Algorithms for computing traffic equilibria

Michael Patriksson*

November 21, 2002

Abstract

This paper surveys the most basic traffic models based on the concept of traffic equilibria. It describes the most fruitful formulations that have been used, together with characterizations and properties of equilibria that can be derived from them. It further discusses the structural properties of these models that can be beneficially utilized, either directly or through model manipulations, in iterative algorithms, as well as some of the most successful such representatives. Extensions to traffic equilibrium models with side constraints and with non-additive route costs are provided.

Key words: User equilibrium, elastic demands, complementarity, decomposition, monotonicity, Lagrangian, merit function

1 Introduction

Wardrop's (1952) famous user equilibrium principle states that for every origin–destination (OD) pair $(p, q) \in \mathcal{C}$, the travel costs of the routes utilized are equal and minimal for each individual user. We denote by $\mathcal{R}_{pq}$ the set of simple (loop-free) routes for OD pair (p, q), by h_r the volume of traffic on route $r \in \mathcal{R}_{pq}$, and by c_r the travel cost on the route as experienced by an individual user. This condition can be written as follows:[1]

$$h_r > 0 \implies c_r = \pi_{pq}, \qquad r \in \mathcal{R}_{pq}, \qquad (p,q) \in \mathcal{C}, \tag{1a}$$
$$h_r = 0 \implies c_r \geq \pi_{pq}, \qquad r \in \mathcal{R}_{pq}, \qquad (p,q) \in \mathcal{C}, \tag{1b}$$

where the value of π_{pq} is the minimal (i.e., equilibrium) route cost in OD pair (p, q). This does however in general not describe an equilibrium, unless we introduce a means to describe the cost perception of the users given the volume of traffic, and the mechanism by which flow demand is generated. Let therefore $c : \Re^{|\mathcal{R}|}_{+} \mapsto \Re^{|\mathcal{R}|}$ be a vector-valued function of route costs. (Its argument is h.) Further, we assume that the demand is given by the vector-valued function $g : \Re^{|\mathcal{C}|} \mapsto \Re^{|\mathcal{C}|}_{+}$. (Its argument is π.) Under the condition posed on the non-negativity of the demands, the following is an equivalent formulation of (1) together with the demand constraints:

$$0^{|\mathcal{R}|} \leq h \perp (c(h) - \Gamma\pi) \geq 0^{|\mathcal{R}|}, \tag{2a}$$
$$\Gamma^{\mathrm{T}} h = g(\pi), \tag{2b}$$

where $a \perp b$, for two arbitrary vectors $a, b \in \Re^n$, means that $a^{\mathrm{T}} b = 0$, and where we introduce the matrix $\Gamma \in \Re^{|\mathcal{R}| \times |\mathcal{C}|}$ to be the route–OD pair incidence matrix (i.e., the element γ_{rk} is 1 if route r joins OD pair $k = (p, q) \in \mathcal{C}$, and 0 otherwise). In the two most standard special cases of this model, we can eliminate the π variables from the formulation. In the first, we consider the inelastic (fixed) demand model (set $g \equiv \bar{d} \in \Re^{|\mathcal{C}|}_{++}$), whence (2) is equivalent to the variational inequality problem (VIP):

$$-c(h) \in N_{\hat{H}}(h), \tag{3}$$

where, for any convex set $C \subseteq \Re^n$,

$$N_C(x) = \begin{cases} \{\, z \in \Re^n \mid z^{\mathrm{T}}(y - x) \leq 0,\ \forall y \in C \,\}, & x \in C, \\ \emptyset, & x \notin C, \end{cases}$$

*Department of Mathematics, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. E-mail: mipat@math.chalmers.se

[1] Most of the derivations in the paper can be found in Patriksson (1994, 2003).

denotes the *normal cone* to C at x (Rockafellar, 1970). Here, flow feasibility is described by the set $\hat{H} := \{\, h \in \Re_+^{|\mathcal{R}|} \mid \Gamma^{\mathrm{T}} h = \bar{d} \,\}$. In the second case, we assume that the demand function g is invertible, that is, that there exists a function $\omega : \Re^{|\mathcal{C}|} \mapsto \Re^{|\mathcal{C}|}$ with $d = g(\pi)$ if and only if $\pi = \omega(d)$. We then obtain the equivalent VIP:

$$[-c(h), \omega(d)] \in N_{\hat{H}_d}(h, d), \tag{4}$$

where $\hat{H}_d := \{\, (h, d) \in \Re_+^{|\mathcal{R}|} \times \Re^{|\mathcal{C}|} \mid \Gamma^{\mathrm{T}} h = d \,\}$. To see that they are equivalent, one can utilize the fact that the VIP is equivalent to the pair (h^*, d^*) solving the linear programming (LP) problem

$$\begin{aligned} \underset{(h,d)}{\text{minimize}}\ & c(h^*)^{\mathrm{T}} h - \omega(d^*)^{\mathrm{T}} d, \\ \text{subject to } & \Gamma^{\mathrm{T}} h - d = 0^{|\mathcal{C}|}, \\ & h \geq 0^{|\mathcal{R}|}, \end{aligned}$$

to which π is the LP dual solution. In the case where there is an upper bound on g, this value, say $\bar{g}$, can be used in an equivalent inelastic demand reformulation of the model (4), where for each OD pair $(p, q) \in \mathcal{C}$, an additional link (p, q) is introduced; with e_{pq} being the flow on this link, its cost is $\omega_{pq}(\bar{g} - e)$, $e \in \Re^{|\mathcal{C}|}$, and the inelastic demand vector for this new network is $\bar{g}$. At equilibrium, the volume on the links e_{pq} are the excess demands that will not be observed in the real network. (See Patriksson, 2003.)

If we wish to model classes of vehicles or users of different types interacting in the same network, we can simply do this in the above framework by letting each class have its own network copy; a link will therefore have several copies, and the cost interaction among the users of that link is modelled through the individual user class costs. In this way, it thus turns out that multi-class user models are really no more general than single-class models, provided of course that we allow the link travel costs to be non-separable.

Two further specializations are of interest to introduce immediately. The first is that of additive route costs, which will enable us to work with link-based models and algorithms. The second is that the travel cost and demand functions are "symmetric," so that we can solve for an equilibrium with algorithms for nonlinear optimization.

