

ANALYTICAL DYNAMIC TRAFFIC ASSIGNMENT MODELS

Terry L. Friesz, Changhyun Kwon, and David Bernstein

The Pennsylvania State University, the Pennsylvania State University, and James Madison
University

1. Introduction

The rapid development of intelligent transportation system technologies and the policy emphasis on their deployment have increased the importance of predictive dynamic network flow models, especially so-called dynamic network loading and dynamic traffic assignment models. In this chapter we provide a critical review of analytic models used in predicting time-varying urban network flows. Specifically, we examine and compare four types of dynamics used as the foundation of dynamic network models:

- dynamics based on arc exit-flow functions
- dynamics for which both exit and entrance flow rates are controlled,
- dynamics based on arc exit-time functions, and
- tatonnement and projective dynamics.

We then describe the other assumptions attached to these dynamics to create dynamic network loading and dynamic traffic assignment models. Our intent is to illustrate the usefulness and limitations of each category of network dynamics as a foundation for predicting dynamic network flows. Our review is not exhaustive, but rather focuses on those network dynamics which, in our opinion, have had the greatest impact and engendered the most scientific debate in recent years. Following our review of network dynamics we describe how these are combined with postulates regarding route and departure-time choice to create predictive models for dynamically loading and assigning traffic to a network. Our discussion is unabashedly not mathematically rigorous in order to make this chapter readable by the widest possible audience. That is to say, although we do use a lot of symbolic notation, one does not need to follow any difficult derivations or be proficient in the calculus, real analysis or functional analysis to follow the key aspects of the presentation.

2. What is dynamic traffic assignment?

Static traffic assignment is, of course, that aspect of the transportation planning process that determines traffic loadings (expressed as flows, i.e., volumes per unit time) on arcs and paths of the road network of interest in a steady state setting. Dynamic traffic assignment (DTA) is concerned

with the same problem in a dynamic setting. DTA models may be either equilibrium or disequilibrium in nature. When the solution of a DTA model is a dynamic equilibrium, the flow pattern is time varying, but the trajectories through the time of arc and path flows are such that an appropriate dynamic generalization of Wardrop's first principle of user optimality is enforced at each instant of time. DTA models may be used to generate forecasts of traffic that illustrate how congestion levels will vary with time; these forecasts are intrinsically useful for traffic control and management in both the near-real time and deliberate planning contexts. The data requirements for DTA models are, on the surface, quite similar to those of static traffic assignment models; however, on closer examination, and as we make clear here, there is one very significant difference. That difference is that DTA models require path delay operators -- rather than the monotonically increasing-with-flow path-delay functions familiar from static assignment. Path delay operators express the delay on a given path in light of the time of departure from the origin of the path and the traffic conditions encountered along the path. This accounts for the fact that path traversal is not instantaneous and a platoon departing now will encounter traffic on subsequent arcs that may have departed previously as well as traffic that may have departed subsequently from the same or other origins. Thus, there is the potential for path delay operators to depend on the complete history (past, present and future) of flows on the network. Such delay operators are, except in certain special cases, not knowable in closed form; that is, delay operators for DTA are known only numerically for the general case and may require a simulation model to determine.

In fact, there are two main categories of DTA models: those that employ rule-based simulation, and those that do not but are instead based entirely on equations and inequalities. This second category is usually referred to as analytical DTA models and is the focus of the balance of this paper. This emphasis has been chosen in light of the growing recognition that analytical DTA models are extremely useful for deliberate planning, and in our view seem likely to dominate such applications in the future due to their relative simplicity and lower manpower costs for implementation.

3. Dynamic network loading and dynamic traffic assignment

Before proceeding, it is important to properly distinguish two overlapping problems that make use of traffic network dynamics. These are the dynamic network loading problem (DNLP) and the DTA problem (DTAP). The DNLP is the problem of finding dynamic arc volumes and flows (i.e., "loads") when time-varying departure rates for paths are known. Although a universally accepted statement of the DTAP has yet to emerge, in this chapter the DTAP will be the problem of simultaneously finding dynamic path departure rates and dynamic arc loadings. Clearly, in light of the above definitions, models for both the DNLP and the DTAP require submodels of how network arc flows change over time; it is these dynamic arc submodels which we call network dynamics. Submodels of route choice are coupled to network dynamics to create a model of dynamic network loading. If one also includes a submodel for departure-time choice, then a DTA model may be created from network dynamics. It is therefore not unreasonable to refer to network dynamics as the fundamental core of any model meant to address the DNLP and the DTAP.

