_{pqr} := \sum_{a \in \mathcal{A}} \delta_{pqra} t_a(f_a), \qquad \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C},$$

we obtain from (4) that the generalized route travel costs may be stated as

$$\bar{c}_{pqr} := \sum_{a \in \mathcal{A}} \delta_{pqra} \left(t_a(f_a^*) + \sum_{k \in \mathcal{K}} \beta_k^* \frac{\partial g_k(f^*)}{\partial f_a} \right), \qquad \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C},$$

and we arrive at the following natural definition.

Definition 2.2 (Generalized link travel costs) *Let f^* be the link flow solution to the problem* [TAP-SC] *and let β^* be a vector of Lagrange multipliers for the side constraints (3e). Then, the generalized link travel costs are defined as*

$$\bar{t}_a(f^*) := t_a(f_a^*) + \sum_{k \in \mathcal{K}} \beta_k^* \frac{\partial g_k(f^*)}{\partial f_a}, \qquad \forall a \in \mathcal{A}. \tag{7}$$

The reader should note that without further assumptions on the properties of the side constraints, neither the values of their multipliers nor, as a consequence, the generalized equilibrium link and route travel costs (7) and (4), respectively, are necessarily uniquely determined. (The statement made on page 439, rows 11–13 in Larsson and Patriksson, 1995, was poorly formulated, since it may lead the reader to believe that the values of the multipliers π_{pq} are unique in the link capacitated model.) Larsson and Patriksson (1998b) provide a characterization [based on the conditions (6)] of the set of multipliers β as a polyhedral set, and a simple example with link capacity side constraints showing that the set is not a singleton in general, and in fact is likely to be unbounded (at least in the link capacitated case). (A solution method for [TAP-SC] will in general produce one such vector; see Section 5.)

This non-uniqueness property has immediate consequences for the applications of side constraints that we consider in this paper. First, it implies that in the case where prescriptive side constraints are used to model some traffic control policy, the resulting link queueing delays are not unique; second, in the case where side constraints are used to model some traffic management goals, there is a freedom-of-choice in the link tolls that should be introduced in order to reach these goals.

Introducing the *Lagrangean function* with respect to the side constraints (3e),

$$L(f, \beta) := T(f) + \beta^{\mathrm{T}} g(f),$$

that is, the objective which is obtained if these constraints are taken into account only implicitly through a Lagrangean penalization (or, dualization) with multipliers β, it is seen that the vector of generalized equilibrium link travel costs may be expressed as

$$\bar{t}(f^*) := \nabla_f L(f^*, \beta^*) = t(f^*) + \nabla g(f^*)^{\mathrm{T}} \beta^*. \tag{8}$$

The generalized equilibrium link and route travel costs (7) and (4), respectively, are thus composed by actual costs and penalty costs, where the latter include in a Lagrangean fashion the impact of the side constraints on the equilibrium problem. This composition of the generalized equilibrium travel costs is a consequence of the equivalence between [TAP-SC] and the equilibrium assignment problem obtained when dualizing the side constraints with penalties that are Lagrange multipliers; this is our second main result.

Theorem 2.2 (An equivalent equilibrium assignment problem) *Let β^* be a vector of Lagrange multipliers for the side constraints (3e). Then, the equilibrium assignment model with the (symmetric) link travel cost mapping*

$$\bar{t}(\cdot) := t(\cdot) + \nabla g(\cdot)^{\mathrm{T}} \beta^*, \tag{9}$$

has the same solution set as [TAP-SC].

Proof. The strict convexity of T and the discussion following Theorem 6.5.1 of Bazaraa *et al.* (1993) yield that f^* is the unique link flow solution to the dualized problem

$$\underset{f \in F}{\text{minimize}} \ L(f, \beta^*) := T(f) + (\beta^*)^{\mathrm{T}} g(f).$$

Since the link travel cost mapping of the dualized problem is $\nabla L(\cdot, \beta^*) := t(\cdot) + \nabla g(\cdot)^{\mathrm{T}} \beta^*$, the result follows. □

Hence, if a vector of Lagrange multipliers for the side constraints were somehow known, then [TAP-SC] could be solved as a standard equilibrium assignment problem. (A corresponding result can be shown to hold in the case of a non-integrable link travel cost mapping, see Larsson and Patriksson, 1994, Theorem 4.1.) Note that this equivalence result implies that the link travel cost mapping (9) is a precise description of the influence of prescriptive traffic flow restrictions on the travel cost perception of the users of the traffic network, and therefore on their route-choice behaviour. (This conclusion highlights the fact that the model [TAP-SC] also rests on the steady-state assumption which underlies the Wardrop conditions.)

The result of Theorem 2.2 can be used to construct travel cost functions that bring the equilibrium solution more into par with a traffic engineer's anticipation of reasonable levels of flow (cf. Hearn, 1980); further, it facilitates the construction of improved travel cost functions, based on (1) some tentative cost function which however does not agree with the trip-makers' travel cost perception (since, for example, the calculated travel times or link flows do not correspond to measured ones), and (2) some descriptive side constraints that incorporate additional knowledge about the traffic conditions. This strategy seems appealing since it should be easy for the engineer to formulate side constraints that will adjust the calculated traffic flow towards a more reasonable one, as compared to the task of predicting how a heuristic adjustment of the tentative travel time functions will affect the equilibrium flow pattern.

The reader may note that the equivalence result of Theorem 2.2 holds also if the link travel cost mapping (9) is simplified into the mapping $t(\cdot) + \nabla g(f^*)^{\mathrm{T}} \beta^*$, that is, if the penalty costs which shall enforce the fulfilment of the side constraints are chosen to be fixed instead of flow-dependent. (If, for example, the side constraints (3e) describe traffic management goals, then $\nabla g(f^*)^{\mathrm{T}} \beta^*$ is, in fact, the vector of fixed link tolls that should be imposed upon the travellers in order to reach these goals. More on this application is found in Larsson and Patriksson, 1997, 1998a, 1998b, and in Section 5.3.)

2.2 Stability results

We will next analyze the side constrained equilibrium assignment model with respect to a (dual) stability property. This property is closely related to the steady-state assumption, that is, that the model describes a steady-state situation which is reached after a transient phase in which the travellers successively adjust their route-choices. We define a solution to [TAP-SC] as being *stable* if the set of shortest routes (with respect to the generalized travel costs) is unique (compare with the similar regularity condition of Nagurney and Zhang, 1995, p. 213, for the user equilibrium case), that is, if it is independent of the choice of values of the Lagrange multipliers β among those which satisfy the conditions (6). The importance of this property lies in the fact that it ensures that the collection of shortest routes in a steady-state is independent of the initial conditions and history of the disequilibrium phase. (If the solution is not stable, then different initial conditions and flow histories can lead to steady-states corresponding to different values of the multipliers and different sets of shortest routes.)

We first make the obvious observation that stability of a solution to [TAP-SC] is ensured if the generalized equilibrium route travel costs (4) are unique, which, in turn, holds if the generalized link travel costs (7) are unique, or if the values of the multipliers of the side constraints are uniquely determined. (Note that the uniqueness of the generalized link travel costs does not, in general, imply the uniqueness of the values of the Lagrange multipliers.) The analysis of the stability property thus leads to the study of uniqueness properties of the generalized travel costs and the multipliers of the side constraints. (These uniqueness properties are of interest for other reasons as well. For example, if the side constraints are link capacities which arise from traffic signals at the links' exits, then unique values of the multipliers correspond to unique and stable link queueing delays.)