Let $\Lambda \in \{0, 1\}^{|\mathcal{A}| \times |\mathcal{R}|}$ be the link–route incidence matrix, whose element λ_{ar} equals one if route $r \in \mathcal{R}$ utilizes link $a \in \mathcal{A}$, and zero otherwise. Route r has an additive route cost $c_r(h)$ if it is the sum of the costs of using all the links defining it. In other words, $c_r(h) = \sum_{a \in \mathcal{A}} \lambda_{ar} t_a(v)$, where $v \in \Re^{|\mathcal{A}|}$ is the total volume of traffic on the links. Also, implicit in this relationship is the assumption that the pair (h, v) is consistent, in the sense that v equals the sum of the route flows: $v = \Lambda h$. The model (3) can then equivalently be written as the link-based model

$$-t(v) \in N_{\hat{F}}(v), \tag{VIP-$\hat{F}$}$$

where $\hat{F} := \{\, v \in \Re^{|\mathcal{A}|} \mid \exists h \in \hat{H} \text{ with } v = \Lambda h \,\}$ is the aggregate of H. Correspondingly, the model (4) can be written as

$$[-t(v), \omega(d)] \in N_{\hat{F}_d}(v, d), \tag{VIP-$\hat{F}_d$}$$

where $\hat{F}_d := \{\, (v, d) \in \Re^{|\mathcal{A}|} \times \Re^{|\mathcal{C}|} \mid \exists (h, d) \in H_d \text{ with } v = \Lambda h \,\}$.

In the case where t is integrable,[2] the model [VIP-$\hat{F}$] defines the first-order optimality conditions for an optimization problem; assuming, further, that t is separable, so that t_a is a function only of v_a, $a \in \mathcal{A}$, the optimization problem has the form

$$\underset{v}{\text{minimize}} \sum_{a \in \mathcal{A}} \int_0^{v_a} t_a(s)\, ds, \tag{5a}$$

$$\text{subject to } \Gamma^{\mathrm{T}} h = \bar{d}, \tag{5b}$$

$$h \geq 0^{|\mathcal{R}|}, \tag{5c}$$

$$v = \Lambda h. \tag{5d}$$

[2] If t is continuously differentiable, then integrability is equivalent to the symmetry of its Jacobian matrix $\nabla t(v)$ everywhere. Integrability is a more general property than this symmetry property, since t need not be always be differentiable; nevertheless, the term "symmetry" is always used in traffic planning when it should really refer to the integrability property.

When demand is elastic and separable, according to [VIP-$\hat{F}_d$] we have the equivalent optimization problem to

$$\underset{(v,d)}{\text{minimize}} \sum_{a\in\mathcal{A}} \int_0^{v_a} t_a(s)\,ds - \sum_{(p,q)\in\mathcal{C}} \int_0^{d_{pq}} \omega(s)\,ds, \tag{6a}$$

$$\text{subject to } \Gamma^{\mathrm{T}} h - d = 0, \tag{6b}$$

$$h \geq 0^{|\mathcal{R}|}, \tag{6c}$$

$$v = \Lambda h. \tag{6d}$$

We may also formulate the Wardrop conditions in terms of link flows only. We introduce the link–node incidence matrix, $E \in \{-1,0,1\}^{|\mathcal{N}|\times|\mathcal{A}|}$, whose element e_{ia} equals -1 if node i is the origin node of link a, 1 if node i is the destination node of link a, and 0 otherwise. This link–node version of Wardrop's equilibrium conditions states that at an equilibrium link flow v, being the aggregate of commodity (OD pair) volumes $w_k \in \Re^{|\mathcal{A}|}$, $k := (p,q) \in \mathcal{C}$, there exists a vector $\pi_k \in \Re^{|\mathcal{N}|}$ of node prices such that

$$0^{|\mathcal{A}|} \leq w_k \perp \left(t(v) - E^{\mathrm{T}}\pi_k\right) \geq 0^{|\mathcal{A}|}, \qquad k \in C; \tag{7}$$

for a link $(i,j) \in \mathcal{A}$, this means that $0 \leq w_{ijk} \perp (t_{ij}(v) - [\pi_{jk} - \pi_{ik}]) \geq 0$. The set of feasible volumes is, for inelastic demands, the set of vectors $w_k \geq 0^{|\mathcal{A}|}$ for which $Ew_k = i_k\bar{d}_k$ holds, where $\bar{d}_k$ is the fixed demand in the OD pair.[3] In the case of elastic demands, the fixed right-hand side is replaced by the demand function, g_k, defined at the value of the vectors of least-costs π_k, $k \in \mathcal{C}$.

In the same two cases as above, then, we have two variational problems of the same form as [VIP-$\hat{F}$] and [VIP-$\hat{F}_d$], respectively, in the inelastic and elastic cases. Here, however, $\hat{F}$ is replaced by $F := \{\, v \in \Re^{|\mathcal{A}|} \mid \exists w_k \in \Re_+^{|\mathcal{A}|} \text{ with } Ew_k = i_k\bar{d}_k,\ k \in \mathcal{C}, \text{ and } v_a = \sum_{k\in\mathcal{C}} w_{ak},\ a \in \mathcal{A}\,\}$. Again, we can use LP duality to show that (7) are the complementarity conditions of a primal–dual pair of LPs corresponding to [VIP-F]. Further, $\hat{F}_d$ is replaced by $F_d := \{\, (v,d) \in \Re^{|\mathcal{A}|} \times \Re^{|\mathcal{C}|} \mid \exists w_k \in \Re_+^{|\mathcal{A}|} \text{ with } Ew_k = i_k d_k,\ k \in \mathcal{C}, \text{ and } v_a = \sum_{k\in\mathcal{C}} w_{ak},\ a \in \mathcal{A}\,\}$. The two optimization link–route formulations (5) and (6) have their corresponding link–node formulations.

The following section discusses the existence and uniqueness of equilibria. Section 3 describes the structures in the above models that can and should be utilized when devising algorithms. Section 4 surveys some classic and more recent algorithms for equilibrium models. The final section extends the notion of an equilibrium to the case of additional, side, constraints, and discusses a non-additive route cost model.

2 Properties of equilibria

We assume throughout that the functions c, t, g, and ω, whenever present, are single-valued and continuous. The nonnegativity assumption on g implies that the model [VIP-$\hat{F}_d$] has at least one solution when, in addition, g is upper bounded; this follows from the convexity and compactness of the set of possible flows, costs and demands. In consequence, the continuity of the cost function c is enough for the existence of solutions to the model (3). Likewise, then, the two optimization formulations (5) and (6) have optimal solutions, when in addition the link cost and inverse demand functions are integrable. The case of the link–node formulation is slightly different; the difference to the link–route case is that cyclic flows are included in F and F_d, and these sets are therefore not bounded. Under the assumption that t is nonnegative along any directed cycle however has the implication that a cycle is either not used in equilibrium, since it would make any route that uses it more expensive, or any cyclic flow of zero cost can be eliminated. So, in the sense of optimality, the link–route and link–node forms are equivalent when costs are cycle-wise nonnegative, and the links used in equilibrium from a given origin will/can therefore describe an acyclic network. In the following in this section, we will therefore not always separate the discussion between the two formulations.