3.1 Notation

Several different notational schemes have been developed by the scholars working on dynamic network models. We employ, insofar as it is possible, a common notation for discussing the models reviewed herein; this notation must necessarily differ somewhat from that employed in the original articulation of certain models of network dynamics. Among the definitions that are employed repeatedly in the following presentation are:

i, j, l indices generally referring to network arcs

a subscript, generally referring to an arc of the network

p subscript, generally referring to a path of the network
 N_o the set of origin nodes of the network
 N_D the set of destination nodes of the network
 A the complete set of arcs of the network
 $A(i)$ the set of arcs having tail node
 $B(j)$ the set of arcs having head node
 P the complete set of paths of interest for the network
 P_{ij} the set of paths connecting origin node i , to destination node j

4. Dynamics based on arc exit-flow functions

Let us posit that it is possible to specify and empirically estimate or to mathematically derive from some plausible theory, functions which describe the rate at which traffic exits a given network arc as a function of the volume of traffic present on that arc. To express this supposition symbolically, we use $x_a(t)$ to denote the volume of traffic on arc a at time t and $g_a(x_a(t))$ to denote the rate at which traffic exits from link a . Where it will not be confusing, we suppress the explicit reference to time t and write the arc volume as x_a and the exit-flow function as $g_a(x_a)$, with the understanding that both entities are time varying. It is also necessary to define the rate at which traffic enters arc a , which we denote by $u_a(t)$.

Again, when it is not confusing we suppress the time dependency of the entrance rate for arc a and simply write u_a . Note that both $g_a(x_a)$ and u_a are rates; that is, they have the units of volume per unit time, so it is appropriate to refer to them as exit flow and entrance flow, respectively. A natural-flow balance equation can now be written for each link:

$$\frac{dx_a}{dt} = u_a - g_a \quad (1)$$

where A denotes the set of all arcs of the network of interest. Although (1) is a fairly obvious identity, it seems to have been first studied in depth by Merchant and Nemhauser (1978a,b) in the context of system optimal DTA. The same dynamics were employed by Friesz et al. (1989) to explore certain extensions of the Merchant--Nemhauser work. Exit-flow functions have been widely criticized as difficult to specify and measure. Exit-flow functions also allow violation of the first-in-first-out (FIFO) queue discipline as illustrated and discussed by Carey (1986, 1987, 1992).

The Merchant--Nemhauser dynamics enforce flow-conservation constraints which for a single destination and multiple origins may be expressed as

$$\sum_{a \in A(k)} u_a(t) - \sum_{a \in B(i)} g_a[x_a(t)] = S_k(t) \quad \forall k \in M \quad (2)$$

where $S_k(t)$ denotes the rate of trip production at origin node k , M is the set of all origin nodes, $A(k)$ is the set of arcs with tail node k , and $B(k)$ is the set of arcs with head node k . Obviously, the arc volumes and arc entrance rates are non-negative: $x(t) \geq 0$ and $u_a(t) \geq 0$ for all arcs. Consequently, if we let

$$(x(t), u(t)) \equiv (x_a(t), u_a(t)) \quad \forall a \in A$$

then the set of feasible solutions corresponding to these dynamics is

$$\Lambda_1 = \{(x(t), u(t)) : \sum_{a \in A(k)} u_a(t) - \sum_{a \in B(i)} g_a[x_a(t)] = S_k(t) \quad \forall k \in M; \quad (3)$$

$$x_a(t) \geq 0, u_a(t) \geq 0, \forall a \in A; \forall t \in [0, T]\}$$

This allows the following shorthand summary of the Merchant--Nemhauser class of dynamics:

$$\frac{dx}{dt} = u - g(x) \quad (4)$$

$$(x, u) \in \Lambda_1 \quad (5)$$

$$x(0) = x^0 \quad (6)$$

where x is the vector of state variables and u is the vector of control variables; $x(0) = x^0$ of course represents the known initial conditions for the state variables with their time dependencies suppressed.

5. Dynamics with controlled entrance and exit flows

A possible modification of the Merchant--Nemhauser arc dynamics that avoids the use of problematic exit-flow functions is to treat both arc entrance and exit flows as control variables; i.e.,

$$\frac{dx}{dt} = u - v, \quad (7)$$

where v is an appropriately defined vector of exit-flow variables. By treating both arc entrance and exit flows as control variables, we do not mean to imply that any kind of normative considerations have been introduced, for these variables are viewed as controlled by the decisions of network users constrained by physical reality and observed only at the level of their aggregate (flow) behavior. Yet considerable philosophical debate has been engendered by dynamics (7).