We first state a result which is immediate from the generalized Wardrop conditions (5); it provides however a nice intuitive basis for the next result to be presented.

Theorem 2.3 (Always used routes are shortest for all multiplier values) *Suppose that a route carries a positive flow in every route flow solution to* [TAP-SC]. *Then, it is a shortest route with respect to the generalized route costs for any vector of Lagrange multipliers for the side constraints (3e).*

Note that this result does *not* imply that the shortest route costs, that is, the values of the multipliers π_{pq}, $(p,q) \in \mathcal{C}$, are independent of the values of the multipliers for the side constraints.

We shall next establish the uniqueness of the set of shortest routes at a solution to [TAP-SC]. Here, for a vector β of Lagrange multipliers for the side constraints (3e), we let

$$\overline{c}_{pqr}(\beta) := \sum_{a \in \mathcal{A}} \delta_{pqra} \left(t_a(f_a^*) + \sum_{k \in \mathcal{K}} \beta_k \frac{\partial g_k(f^*)}{\partial f_a} \right), \qquad \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C}.$$

Theorem 2.4 (Uniqueness of the sets of generalized equilibrium shortest routes) *Consider the following statements.*

(a) *For any $(p,q) \in \mathcal{C}$ and any $r \in \mathcal{R}_{pq}$, either $h_{pqr} > 0$ or $h_{pqr} = 0$ holds for all $h \in H^*$.*

(b) *For any route flow solution $h \in H^*$ and any vectors π and β of Lagrange multipliers for the constraints (3b) and (3e), respectively, strict complementarity holds in the Wardrop conditions (5), that is*

$$h_{pqr} > 0 \implies \overline{c}_{pqr}(\beta) = \pi_{pq}, \qquad \forall r \in \mathcal{R}_{pq}, \tag{10a}$$
$$h_{pqr} = 0 \implies \overline{c}_{pqr}(\beta) > \pi_{pq}, \qquad \forall r \in \mathcal{R}_{pq}, \tag{10b}$$

holds for all OD pairs $(p,q) \in \mathcal{C}$.

(c) *For any $(p,q) \in \mathcal{C}$ and any $r \in \mathcal{R}_{pq}$, either*

$$\overline{c}_{pqr}(\beta) > \min_{s \in \mathcal{R}_{pq}} \overline{c}_{pqs}(\beta)$$

or

$$\overline{c}_{pqr}(\beta) = \min_{s \in \mathcal{R}_{pq}} \overline{c}_{pqs}(\beta)$$

holds for all vectors β of Lagrange multipliers for the side constraints (3e).

(d) *The generalized equilibrium shortest routes are the same for every $h \in H^*$ and every vector β of Lagrange multipliers for the side constraints (3e).*

Then, the following relations hold.

$$\textbf{(a)} \Longleftarrow \textbf{(b)} \Longleftrightarrow \textbf{(c)} \Longleftrightarrow \textbf{(d)} \tag{11}$$

Proof.

(b) $\Longrightarrow$ **(a)**. Let h^1 and h^2 be in H^*. Take an arbitrary OD pair $(p,q) \in \mathcal{C}$ and a route $r \in \mathcal{R}_{pq}$. Suppose that $h^1_{pqr} > 0$ while $h^2_{pqr} = 0$. Then (10) yields a contradiction, since shortest route costs $\overline{c}_{pqr}(\beta)$ are independent of the route flow solution chosen.

(b) $\Longleftrightarrow$ **(c)**. The conditions (10) are simply the Wardrop conditions (5) plus **(c)**.

(c) $\Longleftrightarrow$ **(d)**. The condition **(d)** is simply a restatement of the condition **(c)**. □

Since some routes are likely to consist of single links, and, moreover, the uniqueness of the generalized equilibrium link travel costs (7) implies unique generalized equilibrium route costs, we shall next consider the uniqueness of the link costs. We first give a sufficient condition for the uniqueness of the link travel costs (7) to be equivalent to the uniqueness of the values of the multipliers β. Of interest here are the side constraints that are *active* at the unique link flow solution to [TAP-SC], that is, the set defined as

$$\mathcal{K}(f^*) := \{\, k \in \mathcal{K} \mid g_k(f^*) = 0 \,\}.$$

The proof of the result is elementary and therefore omitted.

Theorem 2.5 (Simultaneous uniqueness of $\overline{t}(f^*)$ and β^*) *If the vectors*

$$\nabla g_k(f^*), \qquad k \in \mathcal{K}(f^*),$$

are linearly independent, then the generalized equilibrium link travel costs (7) are uniquely determined if and only if the Lagrange multipliers β have unique values.

This result may, for example, be invoked for any link capacity side constrained traffic equilibrium assignment problems, since the linear independence assumption always holds for such a problem. (The reader should note that the result does *not* imply that the multipliers for capacity constraints are unique, since both the generalized equilibrium travel costs and the Lagrange multiplier vector may be non-unique; indeed, Larsson and Patriksson, 1998b, contains a counter-example to this occasionally stated claim.)

We finally give a sufficient condition for the Lagrange multipliers β to have unique values, in which case the generalized equilibrium link and route travel costs will also be uniquely determined. Here, given a route flow solution $h^* \in H^*$, Γ_+ denotes the route–OD pair incidence matrix for the routes with a positive flow in h^* (the vector of which is denoted by h_+). The network thus constructed is, by the positivity of the demand vector d, strongly connected, and therefore the matrix Γ_+ has full row rank. A consequence of this is that the orthogonal projection onto the null space of Γ_+ is well-defined. (The null space of Γ_+ is the set of demand-feasible route flow adjustments from h^* using the routes in h_+ only.)

Theorem 2.6 (Uniqueness of generalized equilibrium travel costs) *Suppose that f^* solves* [TAP-SC]*, and that for some route flow solution $h^* \in H^*$ the orthogonal projections of the vectors*

$$\nabla_{h_+} g_k(f^*), \qquad k \in \mathcal{K}(f^*)$$

onto the null space of Γ_+ are linearly independent. Then, the values of the multipliers β for the side constraints are unique and the generalized equilibrium link and route travel costs (4) and (7), respectively, are uniquely determined.

Proof. Let $c_+(h^*)$ be the vector of travel costs $c_{pqr}(h^*)$ for the routes with positive flow in h^*, and let β be a vector of multipliers for the side constraints.

By the generalized Wardrop condition (5a) there is a vector $\pi^* \in \Re^{|\mathcal{C}|}$ such that

$$c_+(h^*) + \nabla_{h_+} g(f^*)^{\mathrm{T}} \beta = \Gamma_+^{\mathrm{T}} \pi^*.$$

Define the projection matrix

$$P := I - \Gamma_+^{\mathrm{T}}(\Gamma_+\Gamma_+^{\mathrm{T}})^{-1}\Gamma_+,$$

where I is the identity matrix of order equal to the number of routes with positive flow in the solution h^*. (The inverse of the matrix $\Gamma_+\Gamma_+^{\mathrm{T}}$ exists since Γ_+ has full row rank.)

Multiplying the above system of equations with the projection matrix yields

$$P\nabla_{h_+} g(f^*)^{\mathrm{T}}\beta = -Pc_+(h^*).$$

Using that, according to the condition (6f),

$$\beta_k = 0, \qquad k \notin \mathcal{K}(f^*),$$

we obtain the equation system

$$\sum_{k \in \mathcal{K}(f^*)} P\nabla_{h_+} g_k(f^*)\beta_k = -Pc_+(h^*). \tag{12}$$

By assumption, the vectors

$$P\nabla_{h_+} g_k(f^*), \qquad k \in \mathcal{K}(f^*),$$

are linearly independent; the vector β of multipliers considered is therefore the unique solution to (12).