The difference between the existence of solutions between VIP and optimization models is slight (the only one refers to costs and demands being integrable), and so is the case with the difference that is

[3] The vector $i_k \in \{-1,0,1\}^{|\mathcal{N}|}$ is an indicator vector which is zero in all positions but two where, by the sign convention introduced earlier, the element with value 1 (-1) corresponds to the sink (source) node. The vector i_k also relates the OD node price vectors $\pi_k \in \Re^{|\mathcal{N}|}$ and the OD least cost vector $\pi \in \Re^{|\mathcal{C}|}$ through the relation $\pi = (\pi_{pq})_{(p,q)\in\mathcal{C}} = (i_k^{\mathrm{T}}\pi_k)_{(k\in\mathcal{C})}$.

present when we consider uniqueness issues. Beginning with the simpler models, the two optimization problems (5) and (6) are convex whenever t and $-\omega$ are monotone (each component then is non-decreasing in its variable). In the problem (5), this is enough for the costs $c(h)$ and $t(v)$ to be unique, hence also the vector π of least costs; this is a property of every differentiable, convex program (Burke and Ferris, 1991). The total link volume v is unique whenever t is strictly monotone (each component strictly monotonically increasing), as this is equivalent to the integral being strictly convex. For the problem (6), the demand d and vector π is unique whenever $-g$ and t are monotone, and, as above, v is also unique when t is strictly monotone as well. In both these problems, uniqueness of the route flow h and commodity link flows w_k cannot follow from making any monotonicity assumptions on these functions: in most networks, there are route overlaps, making it possible to disaggregate a (unique) link volume v into its commodity components h or w_k, $k \in \mathcal{C}$, in an infinite number of ways, and since the route costs are additive by assumption, their costs will be the same.

In the general, non-separable case, most of the properties mentioned still follow, replacing arguments based on (strict) convexity with those of (strict) monotonicity. However, it is not true that the cost mapping of an asymmetric but monotone VIP is unique; in order to establish that c, t and π are unique on the set of equilibria, a slightly stronger property must be introduced, at least in the elastic demand case.[4] Working with nonmonotone cost models must be done with great care, as there may be isolated equilibria, some of which will also be unstable to small changes in the data; if we are not content with just any of them but are interested, for example, in the most stable one (measured, for example, by a function of the real part of the eigenvalues of the link cost Jacobian), then we must combine the equilibrium model with a mechanism for finding a unique solution, a subject which borders on finding the global solution to a nonconvex—perhaps even bilevel—optimization problem (see Patriksson and Rockafellar, 2002, on the subject of bilevel optimization in transportation science).

3 Model structures and the verification of equilibria

Some favourable problem structures are observed when looking at the models (5) and (6):

OD pair separability The feasibility of the demand and traffic volume in one OD pair does not affect that of another. The problems indeed are multi-commodity flow problems, where there are no side constraints acting on more than one commodity, such as link flow capacity constraints. (This property holds for more general cost and demand models as well.)

Cost separability The objective function, as well as the link cost and demand functions, is separable in the link volume and demand variables. (This property does not extend to more general cost and demand models.)

Primal–dual relations In both the link–route and link–node representations, the vector π is a Lagrange multiplier vector, which further measures the least-cost. (This property holds for more general cost and demand models as well.)

The link–route representation has a simpler constraint structure than the link–node one; the constraints in the former describe a simplex, while the latter describe flow conservation. The former, on the other hand, has an exponential number of variables, which must be enumerated iteratively, while the latter has a polynomial number of variables, and a richer constraint structure.

Compared to most VIP and optimization problems, the inherent network structure also provides efficient tests for equilibrium. Keeping the demand fixed, checking if an equilibrium volume has been found is easy through the linear program defining the VIP: at $(\bar{h}, \bar{d}) \in \hat{H}_d$, we define

$$\psi(\bar{h}) := \underset{(h,\bar{d}) \in \hat{H}_d}{\text{minimum}}\; c(\bar{h})^{\mathrm{T}}(h - \bar{h}) = \sum_{k \in C} \underset{(h_k, \bar{d}_k) \in \hat{H}_{k,d}}{\text{minimum}}\; c_k(\bar{h})^{\mathrm{T}}(h_k - \bar{h}_k). \tag{8}$$

(Here, the index k in the second equality designates feasibility sets and sub-vectors for individual OD pairs.) This LP problem defines a set of shortest route problems; if and only if the total cost of transportation along the shortest routes in each OD pair equals the total travel cost of the current OD flow,

[4] In the case of fixed demand models, it is in fact enough that the route or link cost function is pseudo-monotone$^{+}_{*}$, a stronger form of pseudo-monotonicity, cf. Crouzeix et al. (2000) for the definition and Patriksson (2003) for the result; in the case of elastic demands, the route or link cost function must have the stronger property of monotonicity^{+}, also known as paramonotonicity, a property which however reduces to monotonicity whenever the function in question is a gradient mapping, cf. Iusem (1998); alternatively, we can take $-g$ to be monotone and t to be strictly monotone.

then an equilibrium is reached. The value of ψ is always non-positive, and an equilibrium is equivalent to a zero value. (Otherwise, we have identified a set of routes that should be utilized more.) With fixed link costs, checking if the demand is correct requires only the calculation of the demand function. It would seem natural perhaps to use the iteration $d \mapsto g(\pi(t(v(d))))$ to reach the set of equilibria for elastic demand models. It takes the current demand and calculates a corresponding inelastic demand solution, from which the least route cost is calculated; the value of the demand function gives the new demand estimate. For this type of fixed-point type of iteration to converge it is required that the iteration is a contraction, or at least a non-expansion, that is, that, for an equilibrium demand d^* and any feasible demand vector d, $\|g(\pi(t(v(d)))) - d^*\| \leq \|d - d^*\|$ holds (see, e.g., Dunn, 1973, and Baillon, 1975, for basic fixed-point methods). Given the complicated form of this mapping, whether the contraction condition is satisfied is difficult to analyze from the original data, and updating rules for flows and demands are instead based on approximations of one or more of these mappings, examples of which will be given below.

