Equilibrium Characterizations of Solutions to Side Constrained Asymmetric Traffic Assignment Models

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

Michael Patriksson
Department of Mathematics, Box 354350
University of Washington
Seattle, WA 98195-4350, USA

Extended abstract

The classical model of user equilibrium is well studied and frequently applied in transportation analysis. Its popularity among practitioners is partially explained by its simplicity and nice interpretations, which makes it easy to access and apply. The validity of the model, however, rests on assumptions on the real-world traffic flows being modelled which may be too restrictive in certain applications.

The model may also be invalid due to its structural limitations, that is, its inherent simplicity makes it inapplicable to more complex traffic problems. Examples of such limitations is that there is no discrimination between different types of vehicles, and that every road is (implicitly) presumed to be able to carry arbitrarily large volumes of traffic. The effect of the model's failure to produce reasonable predictions of traffic flows is that the traffic engineer either ignores its predictions or, more often, perturbs components of the model (e.g., the parameters of the travel cost functions) in attempting to make it provide an output more in line with the expected one. 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.

When refining the basic model of traffic assignment to capture supplementary flow relationships the traditional modelling strategy is to modify the travel cost mapping. This has been done successfully especially in the modelling of delays at priority junctions, and also in the modelling of interactions between the flows on intersecting links and between vehicles of different types. The resulting models often have nonseparable and asymmetric travel cost functions, and a solution to the Wardrop conditions of user equilibrium can therefore not be found by solving a convex optimization problem in the form of the classical traffic assignment problem; instead, they are solved as variational inequality or complementarity models.

The travel cost function is not the only modelling instrument available, and there are several examples of flow relationships and restrictions for which the travel cost function is not the proper modelling tool; examples include traffic flow restrictions imposed by joint capacities on two-way streets or in junctions, or the presence of a traffic control policy. For such supplementary relationships we believe that it is more appropriate to introduce side constraints into the model. In general, side constraints should be introduced to describe those flow relationships that have more natural interpretations as flow restrictions than as additional travel costs; such a refinement should be easier to construct and calibrate as well as lead to more reliable traffic models than that using the traditional refinement strategy only.

Supplementary constraints on traffic flows may also be introduced in order to improve the quality of an available assignment model by incorporating into the model additional information

about the actual real-world traffic situation, for example link flow observations, and to capture dynamic effects in time-sliced traffic assignment, e.g., to describe the coupling between the assignment problems in successive time-slices.

An interesting application of side constraints in assignment models is that of modelling and deriving price-directive traffic management schemes. Consider a situation in which the managers of a traffic system wish to reach certain goals with respect to the performance of the system; typical examples of performance goals are maximal travel times and traffic volumes on certain links. Further, the tool to be used for reaching these goals is the introduction of link tolls, which divert the travellers' route-choices from the travel-time minimizing ones. The levels of aspiration of the goals are then formulated as a set of side constraints, and the solution of the side constrained problem is the means for calculating the proper link tolls, which, when imposed upon the individual travellers, change their 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 link tolls are derived by using the values of the multipliers of the side constraints.) This technique for deriving link tolls is appealing since it is quite natural to formulate traffic management goals in terms of side constraints, so that, in fact, it supports the process of identifying the proper goals to be formulated, and, further, since it is quite flexible with respect to which goals that can be considered.

The side constraints introduced may differ significantly with respect to their purposes and properties; we may, however, distinguish two principally different types: *prescriptive* (hard) and *descriptive* (weak). The former are imposed upon the travellers by traffic management and control policies (e.g., through speed limit regulations and traffic signals), are known exactly and can (or may) never be violated; examples include link capacities which are used to model saturation flows and side constraints introduced to derive link tolls. Prescriptive constraints may be binding at a steady-state flow and cause stationary link queues to appear. Descriptive side constraints are introduced as a means to bring the calculated flow in better agreement with the anticipated one (e.g., observed flows on some links), as a means to refine the model by including additional (maybe approximate) traffic flow restrictions into the model (e.g., joint capacities in roundabouts), or as a (rough) means for modelling congestion effects. Clearly, because of the nature of the descriptive constraints they do not need to be satisfied exactly, and this fact may also be taken into account in solution procedures.

Side constraints have been considered to a very limited extent (either to model congestion effects or to introduce link capacity constraints); the main reasons for this are that the solutions to the resulting models can no longer be given characterizations as Wardrop equilibria in the classical sense, and that the models are computationally more demanding due to the loss of the Cartesian product structure of the feasible set.

A main motive behind this paper is to show that a solution to a side constrained model has a characterization as a Wardrop equilibrium flow in terms of well-defined generalized travel costs; this cost is composed of the original travel cost and cost components defined by the first partial derivatives of the side constraint functions and the multipliers of the side constraints. Depending on the nature of the side constraints, these additional cost components have different interpretations: in a network with queueing, they are the delays in steady-state link queues, while, in an application to traffic management, they define the link tolls that should be imposed on the travellers in order to divert the travellers from the least-cost routes to reach the traffic management objectives. These two examples serve to establish that the generalized travel cost may be a very natural cost to be minimized by the travellers, i.e, the generalized equilibrium conditions comply with the basic assumption of rational traveller behaviour.