It is then an immediate conclusion that the generalized equilibrium travel costs are uniquely determined. □

2.3 Wardrop-type principles

According to Theorem 2.1, solutions to [TAP-SC] satisfy the Wardrop conditions in terms of generalized travel costs, but they will, in general, not satisfy any similar conditions in terms of *actual* travel costs. One can therefore, in general, not relate the actual travel costs of the unused routes to those of the used ones; for example, the least costly route in an OD pair may be unused because its generalized cost is too high. This deficiency is due to the fact that the problem [TAP-SC] does, in contrast to [TAP], in general, *not* possess a Cartesian product structure.

Wardrop-type principles in terms of actual travel costs may however be established if the following assumption is fulfilled at the solution to [TAP-SC]:

Assumption 2.1 (Nondecreasing side constraint functions) *At the flow $f \in F$,*

$$\frac{\partial g_k(f)}{\partial f_a} \geq 0, \qquad \forall a \in \mathcal{A},\ \forall k \in \mathcal{K}.$$

If this assumption holds for any flow $f \in F$, then an increase in the flow on one or more links can never result in any side constraint being more strictly satisfied. Conversely, this assumption holds whenever, for example, the side constraints are general capacity restrictions, that is, when they state upper bounds on volumes of traffic flow on certain links or routes, or in an area of the traffic system.

Further, we use the notions of links and routes that are *unsaturated* with respect to the side constraints.

Definition 2.3 (Unsaturated link and route) *A link $a \in \mathcal{A}$ is said to be unsaturated at the flow $f \in F$ if for all $k \in \mathcal{K}$,*

$$\frac{\partial g_k(f)}{\partial f_a} > 0 \Longrightarrow g_k(f) < 0.$$

A route $r \in \mathcal{R}_{pq}$, $(p,q) \in \mathcal{C}$, is said to be unsaturated at the flow $f \in F$ if all the links $a \in \mathcal{A}$ on route r are unsaturated.

A route is clearly *saturated* at the flow $f \in F$ if $\frac{\partial g_k(f)}{\partial f_a} > 0$ holds for some link a on the route and some $k \in \mathcal{K}$ such that $g_k(f) = 0$.

Theorem 2.7 (Wardrop-type principles) *Suppose that (h^*, f^*) solves the problem* [TAP-SC] *and that Assumption 2.1 holds at f^*. Then, the following conclusions hold for any OD pair $(p,q) \in \mathcal{C}$.*

(a) *The routes utilized in the OD pair have equal and minimal generalized route costs.*

(b) *Assume, with no loss of generality, that the first ℓ routes are utilized in the OD pair and that m of these are unsaturated. Then, the routes may be ordered so that*

$$c_{pq1} = \ldots = c_{pqm} \geq c_{pq,m+1} \geq \ldots \geq c_{pq\ell}.$$

(c) *For any pair of routes $r, s \in \mathcal{R}_{pq}$,*

$$\left.\begin{array}{c} \text{route } r \text{ is unsaturated} \\ c_{pqs} > c_{pqr} \end{array}\right\} \quad \Longrightarrow \quad h^*_{pqs} = 0. \tag{13}$$

(d) *For any pair of routes $r, s \in \mathcal{R}_{pq}$,*

$$\left.\begin{array}{c} \text{route } r \text{ is utilized} \\ c_{pqs} < c_{pqr} \end{array}\right\} \quad \Longrightarrow \quad \text{route } s \text{ is saturated.}$$

Proof.

(a) Immediate from Theorem 2.1.

(b) Follows directly from **(a)**, Assumption 2.1, and the condition (6h).

(c) Noting that

$$\overline{c}_{pqs} \geq c_{pqs} > c_{pqr} = \overline{c}_{pqr} \geq \pi_{pq},$$

where the first inequality is derived from Assumption 2.1 and the condition (6h), the equality follows from the fact that route r is unsaturated, and the last inequality is given by the condition (6b), it then follows from the condition (6a) that $h^*_{pqs} = 0$.

(d) We first observe that

$$\overline{c}_{pqs} \geq \pi_{pq} = \overline{c}_{pqr} \geq c_{pqr} > c_{pqs},$$

where the first inequality is given by condition (6b), the equality follows from the fact that route r is utilized, and the second inequality is derived from Assumption 2.1 and the condition (6h). From the expression (4), Assumption 2.1, and conditions (6h) and (6f), it then follows that route s must contain at least one saturated link, and that it is therefore saturated. $\square$

If the implication in either of the results **(c)** and **(d)** was not fulfilled for some pair of routes, then some traveller would choose to shift to a less costly and unsaturated alternative route; hence, these results are quite natural. As touched upon above, the OD routes that are unused in a solution to [TAP-SC] are not necessarily more costly (in actual travel cost) than those used in the OD pair; this is implied by the result **(d)** since a route may be saturated at zero flow. (Incidentally, such an example can be used to prove that the missing implication in (11) [**(a)** $\Longrightarrow$ **(b)**] is invalid.)

Maugeri (1994) uses the implication (13) as the definition of a generalized user equilibrium solution for an assignment problem with route flow capacity side constraints. He also draws the conclusion of the result **(d)**, by relating the travel cost of route s to that of the most costly route among those used in the OD pair (which does not cause any loss of generality), and making the additional assumption that route s is utilized (i.e., the conclusion is somewhat less general than our result). Note that the results of Theorem 2.7 hold also in the case of non-separable, and possibly asymmetric, travel cost functions (Larsson and Patriksson, 1994, Theorem 2.10).

The first and second results of Theorem 2.7 generalize those stated in Larsson and Patriksson (1995) for the *link flow capacity* side constrained assignment model, that is, for the case

$$\mathcal{K} := \mathcal{A}, \qquad g_a(f) := f_a - u_a, \qquad u_a \in [0, +\infty], \qquad \forall a \in \mathcal{A}.$$

Clearly, in capacitated traffic assignment problems, the constraint functions g_a are nondecreasing (cf. Assumption 2.1), an unsaturated link has a flow which is strictly less than its capacity, and an unsaturated route contains no saturated links (cf. Definition 2.3).

Further, in this case the generalized link travel cost (7) reduces to the simple expression

$$\bar{t}_a(f_a^*) = t_a(f_a^*) + \beta_a^*, \qquad \forall a \in \mathcal{A}, \tag{14}$$

which has been given nice interpretations. To cite Jorgensen (1963), the Lagrange multiplier values β_a^*, $a \in \mathcal{A}$, "measure the time gained by users of routes filled to capacity compared to the fastest route still available." Hence, they are also the link tolls that drivers on saturated routes are willing to pay for being allowed to continue to use routes that are faster than the non-saturated ones. Beckmann and Golob (1974) make a similar observation for a link capacitated *system* optimum assignment model, in which case the multipliers have the interpretation of the link tolls that produce a system optimum when the individual travellers minimize their respective generalized travel costs.

Such interpretations may of course also be given to the the Lagrangean penalty cost terms of the generalized link and route travel costs (7) and (4), respectively; another interpretation is given in the next section.