4 Basic iterative algorithms

4.1 Decomposition

The separability of the OD pairs (or, user classes, etc.) in the constraints suggests the use of a decomposition algorithm, wherein flow in one type of flow is updated upon while keeping the others fixed. The remaining problem is then a single-commodity network flow problem. Among the advantages of such a scheme we mention first that the link cost function has stronger monotonicity properties when viewed on this smaller subspace than on the entire space of flows; for example, if $v_a \mapsto t_a(v_a)$ is strictly monotonically increasing, then it is non-strictly monotone in the vector of OD link flows w_k, $k \in \mathcal{C}$, but it is strictly monotonically increasing when keeping the flows in all but one OD pair fixed. Second, several single-commodity network flow algorithms are available (see, e.g., Patriksson, 1994, Section 4.3), among which are several (potentially) super-linearly convergent ones. Third, algorithms of this type can be viewed as block versions of Gauss–Seidel (or Jacobi) algorithms, which have a sound convergence theory. They can however be slow when the number of blocks is large, since too much of the cost interaction among the commodities is lost. An advantage of the problem structure then is that one can create larger and fewer blocks of variables such that, for example, all the flow from one given origin is updated upon at any one time. Contrary to the use of decomposition over OD pairs or origins, a decomposition over links is less often useful, as costs are not always defined by separable functions.

A natural means to decompose the problem over OD pairs is to temporarily fix the cost on the links, thereby removing the dependence between OD pairs in the cost. In the fixed demand problem, what is left is a number of shortest route problems, on which the fixed demands are assigned; in the elastic demand case, the demand to be assigned onto the shortest routes are given by the demand function at the shortest route costs. (This way, we will later identify the classic Frank–Wolfe and Evans' algorithms.)

In order to take advantage of the link cost separability in the models (5) and (6), we must eliminate their dependence in the constraints by introducing some form of relaxation. One possibility is to introduce explicit multipliers, $\nu \in \Re^{|\mathcal{A}|}$, for the constraints "$v = \sum_{k \in \mathcal{C}} w_k$." With the (redundant) constraints $v \in \Re^{|\mathcal{A}|}_+$ added to the Lagrangean relaxed problem, we obtain a problem, equivalent to [VIP-F_d], which amounts to finding (v, w, d, π, ν) such that

$$0^{|\mathcal{A}|} \leq v \perp (t(v) - \nu) \geq 0^{|\mathcal{A}|}, \tag{9a}$$

$$Ew_k = i_k d_k, \qquad k \in \mathcal{C}, \tag{9b}$$

$$0^{|\mathcal{A}|} \leq w_k \perp \left(\nu - E^{\mathrm{T}} \pi_k\right) \geq 0^{|\mathcal{A}|}, \qquad k \in \mathcal{C}, \tag{9c}$$

$$\omega(d) = \pi, \tag{9d}$$

$$v - \sum_{k \in \mathcal{C}} w_k = 0^{|\mathcal{A}|} \tag{9e}$$

holds. Fixing $\bar{\nu} \geq t(0)$, and noting that (9e) is the dual optimality condition, the problem left separates into one over v and w_k, $k \in \mathcal{C}$. The first, (9a), in the separable case further decomposes into a set of one-variable problems with the objective $\int_0^{v_a} (t_a(s) - \bar{\nu}_a)\, ds$, to be minimized over $\Re_+$, with a solution given by the inverse of t_a evaluated at $\bar{\nu}_a$, if non-negative, otherwise zero; in the general case, this is a nonlinear complementarity problem (NCP, see, e.g., Ferris and Pang, 1997). The second problem, (9b)–(9d), amounts to finding, for each $k \in \mathcal{C}$, the shortest route based on the vector $\bar{\nu}$ of link costs. The shortest route costs are used to calculate the demand through the demand function; the demand is distributed onto the shortest routes to form the vectors w_k. A step is then made in ν in order to produce a result that better fulfils the last constraint. Although being a primal–dual algorithm, the aggregate of the w_k is a feasible link flow; this is a consequence of the fact that the dualization is made on the definitional constraint (9e). (See Larsson et al., 1997.)

Another means to decompose the problem is to instead relax the other linear equality constraints, which define the network, that is, (9b). Introducing explicit multipliers $\pi_k \in \Re^{|\mathcal{N}|}$ for each commodity $k \in \mathcal{C}$, we obtain a problem, equivalent to [VIP-$\hat{F}_d$], which amounts to finding (v, w, d, π) such that

$$0^{|\mathcal{A}|} \leq w_k \perp \left(t(v) - E^{\mathrm{T}}\pi_k\right) \geq 0^{|\mathcal{A}|}, \qquad k \in \mathcal{C}, \tag{10a}$$

$$Ew_k = i_k d_k, \qquad k \in \mathcal{C}, \tag{10b}$$

$$\omega(d) = \pi, \tag{10c}$$

$$v - \sum_{k \in \mathcal{C}} w_k = 0^{|\mathcal{A}|} \tag{10d}$$

holds. Fixing $\bar{\pi}_k$ (and letting $\bar{\pi}_{pq} := i_k^{\mathrm{T}}\bar{\pi}_k$), and noting that (10b) is the dual optimality condition, the remaining problem is solved as follows. The demand d is given, through (10c), by the demand function at $\bar{\pi}$. The complementarity system (10a) is solved, in the separable case, by first defining, for each link $a = (i, j) \in \mathcal{A}$, $\bar{p}_{ij}(\pi) := \text{maximum}_{k \in \mathcal{C}}\, \{\pi_{jk} - \pi_{ik}\}$ to be the largest node potential difference among the commodities. Let $\mathcal{C}_{ij}(\pi) \subseteq \mathcal{C}$ denote the set of commodities for which $\pi_{jk} - \pi_{ik} = \bar{p}_{ij}(\pi)$ holds. Then, $v_{ij}(\pi) := \text{maximum}\, \{0, t_{ij}^{-1}(\bar{p}_{ij}(\pi))\}$ is the total link flow in link (i, j). The commodity link flows $w_{ij,k}$, $k \in \mathcal{C}$, can be taken as any distribution of the total flow $v_{ij}(\pi)$ onto the commodities in $\mathcal{C}_{ij}(\pi)$. It remains to define an updating rule for the vectors $\bar{\pi}_k$ such that we approach also a solution to (10b). If this Lagrangian relaxation is utilized within a cyclic decomposition algorithm over OD pairs, where the subproblems are single-commodity network flow problems, then even more efficient dual algorithms are available, since the duals then are differentiable (see, e.g., Patriksson, 1994, Section 4.3.2).