Some scholars argue that drivers do not control their exit flows and any model based on that assumption is invalid. This argument seems somewhat specious as the word "control" in this context is merely an artifact of the version of the calculus of variations used to analyze this category of model; namely, optimal control theory. However, a more well-founded criticism is that missing from the unembellished version of eq. (7) is any explanation of the queue discipline for the various origin--destination flows on the same arc. Just as with the exit-flow functions, we have no way of ensuring that the FIFO queue discipline is enforced without additional constraints or assumptions. So the real "benefit" of this formulation seems to be avoiding the specification and measurement difficulties associated with exit-flow functions, not in ensuring FIFO. It is also very important to realize that, as written, eq. (7) does not incorporate any traffic flow theory and so holds the potential of admitting other flow-propagation anomalies besides FIFO violation, a point that we discuss in more detail below.

In fact, early efforts to use dynamics such as (7) resulted in flow-propagation speeds faster than would occur under free flow with no congestion for the arc delay functions used. This particular flow propagation anomaly has been called "instantaneous flow propagation". Other scholars have referred to models based on (7) as "back to the future" models, to emphasize their potential to describe flow propagation at speeds greater than those associated with free (uncongested) flow. Presently, authors using dynamics like (7) have added flow-propagation constraints to their mathematical formulations in an effort to generally ensure physically meaningful model outcomes.

In fact two types of flow propagation constraint for preventing instantaneous flow propagation in were suggested by Ran et al. (1993) for dynamics (7). The first is stated as

$$U_a^p(t) = V_a^p[t + \Delta_a(t)] \quad \forall a \in A, p \in P, \quad (8)$$

where $U_a^p(\cdot)$ and $V_a^p(\cdot)$ are the cumulative numbers of vehicles associated with path p which are entering and leaving link a , respectively, while $\Delta_a(t)$ denotes the time needed to traverse link a at time t , and P is the set of all paths. The meaning of these constraints is fairly intuitive: vehicles entering an arc at a given moment in time must exit at a later time consistent with the arc traversal time. Ran and Boyce (1994) state that they do not actually employ this constraint in their applications but instead use a second type of flow-propagation constraint. However, we shall see in

a subsequent section that, despite the intuitive appeal of (8), these constraints omit a fundamental term.

The second Ran et al. type of flow-propagation constraint is much more notationally complex and is omitted here for the sake of brevity. Suffice it to say that constraints of this second type are articulated in terms of path-specific arc volumes and are meant to express the idea that a path-specific traffic volume on an arc must ultimately visit a downstream arc or exit the network at the destination node of the path in question. Ran et al. argue that by enforcing this consideration they rule out FIFO violations and instantaneous flow-propagation anomalies. The Ran et al. framework for link dynamics, consisting of (7) and one or the other of their two types of flow propagation constraint, omits any effort to ensure consistency among their submodels for link dynamics, arc delay, and flow propagation. Like the models based on the Merchant--Nemhauser dynamics, the Ran et al. family of models also enforces flow conservation and non-negativity constraints, albeit expressed somewhat differently, and specifies initial conditions for the traffic dynamics. Consequently, dynamics of this family have the form

$$\frac{dx}{dt} = u - v, \quad (9)$$

$$\frac{dE}{dt} = e, \quad (10)$$

$$(x, E, u, v, e) \in \Lambda_2, \quad (11)$$

$$x(0) = x^0, \quad (12)$$

$$E(0) = 0 \quad (13)$$

where Λ_2 is the set of feasible solutions satisfying the flow conservation, flow propagation, and non-negativity constraints; while E is the vector of cumulative departures for each path and e is the vector of departure rates for each path. Note that x and E are state variables and u , v and e are control variables; $x(0) = x^0$ and $E(0) = 0$ of course represent the known initial conditions for the state variables. As stated above, these constrained dynamics employ largely independent arguments to motivate each submodel without attending to the possible conflict that can arise among the submodels. As we discuss below, there are in fact fundamental identities that describe the relationship of link dynamics, arc delay, and flow propagation to one another, and these identities are needed to articulate an internally consistent dynamic network user equilibrium model when both arc inflows and outflows are treated as controls.

6. Cell transmission dynamics

The cell transmission model is the name given by Daganzo (1994) to dynamics of the following form:

$$x_j(t+1) = y_j(t) - y_{j+1}(t), \quad (14)$$

$$y_j(t) = \min\{x_{j-1}(t), Q_j(t), \alpha[N_j(t) - x_j(t)]\} \quad (15)$$

where t is now a discrete time index and a unit-time step is employed. In the above, the subscript $j \in C$ refers to a spatially discrete physical "cell" of the highway segment of interest while $(j-1) \in C$ refers to the cell downstream; C is of course the set of cells needed to describe the highway segment of interest. Similarly to before, x_j refers to the traffic volume of cell j . Furthermore, y_j is the actual inflow to cell j , Q_j is the maximal rate of discharge from cell j , N_j is the holding capacity of cell j , and α is a parameter. Daganzo (1995) shows how eqs. (14) and

(15) can be extended to deal with network structures through straightforward bookkeeping.