3 Equilibrium link queueing delays

In the special case of [TAP-SC] where the side constraints are prescriptive link flow capacities (caused by, for example, traffic signals with fixed cycle times at the links' exits), the generalized link travel cost (7) simplifies into the cost (14), and a steady-state link flow can be considered to be in two distinct regimes. In the first part of the link (which includes its entrance), one observes a moving traffic stream; in the second part of the link (which includes its exit), one observes a steady-state queue whenever the link is saturated. (The length of the queue is assumed to be small compared to that of the entire link, so that the travel time in the moving stream can be considered to be unaffected by the presence of the queue.) It is then natural to interpret the value $t_a(f_a^*)$ as being the travel time of the moving traffic stream and the value of the Lagrange multiplier term in (14) as the waiting time in the queue at the link's exit, that is, the link *queueing delay*.

Payne and Thompson (1975) [see also Smith, 1987] use the notion of queue equilibrium to establish a complete equilibrium characterization of solutions to the capacitated problem for the special case of link travel times being constant regardless of the link flows; their result is extended to the case of non-constant link travel times by Miller *et al.* (1975) [see also Inouye, 1987]. In these results, a feasible link flow solution f to a capacitated traffic assignment problem together with a vector $q \in \Re_+^{|\mathcal{A}|}$ of link queueing delays is defined to be a *queue equilibrium* if the links unsaturated at f carry no queues. (Recall that the vector of queueing delays corresponds to Lagrange multipliers for the link capacity constraints and notice that the definition of queue equilibrium is merely a restatement of the complementarity condition (6f) for the special case of capacity side constraints.) The equilibrium characterization of solutions to the capacitated problem may then in our context be stated as follows.

Theorem 3.1 (Equilibrium characterization of solutions to the capacitated model) *Let f be a feasible link flow solution to the capacitated model. It is then an optimal link flow if and only if there is a vector β of non-negative Lagrange multipliers for the capacity constraints such that f is a Wardrop equilibrium with respect to the generalized link travel costs (14) and (f, β) is a queue equilibrium.*

Hence, in this case, the values of the Lagrange multipliers for the capacity constraints may be interpreted as equilibrium link queueing delays, and, further, a steady-state solution has a characterization as a Wardrop equilibrium flow in terms of the sum of travel times and steady-state queueing delays on saturated links. This generalized travel cost is, of course, the natural one to be minimized by the individual travellers in a capacitated network with queueing. (In case the traffic flow is not in a steady-state, the link flows are unstable, the generalized route travel costs and the route flows vary, and the queue at each link's exit is building up if the link is over-saturated and dissolves if it is de-saturating.)

Next, we shall establish a complete equilibrium characterization of solutions to [TAP-SC], that is, we shall generalize the result of Theorem 3.1 to the case of general side constraints. Moreover, the equilibrium link queueing delay formula derived for the general side constrained model turns out to include the equilibrium link queueing delay for the capacitated model as a simple special case. Our development is based on natural generalizations of the link queue and queue equilibrium concepts introduced above.

When a side constraint involves several link flows, it may cause a queue which, in general, is physically distributed on all the links that are affected by the restriction. Hence, each of the traffic flow restrictions may give rise to a *distributed queue.* (One example of such a constraint is given by Ferrari, 1995, in which two traffic streams interfere during the same signal phase.)

Definition 3.1 (Distributed queue equilibrium) *Let f be a feasible link flow solution to* [TAP-SC] *and let $r \in \Re_+^{|\mathcal{K}|}$ be a vector of delays in distributed queues. Then, (f, r) is said to be a distributed queue equilibrium if the traffic flow restrictions which are unsaturated at f have no distributed queues.*

This definition is equivalent to the complementarity condition (6f), and an immediate implication is the following characterization of solutions to [TAP-SC] which corresponds to the equilibrium characterization of solutions to the capacitated model. This is our third main result and includes the partial converse to the result of Theorem 2.1.

Theorem 3.2 (Equilibrium characterization of solutions [TAP-SC]) *Let f be a feasible link flow solution to* [TAP-SC]. *It is then an optimal link flow if and only if there is a vector β of non-negative Lagrange multipliers for the side constraints (3e) such that f is a Wardrop equilibrium with respect to the generalized link travel costs (7) and (f, β) is a distributed queue equilibrium.*

The interpretation of this result is analogous to that of Theorem 3.1.

In the next theorem it is established that a solution to [TAP-SC] is also a generalized Wardrop equilibrium and a link queue equilibrium to a capacitated assignment model; in this model, each link capacity can be viewed as the aggregate effect of all side constraints on the link's capability of carrying flow, and, further, each link queue is composed by contributions from distributed queues.

Theorem 3.3 (Solutions to [TAP-SC] are link queue equilibria) *Suppose that f^* is the optimal link flow solution to* [TAP-SC] *and that Assumption 2.1 holds at f^*. Further, let β^* be a vector of Lagrange multipliers for the side constraints (3e), and let*

$$q_a^* := \sum_{k \in \mathcal{K}} \beta_k^* \frac{\partial g_k(f^*)}{\partial f_a}, \qquad \forall a \in \mathcal{A}. \tag{15}$$

Then, f^* is a Wardrop equilibrium with respect to the generalized link travel costs

$$\bar{t}_a(f_a^*) := t_a(f_a^*) + q_a^*, \qquad \forall a \in \mathcal{A}, \tag{16}$$

and (f^*, q^*) is a queue equilibrium with the respect to the link capacity constraints

$$f_a \leq u_a, \qquad \forall a \in \mathcal{A},$$

where

$$u_a \begin{cases} := f_a^*, & \text{if } q_a > 0, \\ \geq f_a^*, & \text{if } q_a = 0. \end{cases}$$

Proof. By Theorem 2.1 and the expression (7), it follows that f^* is a Wardrop equilibrium with respect to the generalized link travel costs (16).

To establish that (f^*, q^*) is a link queue equilibrium, we first note that $q_a^* \geq 0$ for all $a \in \mathcal{A}$ since the Lagrange multipliers are non-negative and Assumption 2.1 holds at f^*. Next, suppose a link $a \in \mathcal{A}$ is unsaturated at f^*. We then obtain from Definition 2.3 that

$$\beta_k^* \frac{\partial g_k(f^*)}{\partial f_a} > 0 \Longrightarrow \beta_k^* g_k(f^*) < 0, \qquad \forall k \in \mathcal{K}.$$

But the complementarity condition (6f) states that the conclusion of this implication is not true, so that the hypothesis is not true either, and it follows that $q_a^* \leq 0$. Hence, $q_a^* = 0$ must hold.

The result then follows from Theorem 3.1. □

By combining the above result with that of Theorem 3.2, we have established that distributed queue equilibria for [TAP-SC] are also link queue equilibria for a capacitated model. Further, the equilibrium link queueing delays q^* are then the Lagrange multipliers for the link capacity constraints. According to Theorem 3.3, the Lagrangean term of the generalized link travel cost (7) may thus be interpreted as an equilibrium link queueing delay caused by the traffic flow restrictions which are described by the side constraints (3e). The generalized route cost (4) is then the sum of actual travel costs and link queueing delays along the route, and the Lagrange multiplier value π_{pq}^* is the minimal value of these sums over the routes in the OD pair (p, q).

Notice also that the equilibrium link queueing delay formula (15) states that the queue on a certain link may be decomposed into contributions from queueing effects arising from several side constraints. The formula (15) thus provides an equilibrium *link queue representation* result. (If the multipliers β may take alternative values, then the link queue representation is not necessarily unique. In that case, the composition of the link queues that is actually obtained is a consequence of the process which leads to the distributed queue equilibrium.) Further, since the terms of the formula (15) express the physical distributions of queues originating from the saturated traffic flow restrictions among the links, it directly follows that the delays in the distributed queues are given by

$$\tau_k^* := \beta_k^* \sum_{a \in \mathcal{A}} \frac{\partial g_k(f^*)}{\partial f_a}, \qquad \forall k \in \mathcal{K}. \tag{17}$$

Furthermore, interpreting each of the partial derivatives in the expression (15) as a measure of the contribution of the flow on link a to the saturation of the kth side constraint, in the sense that it is a force towards violating the constraint, the expression states that the distribution of the queue is proportional to these forces.