A decomposition (relaxation) is also reached by manipulating the costs. Suppose, for example, that the link costs are fixed at their current values, and consider the resulting relaxation. Taking [VIP-$\hat{F}_d$] as the model for discussion, and the iterate $(v^\tau, d^\tau, (\pi_k)_{k \in \mathcal{C}}^\tau)$ at iteration τ, fixing the link costs to $t(v^\tau)$ yields the following procedure for solving this relaxation of [VIP-$\hat{F}_d$]: first, for each $k = (p, q) \in \mathcal{C}$, calculate a least-cost route, cf. (7), whence the vectors $\bar{\pi}_k^\tau$, $k \in \mathcal{C}$, $(\bar{\pi}^\tau)_{(p,q) \in \mathcal{C}} = (i_k^{\mathrm{T}}\bar{\pi}_k^\tau)_{k \in \mathcal{C}}$ are given; second, calculate the demand through $\bar{d}^\tau := g(\bar{\pi}^\tau)$; third, this demand is assigned onto each OD pair's shortest route, giving rise to the link flows $\bar{w}_k^\tau$, $k \in \mathcal{C}$, and $\bar{v}^\tau = \sum_{k \in \mathcal{C}} \bar{w}_k^\tau$. If link costs and demands are separable, we can perform a line search in the direction of $(\bar{v}^\tau, \bar{d}^\tau, (\bar{\pi}_k)_{k \in \mathcal{C}}^\tau) - (v^\tau, d^\tau, (\pi_k)_{k \in \mathcal{C}}^\tau)$, whence we have defined a partial linearization algorithm, Evans' (1976) algorithm; if further demand is inelastic, this is precisely the Frank–Wolfe algorithm (e.g., LeBlanc et al., 1975). The case of nonseparable functions is much more complicated, and deserves its own subsection below. We first however discuss an interesting improvement over the above algorithm.

The route information generated during the process of calculating shortest routes is badly utilized in line search algorithms, considering the effort in generating it. If, instead, we were to store the routes that carry flow or that are new and promising, and solve the equilibrium problem over those, we would in fact be attacking a restriction to the link–route formulation of the original model (a restricted master problem). As we have remarked earlier, this model form has an advantage in that the constraint structure can be effectively utilized; if we keep down the size of the route set, we also have a chance of avoiding the drawback of this formulation. Consider then the model [VIP-$\hat{F}$], where we have available a restriction $\hat{\mathcal{R}} \subset \mathcal{R}$ of the route set. We recognize that the simplex form can be utilized very well, if we were to devise a decomposition over OD pairs or origins, for example. We draw an example by considering a kind of cyclic Newton algorithm, wherein we keep the flow from all but one group of users fixed, and then utilize a diagonalized second-order approximation of the remaining objective function (Bertsekas, 1976). At a given iteration where the flow is $\hat{h}^\tau$, we would, for one OD pair k, say, solve the following problem:

$$\underset{\hat{h}_k}{\text{minimize}} \sum_{r\in\hat{\mathcal{R}}_k} \left\{ c_r(\hat{h}^\tau)\hat{h}_r + \frac{b_r^\tau}{2\gamma_\tau}\hat{h}_r^2 \right\}, \tag{11a}$$

$$\text{subject to} \sum_{r\in\hat{\mathcal{R}}_k} \hat{h}_r = \bar{d}_k, \tag{11b}$$

$$\hat{h}_k \geq 0^{|\hat{\mathcal{R}}_k|}, \tag{11c}$$

where $0 < b_r^\tau \approx \frac{\partial c_r(\hat{h}^\tau)}{\partial \hat{h}_r}$, and the value of $\gamma_\tau > 0$ is chosen such that the improvement made is the best possible; the resulting solution is taken to be $\hat{h}_k^{\tau+1}$. The above problem has a low complexity: solving for the optimal Lagrange multiplier for the equation can be done in $O(|\hat{\mathcal{R}}_k|)$ time, that is, linear in the number of routes (Brucker, 1984). Reaching subproblems of low complexity is essential in an iterative algorithm, as the size of traffic problems can be very large.

A final remark on algorithms for the basic models is that it seems difficult to devise truly super-linearly convergent algorithms that utilize problem structure very efficiently. One main reason is the presence of multiple commodities; for separable cost nonlinear single-commodity flow problems, super-linearly convergent algorithms are available, which also utilize the network property in an efficient way; network-utilizing (quasi-)Newton methods can be found, for example, in Klincewicz (1983) and Escudero (1986); see also Best and Griffin (1975) and Gafni and Bertsekas (1984). But the Gauss–Seidel algorithm—which is needed to coordinate the search in the multi-commodity setting—is a linearly convergent algorithm, at best (cf. Bertsekas and Tsitsiklis, 1989, for a detailed analysis). There is therefore still plenty of room for algorithmic improvements, in particular the search for new and clever ways of decomposing and coordinating the problem.

4.2 Merit functions and convergence

When solving separable models, we have access to a natural merit function [(5a) or (6a)], with which we can measure the progress of an algorithm and select reasonable step lengths in a line search; in fact, the whole area of nonlinear optimization is open to exploration. When turning to non-separable models, these merit functions are no longer well-defined, and distance measures from an iterate to an equilibrium must be replaced by artificially constructed functions. One such function is the primal gap function (Zuhovickiĭ et al., 1969; Patriksson, 1994), given in (8), which also can be calculated entirely in link flows. As with most other merit functions used for variational inequality problems, its value requires the solution of an approximation of the original problem. The computation of equilibria for non-separable models are in fact much more complex than for separable ones:

Decomposition In the separable case, some of the most natural decompositions lead to low-complexity subproblems. In the non-separable case, however, they either become unavailable altogether, or they become more complex. For example, a decomposition over links will no longer be natural or even possible. Further, if the form of the cost and demand is complex, it also means that computing their values and derivatives will be more complex, and this will also affect the computational complexity of an algorithm.

Search direction → merit function In the separable case, virtually every algorithm proposed is a feasible direction-based descent one from nonlinear optimization which utilizes the natural objective (5a) or (6a); cf. Patriksson (1994, Section 4.3). Establishing convergence of such algorithms is also relatively straightforward, as it is easy to satisfy the descent condition that the negative of the gradient of the merit function makes an acute angle with the search direction, and it relies only mildly on the monotonicity properties of the cost and demand functions. In the non-separable case, however, the angle condition becomes much more difficult to establish for a suitable merit function. The merit function is intimately associated with the search direction, and in some cases it is even defined by the solution of the search direction-finding problem. That also means that, in contrast to the separable case, the merit function and its gradient depends on the search direction chosen. Convergence requires a lot more from the monotonicity properties of the cost and demand functions in the more general case, as well as the choice of step lengths, cf. Patriksson (1998).

As an example of the complications that arise in the general, non-separable case, compared to the separable one, we mention that the Frank–Wolfe algorithm is not convergent for non-separable, asymmetric cost models (see, e.g., Hammond, 1984); no-one has yet found a valid merit function which descends with an iteration of the Frank–Wolfe algorithm.