Another motive is to show that a side constrained model is computationally tractable, by proposing the use of a general pricing scheme. Using linear or nonlinear Lagrangean penalty functions (price functions) for the side constraints (which generalize the principles of Lagrangean methods and augmented Lagrangean methods, respectively, from the field of nonlinear programming), the original problem may be solved as an ordinary user equilibrium problem with an adjusted travel cost function, once multipliers for the side constraints are at hand; schemes for obtaining these multiplier values are embedded in most (augmented) Lagrangean algorithms.

Utilizing the appropriate combination of the two above mentioned modelling strategies results in a variational inequality model of the traffic equilibrium problem augmented with a set of side constraints. The side constraints are assumed to be of the form

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

where the functions $g_k : \Re_+^{|\mathcal{A}|} \mapsto \Re$, $k \in \mathcal{K}$, are continuously differentiable. (The index set $\mathcal{K}$ may, for example, consist of the index set of the links, nodes, routes, or O-D pairs, or any combination of subsets of them.) We denote by G the set of solutions to the system of inequalities (and assume that a constraint qualification holds for G).

The side constrained traffic equilibrium problem then is to find an $f^* \in F \cap G$ such that

[TAP-VIP-SC-F]

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

where $t : \Re_+^{|\mathcal{A}|} \mapsto \Re_{++}^{|\mathcal{A}|}$ is the link travel cost mapping (assumed continuous), and F is the set of demand-feasible link flows. (Standard notation is used to denote the vector of link flows, etc.)

In our first result we show that a solution to [TAP-VIP-SC-F] is a Wardrop equilibrium in terms of generalized travel costs, that is, in terms of the route travel costs

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

where β^* is any vector of multipliers for the side constraints, and (h^*, f^*) is any route and link flow solution to [TAP-VIP-SC-F].

The above relation translates into

$$\overline{t}_a(f^*) \overset{\text{def}}{=} t_a(f^*) + \sum_{k \in \mathcal{K}} \beta_k^* \frac{\partial g_k(f^*)}{\partial f_a}, \qquad \forall a \in \mathcal{A}, \tag{1}$$

which defines the corresponding generalized link travel costs at equilibrium.

In general, the multipliers for the side constraints are not unique, and the sets of least-cost routes (in the sense of the generalized equilibrium travel cost) are not the same for different values of these multipliers; consequently, the generalized equilibrium link and route travel costs $\overline{t}$ and $\overline{c}$, respectively, are not unique. The uniqueness of the multipliers and the sets of least-cost routes is, however, a natural criterion for the well-posedness of the model, in the sense that it describes a stability property.

In our second set of results we establish the uniqueness of the generalized equilibrium travel costs and multiplier values as well as the invariance of the sets of least-cost routes on the solution set of [TAP-VIP-SC-F] under additional assumptions on the travel cost mapping and the nature of the side constraints; most notably, we need to assume that the side constraint functions are nondecreasing in the sense that

$$\frac{\partial g_k(f)}{\partial f_a} \geq 0, \qquad \forall a \in \mathcal{A},\ \forall k \in \mathcal{K} \tag{2}$$

holds at any link flow solution, and also strong constraint qualifications.

Our third set of results establishes that a Wardrop-type principle in terms of actual travel costs exists under the assumption (2). We then utilize the notion of an *unsaturated* link and route, meaning that a link $a \in \mathcal{A}$ is unsaturated at the flow $f \in F$ if for all $k \in \mathcal{K}$,

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

and that a route $r \in \mathcal{R}_{pq}$ in the O-D pair $(p,q) \in \mathcal{C}$ is unsaturated at the flow $f \in F$ if all the links $a \in \mathcal{A}$ on route r are unsaturated. Assuming, with no loss of generality, that the first l routes are utilized in the O-D pair and that m of these are unsaturated, the routes may then be ordered so that

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

Also, for any pair of routes $r, s \in \mathcal{R}_{pq}$,

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

and

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

We next consider a steady-state network with queueing, in which a link flow is 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 (vertical) queue whenever the link is saturated. In the case of simple link flow capacities (when $g_k(f) = f_a - u_a$ for some nonnegative u_a), it is known that the value $t_a(f^*)$ may be interpreted as the travel time of the moving traffic stream, and the value of the multiplier term in (1) [which reduces to β_a^* in this case] as the link queueing delay. Further, by defining a flow to be a *queue equilibrium* whenever unsaturated links carry no queues, a solution f^* to the capacitated assignment problem may be given a complete characterization as a Wardrop equilibrium with respect to the generalized link travel costs and a queue equilibrium with respect to (f^*, β^*). (This is contained in works of Thompson and Payne and of Smith.) In the setting of more general (convex) side constraints, we assume that a side constraint may cause a queue which, in general, is physically distributed on all the links that are affected by the restriction, and hence consider *distributed* steady-state link queues. With this assumption we generalize the above classical result to characterize a solution to [TAP-VIP-SC-F] as a Wardrop equilibrium with respect to the generalized link travel costs (1) and a distributed queue equilibrium in terms of (f^*, β^*).