As a consequence of Theorem 3.2, there is an equivalence between solutions to [TAP-SC] and the link flow pattern in the traffic network, provided that the latter is a Wardrop equilibrium with respect to generalized travel costs and a distributed queue equilibrium. It is therefore of interest to establish conditions under which these equilibria will arise; specifically, we need to make assumptions on the travellers' behaviour and on the nature of the traffic flow restrictions described by the side constraints. Clearly, a Wardrop equilibrium with respect to the generalized travel costs may be guaranteed through the traditional assumption that the travellers have a rational route-choice behaviour. Further, we shall in the next section give conditions on the nature of the traffic flow restrictions described by the side constraints which imply that a distributed queue equilibrium arise in the traffic network. These conditions involve the link distribution and dynamical behaviour of the queues.

4 Queue dynamics

We now introduce assumptions about the physical nature of the traffic flow restrictions which are modelled by the side constraints, and show that stationary states then are distributed queue equilibria. First, we assume that each traffic flow restriction may cause a queue which is distributed among the links as stated below.

Assumption 4.1 (Delays in distributed queues) *There exist parameters $\gamma_k \geq 0$, $k \in \mathcal{K}$, such that, for any flow $f \in F$ and any traffic flow restriction $k \in \mathcal{K}$, the portion of the distributed queue which is physically located on a link $a \in \mathcal{A}$ has queueing delay $\gamma_k \frac{\partial g_k(f)}{\partial f_a}$.*

Second, we assume that the traffic flow restrictions under consideration are prescriptive (hard) and can never be violated in a stationary state, and that each distributed queue appears only when the corresponding traffic flow restriction is non-redundant. Further, if the traffic flow is in a disequilibrium state and the travellers successively adjust their route-choices with respect to the (varying) generalized travel costs (i.e., actual travel costs and delays in distributed queues), then, at any moment, the distributed queue arising from a traffic flow restriction will be building up or dissolving depending on whether or not the restriction is violated.

Assumption 4.2 (Distributed queue dynamics)

(a) (Stationary queueing delays) *If a traffic flow restriction is saturated at some flow $f \in F$, that is, $g_k(f) = 0$ for some $k \in \mathcal{K}$, then the queueing delay of the distributed queue is in a stationary state, that is, the parameter γ_k has a constant value.*

(b) (Unlimited non-stationary queueing delays) *If a traffic flow restriction is violated at some flow $f \in F$, that is, $g_k(f) > 0$ for some $k \in \mathcal{K}$, then the queueing delay of the distributed queue is non-stationary and will eventually become arbitrarily large, that is, the parameter value γ_k tends to infinity.*

(c) (Vanishing non-stationary queueing delays) *If a traffic flow restriction is unsaturated at some flow $f \in F$, that is, $g_k(f) < 0$ for some $k \in \mathcal{K}$, then the queueing delay of the distributed queue is non-stationary and will eventually vanish, that is, the parameter value γ_k tends to zero.*

In the sequel it will be shown that the parameters γ_k, $k \in \mathcal{K}$, play the role of Lagrange multipliers. The following lemma is then needed.

Lemma 4.1 *Suppose that for some $\bar{f} \in F$ and $k \in \mathcal{K}$, $g_k(\bar{f}) > 0$ holds, and that Assumption 2.1 holds at $\bar{f}$. Then, there is an $a \in \mathcal{A}$ such that $\bar{f}_a > 0$ and $\frac{\partial g_k(\bar{f})}{\partial f_a} > 0$.*

Proof. If the conclusion is not true, then, since $\bar{f} \geq 0$, $\nabla g_k(\bar{f})^{\mathrm{T}} \bar{f} = 0$. From the convexity of the function g_k it then follows that, for any flow $f \in F$,

$$g_k(f) \geq g_k(\bar{f}) + \nabla g_k(\bar{f})^{\mathrm{T}}(f - \bar{f}) = g_k(\bar{f}) + \nabla g_k(\bar{f})^{\mathrm{T}} f \geq g_k(\bar{f}) > 0,$$

since $f \geq 0$, which contradicts that [TAP-SC] has a feasible solution. □

The following theorem is our fourth main result.

Theorem 4.1 (Stationary flows solve [TAP-SC]) *Let $f \in F$, and suppose that Assumption 2.1 holds at f. If, in addition, it is a stationary flow with respect to the link travel costs $t_a(f_a)$, $a \in \mathcal{A}$, and the link queueing delays, then, under Assumptions 4.1 and 4.2, it is also an optimal link flow solution to* [TAP-SC].

Proof. To establish the conclusion, we will show that f and $\beta_k := \gamma_k$, $k \in \mathcal{K}$, satisfy the sufficient optimality conditions (6).

(Feasibility in side constraints) Suppose that $g_k(f) > 0$ for some $k \in \mathcal{K}$. Then, according to Lemma 4.1, there is an $a \in \mathcal{A}$ such that $f_a > 0$ and $\frac{\partial g_k(f)}{\partial f_a} > 0$. From Assumptions 4.1 and 4.2 it follows that the queueing delay on link a tends to infinity, which contradicts that $f_a > 0$ in a stationary flow. Hence, $g_k(f) \leq 0$ for all $k \in \mathcal{K}$, that is, condition (6g) is satisfied.

(Complementarity) If $g_k(f) < 0$ for some $k \in \mathcal{K}$, then, according to Assumption 4.2, the value of the parameter γ_k tends to zero, so that condition (6f) becomes satisfied in a stationary state.

(Optimality) Clearly, the remaining sufficient optimality conditions are also fulfilled, and the result follows. □

Hence, under Assumptions 2.1, 4.1, and 4.2, any stationary flow in the transportation network is also an optimal solution to [TAP-SC], and we have thus established that the set of optimal solutions to [TAP-SC] then coincides with the steady-state flows in the transportation network.

We conclude this section with the observation that the characterization of solutions to [TAP-SC] as being Wardrop equilibria (with respect to generalized travel costs) and distributed queue equilibria suggests that the utilization of side constraints in a traffic equilibrium assignment model is consistent with the assumption of rational traveller behaviour. We also note that the equilibrium characterizations of solutions to [TAP-SC] established in this work have immediate counterparts for more general models, such as, for example, models with non-separable travel costs, for which the corresponding results are found in Larsson and Patriksson (1994).

5 Solving the side constrained model

Whenever side constraints are introduced into a traffic assignment model, the traditional solution methods, such as the Frank–Wolfe algorithm and its relatives, either become inapplicable or their efficiency is seriously degraded. In particular, the linear programming subproblem of the Frank–Wolfe type algorithms does no more separate into a number of shortest route calculations; for example, in the simple special case of link capacity side constraints the subproblem becomes a linear multi-commodity network flow problem, which is prohibitively expensive to solve repeatedly. In addition, the existing program packages do not possess the ability to take side constraints into account.

5.1 A price-directive solution strategy

When considering possible solution principles for the side constrained model, it is most natural to aim at exploiting the efficient solution methods and program packages that are available for the basic model. This immediately leads us to a classical approach for handling complicating constraints: the *pricing* strategy (e.g., Lasdon, 1970, Chapter 8).