Note that (15) is a constraint on the variables x_j and y_j . The language introduced previously is readily applicable to the cell transmission model; in particular (14) is arc (cell) dynamics (although now several dummy arcs can make up a real physical arc) and (15) is flowpropagation constraints. The cell transmission model also includes an implicit notion of arc delay. That notion, however, is somewhat subtle: namely, delay is that which occurs from traffic flowing in accordance with the fundamental diagram of road traffic. This is because (15), as explained by Daganzo (1994), is really a piecewise linear approximation of the fundamental diagram of road traffic. The fundamental diagram can be obtained from empirical measurement or from any of several hydrodynamic models of traffic flow. This feature immunizes the cell transmission model against potential inconsistencies among the three submodels: arc dynamics, flow propagation, and arc delay.

Lo (1999) was perhaps the first to use the cell transmission model as the dynamic foundation for a DTA model, in our view the dynamical description (14) and (15) has yet to be successfully coupled with route and departure-time choice mechanisms to yield a mathematically exact model for dynamic network user equilibrium. A major difficulty associated with using the cell transmission model as a foundation for a dynamic network user equilibrium model is the fact that the right-hand side of (14) is non-differentiable; this means that if the path delay operators are nonlinear any kind of control theoretic approach will involve a nonsmooth Hamiltonian and all the attendant difficulties.

7. Dynamics based on arc exit-time functions

Another alternative to the Merchant--Nemhauser dynamics (1) is based on the use of exit-time functions and their inverses. This approach, due to Friesz et al. (1993), allows one to avoid use of exit-flow functions and the pitfalls associated therewith. Friesz et al. (1993) employed arc exit-time functions and their inverses in a manner that has no antecedents in the literature on dynamic network modeling and is wholly original. The resulting formulation of link dynamics and of the dynamic network user equilibrium problem has been recognized by Adamo et al. (1998), Wu et al. (1998a,b), and Zhu and Marcotte (1999) as a mathematically and behaviorally sound formulation.

To understand the exit-time function, let t_e be the time at which flow exits the i^{th} arc of path p when departure from the origin of that path has occurred at time t_d . The relationship of these two instants of time is expressed as

$$t_e = \tau_{a_i}^P(t_d), \quad (16)$$

and we call $\tau_{a_i}^P(\cdot)$ the exit-time function for arc a_i of path p . The inverse of the exit time function is written as

$$t_d = \theta_{a_i}^P(t_e), \quad (17)$$

and describes the time of departure t_d from the origin of path p for flow which exits arc a_i of that path at time t_e . Consequently, the identity

$$t = \theta_{a_i}^P \left[\tau_{a_i}^P(t_d) \right] \quad (18)$$

must hold for all time t for which flow behavior is being modeled. The role of the exit-time function becomes clearer if we describe path p as the following sequence of conveniently labeled arcs:

$$p \quad \{a_1, a_2, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_{m(p)}\} \quad (19)$$

where $m(p)$ is the number of arcs in path p . It then follows immediately that the total traversal time for path p can be articulated in terms of the final exit-time function and the departure time:

$$D_p(t) = \sum_{i=1}^{m(p)} \left[\tau_{a_i}^P(t) - \tau_{a_{i-1}}^P(t) \right] = \tau_{a_{i-1}}^P(t) - t, \quad (20)$$

when departure from the origin of path p is at time t . Construction of the arc dynamics begins by noting that arc volumes are the sum of volumes associated with individual paths using the arc:

$$x_a(t) = \sum_p \delta_{ap} x_a^p(t) \quad \forall a \in A \quad (21)$$

where x_a^p denotes the volume on arc a associated with path p and

$$\delta_{ap} = \begin{cases} 1 & \text{if arc } a \text{ belongs to path } p \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

If we use the notation $h_p(t)$ for the flow entering path p (departing from the origin of path p) at time t , it is possible to express its contribution to the flow on any arc at a subsequent instant in time using the inverse exit-time functions defined previously. This representation takes the form of an integral equation that can be manipulated to yield

$$\frac{dx_{a_i}^P}{dt} = g_{a_{i-1}}^P(t) - g_{a_i}^P(t), \quad \forall p \in P, i \in [1, m(p)] \quad (23)$$

where $g_{a_{i-1}}^P$ is the flow entering arc a_i (which is the same as the flow exiting arc a_{i-1}) and $g_{a_i}^P$ is the flow exiting arc a_i . These entry and exit flows can be proven to obey

$$g_{a_i}^P(t) = \frac{d\theta_{a_i}^P(t)}{dt} h_p \left[\theta_{a_i}^P(t) \right] \quad (24)$$

Note that even though (23) is remarkably similar to (7), the entrance and exit flows eq. (24) have been very carefully and rigorously related to departure rates (i.e., path flows) to avoid internal inconsistencies and flow-propagation anomalies such as instantaneous propagation. Note also that the dynamics (23) are intrinsically complicated, having right-hand sides that are neither explicit functions nor variables but rather operators that involve inverse exit-time functions.