We have concluded that the standard algorithms for the separable case do not favour well to be extended to the non-separable case. There are however variational inequality algorithms that so-far have not reached the attention of the traffic research community, and which are convergent under rather mild assumptions on the travel cost and demand mappings. We name but three: the gradient projection algorithm can be supplied with special line searches which make it convergent for monotone cost mappings (Bruck, Jr., 1977, Patriksson, 1998); the class of extra-gradient algorithms involves two projections, and is convergent under even milder monotonicity requirements, pseudo-monotonicity (Wang et al., 2001); coupled together with a proximal point algorithm (Rockafellar, 1976), also algorithms that, when working on their own, converge only under strong monotonicity[5] assumptions, will converge under a monotonicity requirement. The most natural environment for these algorithms, at least the first two, is the singly constrained restricted master problem in a route (column) generation algorithm, that is, a restriction of the set defined by (5b)–(5c) or (6b)–(6c) to a subset $\hat{\mathcal{R}}$ of the routes. The reason why we prefer this environment is that projections are easy to perform onto such sets, cf. (11). The third algorithm would be defined thus: for a given cost function in the variational inequality, one would add a multiple ($\gamma_\tau > 0$) times an affine cost of the form $v \mapsto v - v^\tau$, if we work in link space. The resulting mapping is strongly monotone with a modulus of at least γ_τ, provided that the original cost function is monotone. Therefore, if we have an algorithm which requires strong monotonicity, we apply it instead to this perturbed problem. At the (approximate) solution to the perturbed problem, we again perturb the original function at the new iterate, $v^{\tau+1}$, and proceed, perhaps with a different value of γ.

Diagonalization (Sheffi, 1985; Patriksson, 1994) is a favourite among heuristics for non-separable problems. It amounts to, at an iteration, temporarily removing the dependency of a link's cost on the flow in other links. This is a kind of Jacobi algorithm, and may suffer from quite poor convergence characteristics: if the correlation in reality is quite large, one can expect the algorithm to diverge, or at least converge slowly; on the other hand, if the cost dependence between links is mild, or local to a few links only, then convergence can be expected to be fast.

4.3 Implementations and computational experience

The commercially most successful implementation of an equilibrium solver is EMME/2, which uses the Frank–Wolfe algorithm. Although it has been known for quite some time that the Frank–Wolfe algorithm has the worst convergence characteristics among equilibrium solvers, its relatively high initial efficiency, small memory requirements, and the natural interpretation behind its workings—it is reminiscent of heuristics that were devised already in the mid 1960's—has made it popular. Its counterpart for elastic demand models is Evans' algorithm, which has similar convergence characteristics: the demand quickly converges, and once the demand no longer changes much, the algorithm behaves like the Frank–Wolfe algorithm.

Among the best extensions of the Frank–Wolfe algorithm is the class of simplicial decomposition/column generation algorithms (e.g., Larsson and Patriksson, 1992) which are based on shortest route subproblems. While the simplicity of the subproblem is kept, the results of the computations are better utilized, as the shortest routes are kept in memory and the demand is optimally assigned onto them. This restricted master problem, which has the same form as that of the original problem except that the set $\mathcal{R}$ is replaced by the known subset $\hat{\mathcal{R}}$, is then solved using a specialized, perhaps second-order, method. In later implementations of the DSD algorithm (Larsson and Patriksson, 1992), a diagonalized Newton algorithm is applied, which also efficiently detects routes with zero flow and which may be removed from the set $\hat{\mathcal{R}}$ in order to save space. Other implementations use different versions of reduced gradient and gradient projection methods (e.g., Chen et al., 2002). As projection methods are available also for non-separable models, it is anticipated that the best versions can be transferred to solve also such more general models.

A recent contribution is the quasi-Newton algorithm of Bar-Gera (2002a, 2002b). It utilizes the fact that the equilibrium flow from a given origin is acyclic, and tries in a combinatorial fashion to select the correct acyclic networks. For each origin and acyclic network, a quasi-Newton algorithm is implemented for finding the best single-commodity flow, based not only on the flow on the links but also on the flow through the nodes. The numerical experience so-far is impressive, and it is believed that it will, if not

[5]A mapping $F : \Re^n \mapsto \Re^n$ is strongly monotone on $\Re^n$ if there exists an $m > 0$ (the modulus of strong monotonicity) such that $[F(x) - F(y)]^{\mathrm{T}}(x - y) \geq m\|x - y\|^2$ holds for all $x, y \in \Re^n$.

replace, then at least become an alternative to the state-of-the art algorithms used in practice. The drawback of this algorithm is that it has not been shown how to make it work for non-separable models.

Perhaps somewhat surprisingly, the field of computational testing, in particular comparative ones, has remained rather unexplored; see, however, the two recent web sites

`http://www.bgu.ac.il/~bargera/tntp/`

and

`http://www-rocq.inria.fr/scilab/CiudadSim`

for some test problems and computational results. Comparisons are still quite often performed against the Frank–Wolfe algorithm; such results are not very enlightening, given the bad convergence behaviour of the latter. Several interesting alternatives to the classic algorithms have never been tested, and those that are are quite often tested in a way that would not pass the scrutiny of a numerical analysis specialist. The latter is not so surprising, considering that most of the algorithmic development is made within applications to specific problems. The former is not so surprising either, given that most researchers in the field come from an area different from mathematical programming.

Important test objects are also the low-complexity subproblems, like the shortest route problem and the quadratic knapsack problem (11): although there are algorithms for these problems which have the lowest worst-case complexity, the fact that these problems must be solved repeatedly makes it possible that algorithms with a higher worst-case complexity are better to use in practice, if they have a much better reoptimization capability; sorting is a famous such example.

5 Extensions

5.1 Side constraints

Suppose the feasible set defining the traffic flow polyhedron is further constrained, for example by link flow capacities arising from a stationary representation of a traffic control system or from bounds on the quality of service in a telecommunication network. We define convex and differentiable functions $g_k(v)$, $k \in \mathcal{K}$, requiring in addition that (Larsson and Patriksson, 1994, 1998, 1999)

$$g_k(v) \le 0, \qquad k \in \mathcal{K} \tag{12}$$

holds. Flow capacities on some links then correspond to letting $g_a(v) := v_a - c_a$, $a \in \bar{\mathcal{A}} \subseteq \mathcal{A}$, where $c_a > 0$ is the link volume capacity. With this addition, it is clear that the Cartesian product structure in the feasible sets of, for example, (5) and (6), is lost. This will have rather far-reaching effects on the equilibrium characterization of an optimal flow. An equilibrium, in the sense of Wardrop, does no longer exist in terms of the original costs. If we however introduce Lagrange multipliers for the side constraints (12), then we can describe an equilibrium in terms of a generalized cost which, under some circumstances, can be given a natural interpretation. With multipliers $\beta_k \ge 0$, the new optimality condition is the complementarity condition that