The notion of an equilibrium should be interpreted as a steady-state situation evolving after a disequilibrium state whose varying traffic flows eventually stabilize. During the process when the traffic flow gradually adjusts towards an equilibrium state, the travellers make route choice adjustments with the goal to minimize the sums of actual link travel costs and delays in distributed queues under the prevailing traffic conditions. Further, at any given moment a distributed queue will be building up or dissolving depending on whether or not the corresponding traffic flow restriction is violated by the current flow pattern; the generalized route travel costs and the route flows will of course also vary during the disequilibrium state. Under these assumptions on the dynamical behaviour of the queues we establish that a distributed queue equilibrium exists.

The queue equilibrium characterization made above can also be made in the application of traffic management schemes, where travellers adjust their route choices towards an equilibrium state while, simultaneously, link tolls are dynamically adjusted downwards or upwards depending on whether or not a traffic management goal is met (i.e., a side constraint is satisfied).

Our final results concern the use of existing algorithms that are available for the basic variational inequality problem when considering possible solution principles for the side constrained model [TAP-VIP-SC-F]. This leads us to consider pricing strategies, where we associate with the side constraints nonnegative prices β_k, $k \in \mathcal{K}$, for violating them. Given certain values of these prices, the side constraints are priced-out, i.e., handled implicitly in the following way. We consider the problem of finding a flow $f(\beta) \in F$ such that

[TAP-VIP-F, β]

$$[t(f(\beta)) + \nabla g(f(\beta))^{\mathrm{T}} \beta]^{\mathrm{T}} (f - f(\beta)) \geq 0, \qquad \forall f \in F,$$

which is a traffic equilibrium problem with link travel cost functions of the form (1) and which is solvable with standard methods for the user equilibrium model. When a vector of multipliers of the side constraints are at hand, the problem [TAP-VIP-F, β] is equivalent to the original problem, provided that the travel cost mapping t is strictly monotone. Multiplier values are found by solving the dual variational inequality of finding a $\beta^* \in \Re_+^{|\mathcal{K}|}$ such that

[DTAP-VIP-SC-F]

$$f(\beta^*)^{\mathrm{T}} (\beta - \beta^*) \geq 0, \qquad \forall \beta \in \Re_+^{|\mathcal{K}|},$$

where the mapping $\Re_+^{|\mathcal{K}|} \ni \beta \mapsto f(\beta)$ is defined through the solution of [TAP-VIP-F, β]. As the values of the prices tend to a vector of multipliers in the dual solution procedure, the primal solution $f(\beta)$ will simultaneously approach feasibility with respect to the side constraints and tend to a solution to [TAP-VIP-SC-F]. (This scheme may be viewed as a mathematical simulation of the above described dynamical process by which the proper link tolls are derived.)

The advantages of using nonlinear prices instead of linear ones as above, that is, to penalize a violation of a side constraint $k \in \mathcal{K}$ by means of a nonlinear function of g_k, are that a nonlinear penalty function enforces proximity to feasibility with respect to the side constraint stronger than a linear penalty function does and that the resulting algorithm may be convergent under weaker monotonicity properties of the link cost mapping. One such scheme is the (proximal) method of multipliers, in which the mapping $\bar{t}$ of [TAP-VIP-F, β] is replaced by

$$\tilde{t}(f) = t(f) + \sum_{k \in \mathcal{K}} \max\{0, \tilde{\beta} + \tilde{c} g_k(f)\} \nabla g_k(f) + \frac{\mu^2}{\tilde{c}}(f - \tilde{f}),$$

where $\tilde{c} > 0$ is a penalty parameter, $\mu \geq 0$ is a parameter used to control the nonlinearity of the mapping $\tilde{t}$, and $\tilde{f}$ and $\tilde{\beta}$ are current estimates of the link flow and dual solutions to [TAP-VIP-SC-F] respectively.

Suppose that a model is to be refined through the inclusion of some traffic flow relationships that are best represented *explicitly* by side constraints, but that one instead attempts to capture them *implicitly* through modifications of the travel cost functions, without first identifying the proper side constraints; we then notice the following: the proper travel cost mapping to be used has the form (1), this travel cost mapping involves gradients of the side constraint functions, which however are unknown since the side constraints have not been formulated explicitly, and the appropriate values of the parameters β are unknown, since the side constrained problem is not solved. We believe that these observations at least partially explain why traffic assignment models with complex travel cost functions may be difficult to calibrate. In contrast, the refinement of an assignment model through the inclusion of side constraints constitutes a very direct approach, which is also very flexible since the side constraints may be nonlinear as well as nonseparable; further, their physical interpretations may facilitate the estimation of the proper values of their coefficients. Hence, we believe that this approach is appealing from the practical viewpoint.