There is, however, a way of re-expressing the exit-time function based model of arc dynamics eq. (23) to obtain an alternative formulation involving constrained differential equations, state-dependent time lags, and arc entrance and exit flows that are control variables rather than operators. We will see that this alternative formulation obviates the need to explicitly know exit-time functions and their inverses, but nonetheless preserves all the main features of the Friesz et al. (1993) model of link dynamics. Illustration of these claims requires that we introduce some model of link delay. To this end we introduce a simple deterministic link delay model first suggested by Friesz et al. (1993) and named the "point queue model" by Daganzo (1995). To articulate this delay model, let the time to traverse arc a_i for drivers who arrive at its tail node at time t be denoted by $D_{a_i} [x_{a_i}(t)]$. That is, the time to traverse arc a_i is only a function of the number of vehicles in front of the entering vehicle at the time of entry. As a consequence, we have

$$\tau_{a_i}^P = t + D_{a_i} [x_{a_i}(t)] \quad \forall p \in P, \quad (25)$$

$$\tau_{a_i}^P = \tau_{a_{i-1}}^P(t) + D_{a_i} [\tau_{a_{i-1}}^P(t)], \quad \forall p \in P, i \in [2, m(p)] \quad (26)$$

By employing expressions (24), (25), and (26) together with the chain rule, as explained in Friesz et al. (1999), the following flow-propagation constraints are obtained:

$$g_{a_1}^P(t + D_{a_1}(x_{a_1}(t)))(1 + D_{a_1}'(x_{a_1}(t))\dot{x}_{a_1}) = h_p(t), \quad (27)$$

$$g_{a_i}^P(t + D_{a_i}(x_{a_i}(t)))(1 + D_{a_i}'(x_{a_i}(t))\dot{x}_{a_i}) = g_{a_{i-1}}^P(t), \quad (28)$$

where the overdot refers to a total time derivative. Expressions (27) and (28) are proper flow progression constraints derived in a fashion that makes them completely consistent with the chosen exit-time function dynamics and point queue model of arc delay. Note that these constraints involve a state-dependent time lag $D_{a_i}[x_{a_i}(t)]$ but make no explicit reference to the exit-time functions and their inverses. Expressions (27) and (28) may be interpreted as describing the expansion and contraction of vehicle platoons or wave packets moving through various levels of congestion en route to their destinations. These flow propagation constraints were first pointed out by Tobin (1993) and presented by Friesz et al. (1995). Astarita (1995, 1996) independently proposed flow-propagation constraints that may be readily placed in the form of (27) and (28).

To complete our development, we introduce some additional categories of constraints. The first of these are flow-conservation constraints, which we express as

$$\sum_{p \in ij} \int_0^T h_p(t) dt = Q_{ij}, \quad \forall i \in N_O, j \in N_D, \quad (29)$$

where Q_{ij} is the fixed travel demand for an origin-destination pair (i, j) associated with the fixed trip matrix

$$Q = (Q_{ij})_{j=1, |N_O|}^{i=1, |N_D|} \quad (30)$$

and T is the departure time horizon. Finally, we impose the non-negativity restrictions

$$x \geq 0, \quad g \geq 0, \quad h \geq 0 \quad (31)$$

where x , g , and h are the relevant vectors of state variables and control variables. We may now define

$$\Lambda_3 = \{(x, g, h) \geq 0 : \text{eqs. (27), (28) and (29) hold}\}, \quad (32)$$

which is the set of state and control variables that represent physically meaningful flow propagation.

As a consequence of the preceding development we can now state a third type of network dynamics based on proper flow propagation constraints and which is completely self-consistent:

$$\frac{dx_{a_1}^P(t)}{dt} = h_p(t) - g_{a_1}^P(t), \quad \forall p \in P, \quad (33)$$

$$\frac{dx_{a_i}^P(t)}{dt} = g_{a_{i-1}}^P(t) - g_{a_i}^P(t), \quad \forall p \in P, i \in [2, m(p)] \quad (34)$$

$$(x, g, h) \in \Lambda_3, \quad (35)$$

$$x(0) = x^0, \quad (36)$$

which makes clear that the link volumes $x_{a_i}^P$ are natural state variables while the path flows h_p and link entrance (exit) flows $g_{a_i}^P$ are natural control variables in this formulation.