$$0^{|\mathcal{K}|} \le \beta \perp g(v) \le 0^{|\mathcal{K}|}. \tag{13}$$

Letting $G := \{\, v \in \Re^{|\mathcal{A}|} \mid g_k(v) \le 0, \quad k \in \mathcal{K} \,\}$, the side constrained inelastic demand model becomes

$$-t(v) \in N_{F\cap G}(v) = N_F(v) + N_G(v) = N_F(v) + \nu, \quad \nu \in N_G(v);$$

if G satisfies a constraint qualification (CQ; see Bazaraa et al., 1993, Chapter 5), for example, that each function g_k is affine, then for some vector β satisfying (13), we have that, at equilibrium,

$$-\left[t(v) + \sum_{k\in\mathcal{K}} \beta_k \nabla g_k(v)\right] \in N_F(v)$$

holds. Hence, for any optimal vector of Lagrange multipliers, v is a Wardrop equilibrium volume in terms of the generalized link costs $\bar{t}(v) := t(v) + \sum_{k\in\mathcal{K}} \beta_k \nabla g_k(v)$.

In the case of link capacities, this simply becomes $\bar{t}(v) := t(v) + \beta$ (Hearn, 1980), where the elements of β for $a \notin \bar{\mathcal{A}}$ are fixed to 0. This expression has been interpreted as an equilibrium link queueing delay, where β is a vector of additional delays at intersections. This interpretation may sometimes be valid, but it is important to note that the vector β of Lagrange multipliers very seldom can be guaranteed to be unique, even in this simple case, which means that the link queueing delay is not unique, even if the link flow v is. (See Larsson and Patriksson, 1994, 1998, 1999, for analyses and examples.)

The above development suggests a natural algorithm for side constrained traffic equilibrium models: guess a value of the multiplier vector β, solve a standard traffic equilibrium model in terms of generalized link costs, and update the guess on β based on a convergent algorithm for solving the Lagrangian dual problem. (See Larsson and Patriksson, 1995, for an example augmented Lagrangian algorithm for the link capacitated case.) Whether it will be successful is determined in part by the number of side constraints, and the convexity and separability properties of the functions g_k.

5.2 Non-additive route costs

In the above development, we have assumed that the route cost is additive, that is, that $c_r(h) = \sum_{a\in\mathcal{A}} \lambda_{ar} t_a(v)$ holds for all $r \in \mathcal{R}$ and consistent flows (h, v). This is not true, for example, in the logit-based stochastic user equilibrium model (Fisk, 1980), nor is it true if a route has a route-specific monetary cost associated with it, such as a route-based toll, or if we want c_r to include the cost of emissions. Cost additivity also fails in loss networks arising in, for example, telecommunication systems, cf. Altman et al. (2002).

We consider a particular case of a non-additive route cost model with two attributes: time and money.

Although route cost usually has several attributes (time, monetary outlays, distance, etc.), they are usually linearly combined into one generalized cost. Empirical studies however show that the value of time can be nonlinear. Consider, for example, that $c_r(h) = f_r + \varphi_r(\sum_{a\in\mathcal{A}} \lambda_{ar} t_a(v))$, where f_r is the monetary outlay and φ_r is the value of time, for route $r \in \mathcal{R}$ (Gabriel and Bernstein, 1997). In this model, time is converted into money through the function φ. It is also possible to work with money as the cost unit, thus considering instead a cost model of the form (Larsson et al., 2002)

$$c_r(h) = \sum_{a\in\mathcal{A}} \lambda_{ar} t_a(v) + \nu_r(f_r), \tag{14}$$

where the function ν_r converts money into time. Interestingly, even if we choose $\nu_r = \varphi_r^{-1}$, the two models are not equivalent (Bernstein and Wynter, 2000). The latter model has the advantage that there exists an equivalent optimization formulation in the case of separable link costs: provided f_r is flow-independent, and that the value of ν_r is the same within one OD pair, the partial derivative of the objective function

$$\sum_{a\in\mathcal{A}} \int_0^{v_a} t_a(s) + \sum_{k\in\mathcal{C}} \sum_{r\in\mathcal{R}_k} h_r \nu_k(f_r)$$

with respect to h_r is $c_r(h)$ given in (14), and Wardrop's conditions are equivalent to the optimality conditions of this optimization problem. We also see that the objective function is convex under the same conditions as for the additive model.

Algorithms for non-additive traffic equilibrium models need explicit route information; a simplicial decomposition/route (column) generation algorithm à la DSD (Larsson and Patriksson, 1992) therefore seems natural. In the case of the above cost model, the only place where some complications arise is in the generation of new, profitable routes. The Wardrop conditions suggest the following extension of the shortest route problem, for an OD pair $k \in \mathcal{C}$, at some fixed link volume $\bar{v}$:

$$\underset{r\in\mathcal{R}_k}{\text{minimize}} \left\{ \sum_{a\in\mathcal{A}} \lambda_{ar} t_a(\bar{v}) + \nu_k(f_r) \right\}.$$

If and only if there exists a route in some OD pair that makes this value negative can the current solution be improved through the reallocation of flow onto one such route. The above problem cannot be solved using Dijkstra's shortest route algorithm, as each link has two attributes: travel time and money outlay; in particular, the dynamic programming principle—that sub-routes are optimal if the entire routes are—does not necessarily apply here. Instead, multi-label shortest route methods are used. Suppose that $f_r = \sum_{a\in\mathcal{A}} \lambda_{ar} f_{ar}$, so that the monetary outlay on the route is the sum of link tolls. Then we can devise an algorithm in which we have node labels with accumulated travel times and monetary outlays, coming from a specific node. Domination tests are used to keep down the number of labels stored, and those remaining at the terminal node are compared with respect to the function ν_k to determine the best route. The corresponding extension of the DSD algorithm was in Larsson et al. (2002) found to work quite well for this problem.

References

[AEA02] E. Altman, R. El Azouzi, and V. Abramov, "Non-cooperative routing in loss networks," *Performance Evaluation*, 49 (2002), pp. 43–55.

[Bai75] J. B. Baillon, "Un théorème de type ergodique pour les contractions nonlinéaires dans un espace de Hilbert," *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences (Paris), Série A*, 280 (1975), pp. A1511–A1514.

[Bar02a] H. Bar-Gera, "Origin-based algorithm for the traffic assignment problem," *Transportation Science*, 36 (2002) (to appear).

[Bar02b] H. Bar-Gera, "Origin-based network assignment," In: *Transportation Planning—State of the Art*, M. Patriksson and M. Labbé (eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002, pp. 1–17.

[BSS93] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons, New York, NY, second ed., 1993.