We associate with the side constraints (3e) non-negative prices β_k, $k \in \mathcal{K}$, for violating them. Given certain values of these prices, the side constraints are priced-out, that is, handled implicitly by being included in the objective function only. The *priced-out* problem,

[TAP(β)]

$$\underset{f \in F}{\operatorname{minimize}} \; L(f,\beta) := T(f) + \beta^{\mathrm{T}} g(f),$$

is an equilibrium assignment problem with the link travel cost mapping

$$\nabla L(\cdot,\beta) := t(\cdot) + \nabla g(\cdot)^{\mathrm{T}} \beta,$$

that is, with the generalized link travel cost functions

$$t_a(f_a) + \sum_{k \in \mathcal{K}} \beta_k \frac{\partial g_k(f)}{\partial f_a}, \qquad \forall a \in \mathcal{A},$$

which, in general, are non-separable though. The priced-out problem is solvable with most standard methods for the basic model, and the resulting link flow, denoted $f(\beta)$, is unique, since the objective $L(\cdot,\beta)$ is strictly convex with respect to the link flows.

The solution to the priced-out problem may be characterized as the solution to a side constrained assignment problem where the right-hand sides of the original side constraints are modified through certain perturbations; this result follows immediately from Everett's Theorem (e.g., Lasdon, 1970, Thm. 8.3).

Theorem 5.1 (An Everett-type result) *For any price vector $\beta \in \Re_+^{|\mathcal{K}|}$, the solution $f(\beta)$ to the priced-out problem* [TAP(β)] *solves the side constrained traffic equilibrium assignment problem*

[TAP-SC(β)]

$$\operatorname{minimize} T(f),$$

subject to

$$\begin{aligned}
\sum_{r \in \mathcal{R}_{pq}} h_{pqr} &= d_{pq}, && \forall (p,q) \in \mathcal{C}, \\
h_{pqr} &\geq 0, && \forall r \in \mathcal{R}_{pq},\ \forall (p,q) \in \mathcal{C}, \\
\sum_{(p,q) \in \mathcal{C}} \sum_{r \in \mathcal{R}_{pq}} \delta_{pqra} h_{pqr} &= f_a, && \forall a \in \mathcal{A}, \\
g_k(f) &\leq \overline{g}_k(\beta), && \forall k \in \mathcal{K},
\end{aligned}$$

where

$$\overline{g}_k(\beta) := \begin{cases} g_k(f(\beta)), & \textit{if } \beta_k > 0, \\ \max\{0, g_k(f(\beta))\}, & \textit{if } \beta_k = 0. \end{cases}$$

This result is useful in case the side constraints are descriptive (soft) and thus not need to be fulfilled exactly (see below). Further, if the side constraints state traffic management goals which are to be achieved through the introduction of link tolls, then the solution of [TAP(β)] may be seen as a simulation of the effect of a tentative toll schedule on the traffic flow, and the result of Theorem 5.1 provides a means for evaluating the toll schedule with respect to the goals formulated.

The reader may note that the equivalence result stated in Theorem 2.2 is a special case of the conclusion in Theorem 5.1; as the price vector tends to a vector of Lagrange multipliers, the solution $f(\beta)$ will, because of the strict convexity of T, tend continuously to the link flow f^*, so that the right-hand sides $\overline{g}_k(\beta)$ of the side constraints of the problem [TAP-SC(β)] tend continuously to zero, and the problem [TAP-SC(β)] tends to [TAP-SC].

In order to find correct values of the prices (i.e., Lagrange multipliers) one may solve the *Lagrangean dual* problem

[TAP-SCD]

$$\underset{\beta \geq 0}{\text{maximize}} \;\; L(\beta),$$

where β is now interpreted as a vector of dual variables and the Lagrangean dual objective function is given by

$$L(\beta) := \underset{f \in F}{\text{minimum}} \;\; L(f, \beta).$$

Lagrangean dual problems are typically solved using simple iterative search methods for (essentially) unconstrained optimization. Within a dual solution procedure for [TAP-SCD], the result of Theorem 5.1 may be utilized for monitoring the progress of the procedure with respect to the aim of finding a solution to [TAP-SC]. This result also facilitates the finite termination of the dual algorithm when the solution is near-feasible with respect to the side constraints. Clearly, near-feasible solutions are often satisfactory considering the uncertainties in the input data; near-feasibility is, of course, also satisfactory when the side constraints are soft.

Larsson and Patriksson (1995) develop and evaluate an *augmented Lagrangean* dualization (i.e., *nonlinear* pricing) technique for the link capacity side constrained equilibrium model. They establish the efficiency of this technique for finding the correct values of the multipliers β and for solving the capacitated model, and also conclude that this dualization scheme is in both these respects more efficient than the traditional Lagrangean dualization. For certain instances of this augmented Lagrangean scheme, and under additional technical assumptions, the sequence of dual iterates generated can be shown to converge (at least linearly) although the set of dual solutions is not a singleton in general; further, under some additional assumptions, the limit point of the iterates is the dual solution which has the least Euclidean norm.

5.2 Interpretations of the solution strategy

Using the result of Theorem 2.2 for constructing improved travel cost functions, we note that the application of an iterative solution procedure to the dual problem [TAP-SCD] then can be given the nice interpretation of an automatized process of adjusting the travel cost functions towards the correct ones, which are reached in the limit.

With reference to the equilibrium characterizations of solutions to [TAP-SC] stated in Section 3, the Lagrangean dual problem [TAP-SCD] has the interesting interpretation of being the problem of finding a distributed queue equilibrium (and also a link queue equilibrium), under the implicit assumption that the traffic flow pattern is a Wardrop equilibrium with respect to actual travel costs and link queueing delays. The evaluation of the dual objective value $L(\beta)$ [where β is given and possibly non-optimal] then amounts to finding a Wardrop equilibrium with respect to the generalized link travel costs

$$t_a(f_a) + q_a(f), \qquad \forall a \in \mathcal{A},$$

with the flow-dependent link queueing delays

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

[cf. the expressions (16) and (15)]. Further, a non-negative price vector solves the dual problem if and only if the resulting distributed queues [whose delays are of the form (17) and are calculated for the given β and at the generalized Wardrop equilibrium flow] are at an equilibrium in accordance with Definition 3.1; moreover, the resulting link queues (i.e., those given by the expression above and calculated at the generalized Wardrop equilibrium flow) are then at an equilibrium in the sense of Theorem 3.3.

5.3 Price-directive traffic management through link tolls

We conclude this section by discussing the derivation of price-directive traffic management schemes through the solution of the side constrained equilibrium problem. The concept of price-directive decentralized planning is well known within the fields of economics and operations research (see, for example, Dirickx and Jennergren, 1979), and is closely related to that of Lagrangean duality. Often, it aims at coordinating decisions in a system with decentralized decision-making by the means of prices (typically, on scarce resources) so that an optimal policy for the system as a whole is obtained. (See, for example, Bernstein and Smith, 1994, for an overview of the use of pricing in network equilibria, and the references cited therein for more details.)

We consider a situation in which the managers of a traffic system wish to reach certain goals with respect to the performance of the system. The goals may, for example, involve travel times on certain links or routes (for example, in order to provide small travel times in certain OD pairs), or volumes of traffic flow on certain links or routes, or into a central area of the traffic system (for example, in order to avoid congestion or high concentrations of exhaust fumes or levels of noise in sensitive areas). Further, the tool to be used for reaching these goals is the introduction of link tolls, which divert the travellers' route-choices from the natural, travel-time minimizing, ones.