8. Dynamic user equilibrium

The development of a formal model of DTA requires a precise definition of the flow state that

characterizes dynamic traffic networks. The most widely used characterization in the technical literature is the notion of dynamic user equilibrium (DUE). There are a number of measure theoretic subtleties associated with the description of a dynamic equilibrium. We omit these for the sake of brevity, but see Friesz et al. (1993) for a discussion of the formal measure theoretic definition of DUE. For our purposes here it suffices to say that a dynamic network user equilibrium is a flow state for which no group of similar travelers can elect a different departure time and choose a different route which will shorten the effective delay they experience.

Each of the dynamics reviewed above may be combined with the definition of DUE to create a predictive dynamic traffic network model which will determine time-varying flows on all arcs and paths of the network. Predictive models meant to solve the DNLP and DTAP differ from one another primarily in the dynamics chosen and the particular mathematical formalism used to express DUE. Moreover, since the demonstration by Friesz et al. (1993) that DUE may be represented as a variational inequality, most continuous-time, continuous-choice models for the DNLP and the DTAP have used the variational inequality formalism to express route and departure choices. The way in which the variational inequality formalism is employed depends on whether approximate or exact notions of path delay are employed and whether the intent is to model only route choice (the DNLP) or route and path choice (the DTAP).

The recursive relationships (25) and (26), when combined with the arc delay functions, lead after some fairly straightforward manipulations to closed-form path delay operators:

$$D_p(t, x) \equiv \text{unit delay on path } p \text{ for traffic conditions } x \quad (37)$$

$$= \sum_{i=1}^{m(p)} \delta_{a_i p} \Phi_{a_i}(t, x), \quad (38)$$

where the arc delay operators obeying

$$\begin{aligned} \Phi_{a_1}(t, x) &= D_{a_1}(x_{a_1}(t)) \\ \Phi_{a_2}(t, x) &= D_{a_2}(x_{a_2}(t + \Phi_{a_1})) \\ \Phi_{a_3}(t, x) &= D_{a_3}(x_{a_3}(t + \Phi_{a_1} + \Phi_{a_2})) \\ &\vdots \\ \Phi_{a_i}(t, x) &= D_{a_i}(x_{a_i}(t + \Phi_{a_1} + \dots + \Phi_{a_{i-1}})) \\ &= D_{a_i}(x_{a_i}(t + \sum_{j=1}^{i-1} \Phi_{a_j})). \end{aligned} \quad (39)$$

Typically, a penalty

$$[\Pi t + D_p(t, x) - T_A],$$

where T_A is a prescribed arrival time, is added to the path delay operator to obtain the effective unit path-delay operator

$$\Psi_p(t, x) = D_p(t, x) + F \{ t + D_p(t, x) - T_A \} \quad (40)$$

for each path p . This form of the path-delay operator was first stated by Friesz et al. (1993). It is now well understood that the variational inequality problem constructed from constrained dynamics (33) to (36) and the effective path-delay operators (40) is an exact representation of DUE and is, therefore, the foundation model for specific efforts to solve the DNLP and the DTAP. That variational inequality is

$$\begin{aligned} \text{find } (x^*, g^*, h^*) &\in \Lambda_3 \text{ such that} \\ \langle \Psi(t, x^*), h - h^* \rangle &= \sum_{p \in P} \int_0^T \Psi_p(t, x^*) [h_p(t) - h_p^*(t)] dt \geq 0 \end{aligned} \quad (41)$$

$$\text{for all } (x, g, h) \in \Lambda_3$$

where

$$\Psi(t, x^*) = \left(\Psi_p(t, x^*) : p \in [1, |P|] \right)$$

Many analytical models proposed to date for solving the DNLP and the DTAP are either special cases or extensions of (41). The formal proof of the correctness of (41) was first given by Friesz et al. (1999). Several successful algorithms for variants of (41) have been developed and tested. See, in particular, Xu et al. (1999) and Wu et al. (1998a,b).

9. Tatonnement and projective dynamics

In this section we discuss a very different kind of network dynamics which are not at all similar to the Merchant--Nemhauser dynamics, but rather are related to the concept of tatonnement from microeconomics. Classical microeconomic equilibrium models are often motivated with stories of an auctioneer who calls out a price, observes the potential demand and supply at that price, and adjusts the price upward or downward accordingly. Such tatonnement adjustment processes (from the French *tâtonner*, which means to grope or feel one's way) are not meant to describe reality, but are merely a convenient fiction used to motivate movements from disequilibria to equilibria. In general, a real or, as we say, a realizable price adjustment process must include mechanisms for handling the inventories (or other types of excess supply) and back-orders (or other types of excess demand) that often result from a market being out of equilibrium.