[BeW00] D. Bernstein and L. Wynter, "Issues of uniqueness and convexity in non-additive bi-criteria equilibrium models," Conference paper, 8th Meeting of the EURO Working Group on Transportation, Rome Jubilee 2000 Conference, La Sapienza, Rome, Italy, September 11–14, 2000.

[Ber76] D. P. Bertsekas, "On the Goldstein–Levitin–Polyak gradient projection method," *IEEE Transactions on Automatic Control*, AC-21 (1976), pp. 174–184.

[BeT89] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, London, 1989.

[BeG75] M. J. Best and V. J. Griffin, "Optimisation of nonlinear performance criteria subject to flow constraints," Conference paper presented at the ORSA/TIMS National Meeting, Chicago, IL, USA, 30 April–2 May 1975.

[Bru77] R. E. Bruck, Jr, "On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space," *Journal of Mathematical Analysis and Applications*, 61 (1977), pp. 159–164.

[Bru84] P. Brucker, "An O(n) algorithm for quadratic knapsack problems," *Operations Research Letters* 3 (1984), pp. 163–166.

[BuF91] J. V. Burke and M. C. Ferris, "Characterization of solution sets of convex programs," *Operations Research Letters*, 10 (1991), pp. 57–60.

[CLJ02] A. Chen, D.-H. Lee, and R. Jayakrishnan, "Computational study of state-of-the-art path-based traffic assignment algorithms," *Mathematics and Computers in Simulation*, 2060 (2002), pp. 1–10.

[CMZ00] J.-P. Crouzeix, P. Marcotte, and D. Zhu, "Conditions ensuring the applicability of cutting-plane methods for solving variational inequalities," *Mathematical Programming*, 88 (2000), pp. 521–539.

[Dun73] J. C. Dunn, "On recursive averaging processes and Hilbert space extensions of the contraction mapping principle," *Journal of the Franklin Institute*, 295 (1973), pp. 117–133.

[Esc86] L. F. Escudero, "A motivation for using the truncated Newton approach in a very large scale nonlinear network problem," *Mathematical Programming Study*, 26 (1986), pp. 240–244.

[Eva76] S. P. Evans, "Derivation and analysis of some models for combining trip distribution and assignment," *Transportation Research*, 10 (1976), pp. 37–57.

[FeP97] M. C. Ferris and J.-S. Pang, "Engineering and economic applications of complementarity problems," *SIAM Review*, 39 (1997), pp. 669–713.

[Fis80] C. S. Fisk, "Some developments in equilibrium traffic assignment," *Transportation Research*, 14B (1980), pp. 243–255.

[GaB97] S. A. Gabriel and D. Bernstein, "The traffic equilibrium problem with non-additive path costs," *Transportation Science*, 31 (1997), pp. 337–348.

[GaB84] E. M. Gafni and D. P. Bertsekas, "Two-metric projection methods for constrained optimization," *SIAM Journal on Control and Optimization*, 22 (1984), pp. 936–64.

[Ham84] J. H. Hammond, "Solving asymmetric variational inequality problems and systems of equations with generalized nonlinear programming algorithms," PhD thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, 1984.

[Hea80] D. W. Hearn, "Bounding flows in traffic assignment models," Research Report 80-4, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, 1980.

[Ius98] A. N. Iusem, "On some properties of paramonotone operators," *Journal of convex analysis*, 5 (1998), pp. 269–278.

[Kli83] J. G. Klincewicz, "A Newton method for convex separable network flow problems," *Networks*, 13 (1983), pp. 427–442.

[Lar+02] T. Larsson, P.-O. Lindberg, J. Lundgren, M. Patriksson, and C. Rydergren, "On traffic equilibrium models with a nonlinear time/money relation," In: *Transportation Planning—State of the Art*, M. Patriksson and M. Labbé (eds.), Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002, pp. 19–31.

[LLP97] T. Larsson, Z.-W. Liu, and M. Patriksson, "A dual scheme for traffic assignment problems," *Optimization*, 42 (1997), pp. 323–358.

[LaP92] T. Larsson and M. Patriksson, "Simplicial decomposition with disaggregated representation for the traffic assignment problem," *Transportation Science*, 26 (1992), pp. 4–17.

[LaP94] T. Larsson and M. Patriksson, "Equilibrium characterizations of solutions to side constrained asymmetric traffic assignment models," *Le Matematiche*, 49 (1994), pp. 249–280.

[LaP95] T. Larsson and M. Patriksson, "An Augmented Lagrangean dual algorithm for link capacity side constrained traffic assignment problems," *Transportation Research*, 29B (1995), pp. 433–455.

[LaP98] T. Larsson and M. Patriksson, "Side constrained traffic equilibrium models—traffic management through link tolls," in *Equilibrium and Advanced Transportation Modelling*, P. Marcotte and S. Nguyen, eds., Kluwer Academic Publishers, New York, NY, 1998, pp. 125–151.

[LaP99] T. Larsson and M. Patriksson, " Side constrained traffic equilibrium models: Analysis, computation and applications," *Transportation Research*, 33B (1999), pp. 233–264.

[LMP75] L. J. LeBlanc, E. K. Morlok, and W. P. Pierskalla, "An efficient approach to solving the road network equilibrium traffic assignment problem," *Transportation Research*, 9 (1975), pp. 308–318.

[Pat94] M. Patriksson, *The Traffic Assignment Problem—Models and Methods*, Topics in Transportation, VSP BV, Utrecht, The Netherlands, 1994.

[Pat98] M. Patriksson, *Nonlinear Programming and Variational Inequalities—A Unified Approach*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[Pat03] M. Patriksson, *Traffic Equilibrium Models: Analysis, Applications, and Optimization Algorithms*, Springer-Verlag (under preparation).

[PaR02] M. Patriksson and R.T. Rockafellar, "A mathematical model and descent algorithm for bilevel traffic management," *Transportation Science*, 36 (2002), pp. 271–291.

[Roc70] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[Roc76] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM Journal on Control and Optimization*, 14 (1976), pp. 877–898.

[She85] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[WXW01] Y. J. Wang, N. H. Xiu, and C. Y. Wang, "Unified framework of extragradient-type methods for pseudomonotone variational inequalities," *Journal of Optimization Theory and Applications*, 111 (2001), pp. 641–656.

[War52] J. G. Wardrop, "Some theoretical aspects of road traffic research," *Proceedings of the Institute of Civil Engineers, Part II*, (1952), pp. 325–378.

[ZPP69] S. I. Zuhovickiĭ, R. A. Polyak, and M. E. Primak, "Two methods of search for equilibrium points of n-person concave games," *Soviet Mathematics Doklady*, 10 (1969), pp. 279–282.