The levels of aspiration of the goals are formulated as (one or more) side constraints, and the solution of the side constrained assignment problem is a means for calculating the proper link tolls [that is, the Lagrangean penalty cost terms of the generalized link travel costs (7)], which, when imposed upon the individual travellers, change the route-choice behaviour so that the traffic management goals are reached, without the need to resort to a more direct or centralized traffic control.

The proper link tolls can alternatively be determined through the solution of the dual problem [TAP-SCD] (which can be thought of as the problem of finding the tolls which make the side constraints satisfied and optimally utilized). Moreover, the solution of the dual problem using an iterative search procedure may be interpreted as a mathematical *simulation* of a real-life process in which a traffic engineer attempts to enforce some traffic management goals by introducing link tolls and modifying them until the travellers' behavioural response is the intended one.

(This strategy for finding suitable link tolls can certainly not be implemented in the real-life traffic system.)

In the context of price-directive traffic management it may also be of interest to exploit the fact that the link tolls which lead to the fulfilment of the management goals are not necessarily uniquely determined, and that there may therefore be an option to choose a link toll schedule (among the proper ones) that optimizes some secondary criterion (e.g., the one which minimizes the total toll revenues). These considerations are a subject for ongoing research; see Larsson and Patriksson (1997, 1998a, and 1998b) for results obtained so far.

6 Extensions

Most of the results presented in this paper have immediate generalizations to a side constrained non-separable and asymmetric traffic equilibrium model, that is, a side constrained extension to [VIP]; these results are reported in Larsson and Patriksson (1994).

The model [TAP-SC] may be extended to incorporate elastic demands; in this case, in contrast to the fixed demand model considered here, the multiplier vector β for the side constraints is (essentially) unique, due to the fact that the demand and link flow solution is unique and the demand function is sensitive to the value of the generalized travel cost. For characterizations of the solutions to this model and its application in the context of price-directive traffic management through link tolls, we refer to Larsson and Patriksson (1998b).

We next describe in brief possible extensions of the model to more general side constraints, and the consequences of the extensions for the assumptions needed in the underlying equilibrium model and on the toll scheme. Suppose that some management goals are described in terms of the individual commodity flows. Such side constraints could, for example, describe a desired distribution of flow among different modes of transport, and may be used to distinguish between, and differentiate the tolls imposed upon, different classes of users of the network. For example, we may consider restrictions of the form $g_k([f_{pq}]) \leq 0$, where f_{pq} denotes the vector of link flows in commodity (p,q) and $[f_{pq}]$ is the concatenated vector of all the commodity link flow vectors. If we are interested in obtaining flows satisfying these restrictions through the use of tolls, then the tolls would obviously have to be differentiated between the individual commodities, that is, a link toll that is uniform over all the commodity flows on a link would be inadequate. This is quite evident from the equilibrium characterizations of such an extension of the side constrained traffic equilibrium model [TAP-SC], in which the generalized link travel cost $\bar{t}_a(f^*)$, $a \in \mathcal{A}$, in (7) is replaced by

$$\bar{t}_{apq}(f^*) := t_a(f_a^*) + \sum_{k \in \mathcal{K}} \beta_k^* \frac{\partial g_k([f_{pq}^*])}{\partial f_{apq}}, \qquad a \in \mathcal{A}, \quad (p,q) \in \mathcal{C}; \tag{18}$$

clearly, the toll, given by the second term of (18), is in general different for the different commodities on the same link, and can therefore in general not be represented by a uniform toll.

Having concluded that it is necessary to levy individual tolls for the different commodities, we next turn to the question of whether such a toll will achieve the traffic management goals. Solving the side constrained traffic equilibrium model [TAP-SC] with the above commodity-specific side constraints, we obtain a unique link flow solution, and a consistent commodity flow that satisfies these constraints. From the use of a dual scheme for the solution of [TAP-SC], we may also automatically generate a vector of multipliers for the side constraints, and thus a toll vector, equal to the second term of (18). This commodity-specific, fixed link toll will, however *not* necessarily influence the flow to satisfy the side constraints, even though they will produce the unique link flow solution. The reason for this seemingly counter-intuitive result is that while the link flow solution to [TAP-SC] (likewise the optimal Lagrange subproblem [TAP-SC(β^*)])

is unique, its decomposition into commodity flows is not, under the given assumptions; among the commodity link flows that are consistent with the optimal total link flows, there are some that do not satisfy all the side constraints, and some that do but are non-optimal in [TAP-SC]. In the context of price-directive, decentralized planning, this property is known as the *non-coordinability* phenomenon (cf. Dirickx and Jennergren, 1979, Chapter 2). In technical terms, the reason is that while the objective function T of [TAP-SC] (likewise the objective $L(\cdot, \beta^*)$ of [TAP-SC(β^*)]) is strictly convex in the total link flows, it is non-strictly convex in the individual commodity link flows. It thus becomes clear that if commodity-specific goals are to be achieved, not only do we need to consider individual commodity link tolls, but the equilibrium model must have the property that the commodity flow is uniquely determined. (This property holds, for example, when the model includes individual link travel cost functions t_{apq} for the different commodities which are also strictly increasing in the respective flow variables f_{apq}.)

We may generalize the above discussion to consider even more general side constraints of the form $g_k(h) \leq 0$, in which the goals are described in terms of the individual route flows. In this case, the generalized equilibrium route travel costs given in Definition 2.1 are extended to

$$\bar{c}_{pqr} := c_{pqr}(h^*) + \sum_{k \in \mathcal{K}} \beta_k^* \frac{\partial g_k(h^*)}{\partial h_{pqr}}, \qquad r \in \mathcal{R}_{pq}, \quad (p,q) \in \mathcal{C}. \tag{19}$$

We conclude that in this case, the tolls would have to be imposed upon the travellers on the basis of their specific route-choices, and the underlying equilibrium model must have the property that the route flow is uniquely determined. We remark that in order for such a result to be established, the travel cost functions must have the property that the travel cost on a route cannot be calculated as the sum of the travel costs on the links defining the route, that is, travel costs must be *non-additive*, since otherwise individual route flows cannot be distinguished with respect to their individual costs; clearly, any travel cost functions that are based on total link flows, such as the ones utilized in the models [TAP] and [TAP-SC], violate this condition, and thus will not produce unique equilibrium route flow solutions.

There is a class of equilibrium models which do produce such solutions, the class of *stochastic user equilibrium* models (e.g., Sheffi, 1985). In such models, the perception of the least travel cost is assumed to vary among the trip-makers according to some (known) probability distribution, and the distributions most often considered lead to unique stochastic user equilibrium route flows. (Stochastic user equilibrium models most often utilize the same, link-flow based, travel cost functions as in the deterministic model [TAP], but when viewing the corresponding deterministic equivalent models, one can not identify such travel cost functions.) We remark that most of the results presented in this paper are transferable from the deterministic case considered to the framework of stochastic user equilibrium models.

We conclude this discussion with the remark that if we wish to achieve traffic management goals stated at a given level of flow disaggregation (into commodity link flows or route flows), then the tolls must be levied at a flow disaggregation level that is at least as high as that at which the goals were formulated, and the equilibrium model used as a basis for the derivation of these tolls most produce unique equilibrium flow solutions at the same level of disaggregation.

We finally remark that it is possible to combine the descriptive and prescriptive modelling approaches into a single model, for example to describe a traffic management model for a traffic network with queueing (e.g., Yang and Lam, 1996; Yang and Bell, 1997; Larsson and Patriksson, 1998b).