However, this is not the case for traffic equilibria, since there are no inventories or back-orders of commuter trips. In fact, this lack of inventories and back-orders allows the construction of relatively simple mathematical descriptions of traffic network disequilibria. The realizable generalizations of the tatonnement adjustment processes reported in the economics literature can be applied to the study traffic network disequilibria at the level of path and arc flows, as first suggested by Friesz et al. (1996). As such, these models are a type of DTA model that describe disequilibrium traffic states and their trajectories as they adjust toward equilibrium.

Space considerations permit us to only mention some of the salient features of this class of models. (A detailed discussion is given in Friesz et al. (1996). The fundamental decision variables are path flows (departure rates) and perceived costs. Path flows are considered to have rates of change proportional to excess travel cost, where excess travel cost is the difference between experienced travel cost and perceived travel cost. Similarly, perceived costs have rates of change that are proportional to excess travel demand, defined as the difference between demand and supply of transportation services for the experienced level of congestion. These simple concepts lead to network dynamics that, for appropriate regularity conditions, are asymptotically stable in the sense that flows tend toward either a static or a dynamic user equilibrium, depending on whether the time-scale employed is day-to-day or within-day. That is, the natural form of a tatonnement network dynamics is as a system of simultaneous differential equations; the articulation of these equations requires a priori a model of network user equilibrium. Because network equilibrium flows are constrained by nonnegativity, flow propagation and flow conservation considerations, as explained above, it is generally necessary to modify the tatonnement dynamics based on the rules described above in such a way that trajectories are deformed to obey the imposed constraints. This is done using the concept of a projection operator, which can be loosely defined as the mathematical apparatus for finding a trajectory in the feasible region that is mathematically "closest" to the unconstrained trajectory. The resulting projective network dynamics represent a promising alternative perspective for formulating and solving the DNLP and the DTAP, which allow, unlike the other perspectives described above, disequilibrium states to be explored. As such we expect this class of models to be the subject of much future research and development.

10. A Numerical Method for DUE Problems

Friesz and Mookherjee (2006) have shown that the theory of infinite dimensional mathematical programming may be used to solve the DUE problem. Space prevents us from giving

a detailed treatment of that algorithm and the regularity conditions invoked to assure convergence.

We consider the problem

find $u^* \in U$ such that

$$\langle F(x(u^*, u_D^*), u^*, u_D^*, t), u - u^* \rangle \geq 0 \text{ for all } u \in U \quad (42)$$

where

$$x(u, u_D) = \arg \left\{ \frac{dx}{dt} = f(x, u, u_D, t), x(t_0) = x^0, G(x, u, u_D) = 0, \Gamma[x(t_f), t_f] = 0 \right\} \quad (43)$$

and

$$u_D = \begin{pmatrix} u_1(t + D_1(x_1)) \\ u_2(t + D_2(x_2)) \\ \vdots \\ u_m(t + D_m(x_m)) \end{pmatrix}$$

Problem (42) is a general differential variational inequality of which the DUE problem presented previously is a special case. Friesz and Mookherjee (2006) have shown that an iterative fixed point algorithm on involving the minimum norm projection may be used to solve (42). We give (42) a symbolic name, $DVI(F, f, U, x^0, D)$.

10.1 The Fixed Point Problem

We are now in a position to state the following result.

Theorem 1 (*fixed point problem*) *When some regularity conditions hold, any solution of the fixed point problem*

$$u = P_U[u - \alpha F(x(u, u_D), u, u_D, t)]$$

where $P_U[\cdot]$ is the minimum norm projection onto $U \subseteq (L^2[t_0, t_f])^m$ and $\alpha \in \mathfrak{R}_{++}^1$, is also a solution of $DVI(F, f, U, x^0, D)$.

Naturally there is a fixed point algorithm associated with Theorem 1; that algorithm is summarized by the following iterative scheme:

$$u^{k+1} = P_U[u^k - \alpha F(x(u^k, u_D^k), u^k, u_D^k, t)] \quad (44)$$

The detailed structure of the fixed point algorithm is:

The Fixed Point Algorithm

Step 0. Initialization. Identify an initial feasible solution $u^0 \in U$ and set $k = 0$.

Step 1. Optimal control subproblem. Solve

$$\min_v J^k(v) = \gamma^T \Gamma[x(t_f), t_f] + \int_0^t \frac{1}{2} [u^k - \alpha F(x^k, u^k, u_D^k, t) - v]^2 dt \quad (45)$$

$$\text{subject to } \frac{dx}{dt} = f(x, v, v_D, t) \quad (46)$$

$$G(x, v, v_D) = 0 \quad (47)$$

$$v \in U \quad (48)$$

$$x(0) = x^0 \quad (49)$$

and call the solution u^{k+1} . Note that in this step it is advantageous to "unfold" and explicitly state the constraints embedded in the operator $x(u, u_D)$ of statement (43) in order to facilitate computation.