7 Conclusions and further research

It is our belief that it should in many traffic assignment contexts, and for different purposes, be beneficial to utilize side constraints for refining a model, especially since the side constraints may in many situations be relatively easy to derive and calibrate. The results presented in this paper provide a theoretical justification for the utilization of side constraints as a means for refining traffic equilibrium assignment models, as well as a solid basis and a strong motivation for a continued, theoretical and practical, exploration of this modelling strategy; a particularly strong motive for a continued study of this strategy is the observation that the inclusion of side constraints in an equilibrium assignment model is equivalent to a well-defined adjustment of the travel cost functions. The use of side constraints to derive price-directive traffic management schemes is also an interesting topic for further study.

It would also be of interest to further study the relationship between side constraints that arise as a result of traffic control and the policies employed in the control. An example of this is a side constraint which describes the maximal traffic flow through a junction with vehicle-responsive signals, in which case it is of interest to establish the existence of a control policy which is consistent with the equilibrium link queues which can be calculated from the expression for the side constraint; such a result can probably be based on the work by Smith (1987).

To facilitate the practical utilization of side constraints for refining assignment models and the real-life exploitation of this modelling strategy, effective and efficient computational tools for the solution of the resulting models need to be developed and incorporated in assignment packages. Such tools can probably be based on the application of an augmented Lagrangean solution principle, which was in Larsson and Patriksson (1995) successfully applied to the link capacitated model. The results obtained in that study suggest that also other classes of side constrained traffic assignment models can be efficiently dealt with computationally, although the side constraints destroy the Cartesian product structure of the traditional assignment models. Further, the augmented Lagrangean solution principle does not impose any significant limitations on the design of the model since it can handle both nonlinear and non-separable side constraints.

Acknowledgements

The research leading to this paper started while the second author was visiting the Department of Mathematics, University of York, Heslington, York, England. He wishes to thank Professor Michael J. Smith of the University of York for his hospitality and many interesting discussions, and the Swedish Transport and Communications Research Board (KFB) (Grant TFB 93-131-63), the Swedish Institute (Grant 303 GH/MLH), and the Royal Swedish Academy of Sciences for financing the visit. The first author was sponsored by the Swedish Transport and Communications Research Board (Grants TFB 92-128-63 and KFB 95-118-63).

Bibliography

[1] Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network Flows: Theory, Algorithms, and Applications.* Englewood Cliffs, NJ: Prentice-Hall.

[2] Akcelik, R. (1988). Capacity of a shared lane. *Australian Road Research Board Proceedings, 14(2),* pp. 228–241.

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

[4] 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.

[5] Beckmann, M., McGuire, C. B., and Winsten, C. B. (1956). *Studies in the Economics of Transportation.* New Haven, CT: Yale University Press.

[6] Bernstein, D., and Smith, T. E. (1994). Equilibria for networks with lower semicontinuous costs: With an application to congestion pricing. *Transportation Science, 28,* pp. 221–235.

[7] Charnes, A., and Cooper, W. W. (1961). Multicopy traffic network models. In R. Herman (ed.), *Theory of Traffic Flow*, Proceedings of the Symposium on the Theory of Traffic Flow Held at the General Motors Research Laboratories, Warren, MI, December 7–8, 1959 (pp. 85–96). Amsterdam, The Netherlands: Elsevier.

[8] Choi, W., Hamacher, H. W., and Tufekci, S. (1988). Modeling of building evacuation problems by network flows with side constraints. *European Journal of Operational Research, 35,* pp. 98–110.

[9] Dafermos, S. C. (1972). The traffic assignment problem for multiclass-user transportation networks. *Transportation Science, 6,* pp. 73–87.

[10] Dirickx, Y. M. I., and Jennergren, L. P. (1979). *Systems Analysis by Multilevel Methods.* Chichester, England: John Wiley & Sons.

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

[12] Friesz, T. L., Bernstein, D., Mehta, N. J., Tobin, R. L., and Ganjalizadeh, S. (1994). Day–to–day dynamic network disequilibria and idealized traveler information systems. *Operations Research, 42,* pp. 1120–1136.

[13] Harker, P. T., and Pang, J.-S. (1990). Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. *Mathematical Programming, 48,* pp. 161–220.

[14] 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.

[15] Inouye, H. (1987). Traffic equilibria and its solution in congested road networks. In R. Genser (ed.), *Control in Transportation Systems*, Proceedings of the IFAC Conference on Control in Transportation Systems, Vienna, Austria, 1986 (pp. 267–272). Oxford, England: Pergamon.

[16] Jorgensen, N. O. (1963). *Some aspects of the urban traffic assignment problem.* Graduate Report, Institute of Transportation and Traffic Engineering, University of California, Berkeley, CA.

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

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

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

[20] Larsson, T., and Patriksson, M. (1997). Price-directive traffic management: an approach utilizing side constrained traffic equilibrium models. *Rendiconti del Circolo Matematico di Palermo, Serie II, 48*, pp. 147–170.

[21] Larsson, T., and Patriksson, M. (1998a). Price-directive traffic managemen: applications of side constrained traffic equilibrium models. In R. E. Allsop and M. G. H. Bell (eds.), *Transportation Networks: Recent Methodological Advances*, Proceedings of the 4th Meeting of the EURO Working Group on Transportation, Newcastle University, Newcastle, UK, September 9–11, 1996 (to appear).

[22] Larsson, T., and Patriksson, M. (1998b). Side constrained traffic equilibrium models: traffic management through link tolls. In P. Marcotte and S. Nguyen (eds.), *Equilibrium and Advanced Transportation Modelling*, Proceedings of the Equilibrium and Advanced Transportation Modelling Colloqium, Centre de recherche sur les transport, Université de Montréal, Montréal, Canada, October 10–11, 1996 (to appear). New York, NY: Kluwer Academic Publishers.

[23] Lasdon, L. S. (1970). *Optimization Theory for Large Systems.* New York, NY: Macmillan.

[24] Maugeri, A. (1994). Optimization problems with side constraints and generalized equilibrium principles. *Le Matematiche, 49,* pp. 305–312.

[25] 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.

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

[27] Nagurney, A., and Zhang, D. (1996). *Projected Dynamical Systems and Variational Inequalities with Applications.* Boston, MA: Kluwer Academic Publishers.

[28] Patriksson, M. (1994). *The Traffic Assignment Problem: Models and Methods.* Utrecht, The Netherlands: VSP.

[29] 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.

[30] Sender, J. G., and Netter, M. (1970). *Équilibre offre-demande et tarification sur un réseau de transport.* Rapport de recherche 3, Département Economie, Institut de Recherche des Transports, Arcueil, France.

[31] Sheffi, Y. (1985). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods.* Englewood Cliffs, NJ: Prentice-Hall.

[32] Shepardson, F., and Marsten, R. E. (1980). A Lagrangean relaxation algorithm for the two duty period scheduling problem. *Management Science, 26,* pp. 274–281.

[33] Smith, M. J. (1987). Traffic control and traffic assignment in a signal-controlled network with queueing. Paper presented at the *Tenth International Symposium on Transportation and Traffic Theory*, Boston, MA, 1987.

[34] Toint, Ph., and Wynter, L. (1996). Asymmetric multiclass traffic assignment: a coherent formulation. In J.-B. Lesort (ed.), *Transportation and Traffic Theory, Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France, 24–26 July, 1996 (pp. 237–260). Oxford, Great Britain: Pergamon.

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

[36] Weigel, H. S., and Cremeans, J. E. (1972). The multicommodity network flow model revised to include vehicle per time period and node constraints. *Naval Research Logistics Quarterly, 19,* pp. 77–89.

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

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

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