Step 2. Stopping test. If

$$\|u^{k+1} - u^k\| \leq \varepsilon_1$$

where $\varepsilon_1 \in \mathfrak{R}_{++}^1$ is a preset tolerance, stop and declare $u^* \approx u^{k+1}$. Otherwise set $k = k + 1$ and go to Step 1.

10.2 Descent in Hilbert Space for the Projection Sub-Problems

It is important to realize that the fixed point algorithm of Section 10.1 can be carried out in continuous time provided we employ a continuous time representation of each subproblem (45) through (49) from Step 1 of the fixed point algorithm. This may be done using a continuous time gradient projection method or a discrete time gradient projection method supplemented by spline approximations. For our present circumstances, that algorithm may be stated as:

Gradient Projection in Hilbert Space

Step 0. Initialization.. Pick $v^{k,0}(t) \in U$ and set $j = 0$.

Step 1. Finding state variables. Solve the state dynamics

$$\begin{aligned} \frac{dx}{dt} &= f(x, v^{k,j}, v_D^{k,j}, t) \\ x(0) &= x^0 \end{aligned} \quad (50)(51)$$

and call the solution

$$x^{k,j}(t) \quad (52)$$

In the event a discrete time method is used to solve the state dynamics (50) and (51), curve fitting is used to obtain the continuous time state vector (52).

Step 2. Finding adjoint variables. Solve the adjoint dynamics

$$\begin{aligned} (-1) \frac{d\lambda}{dt} &= \nabla_x H^k \Big|_{x=x^{k,j}} \\ \lambda(t_f) &= \frac{\partial \Gamma[x^{k,j}(t_f), t_f]}{\partial x(t_f)} \end{aligned} \quad (53)(54)$$

where H is the augmented Hamiltonian for the optimal control subproblem of the fixed point algorithm based on current information:

$$H^k = \frac{1}{2} [u^k - \alpha F(x^k, u^k, u_D^k, t) - v]^2 + \lambda^T f(x, v^{k,j}, v_D^{k,j}, t) + \phi^T G(x, v^{k,j}, v_D^{k,j})$$

Call the solution

$$\lambda^{k,j}(t) \quad (55)$$

In the event a discrete time method is used to solve the adjoint dynamics (53) and (54), curve fitting is used to obtain the continuous time adjoint vector (55).

Step 3. Finding the gradient. Determine

$$\nabla_{\mathbf{v}} J^{k,j}(t) \equiv \nabla_{\mathbf{v}} H^k$$

Step 4. Stopping test. For a fixed and suitably small fixed step size

$$\theta_k \in \mathfrak{R}_{++}^1$$

determine

$$\mathbf{v}^{k,j+1}(t) = P_U \left[\mathbf{v}^{k,j}(t) - \theta_k \nabla_{\mathbf{v}} J^{k,j} \right] \quad (56)$$

In the event a discrete time method is used to solve the above projection subproblem, curve fitting is used to obtain the continuous time control vector (56).

Step 5. Stopping test. For $\varepsilon_2 \in \mathfrak{R}_{++}^1$, a pre-set tolerance, stop if

$$\left\| \mathbf{v}^{k,j+1} - \mathbf{v}^{k,j} \right\| < \varepsilon_2$$

and declare

$$\mathbf{v}^{k*} \approx \mathbf{v}^{k,j+1}$$

Otherwise set $j = j + 1$ and go to Step 1.

Note that the above algorithm overcomes the two-point boundary value problem difficulty that is typically associated with the simultaneous determination of state and control variables and that is characteristic of optimal control problems. This is a direct result of determining controls, states and adjoints in a sequential manner. This sequentialness, however, is not an approximation; rather it is a result of the way the original DVI is represented in terms of mappings between appropriately defined Hilbert spaces.

10.3 Applying the Fixed Point Algorithm to the DVI Formulation

In order to apply the formalism developed in Section 10.1, we make the following observations/assumptions:

1. the controls are g and h ;
2. the state variables are the traffic volumes $x_{a_i}^p : \forall p \in P, i \in [1, m(p)]$;
3. the arc delays $D_{a_i}(x_{a_i}) : \forall p \in P, i \in [1, m(p)]$, which appear as explicit time shifts in the flow propagation constraints (27) and (28);
4. the operator $x(g, g_D, h, h_D)$ has the properties of continuity and G-differentiability, where g_D and h_D have the obvious definitions

$$g_D \equiv g(t + D(x)), \quad h_D \equiv h(t + D(x))$$

and of course

$$x \equiv \left(x_{a_i}^p : p \in P, i \in [1, m(p)] \right), \quad D \equiv \left(D_{a_i} : p \in P, i \in [1, m(p)] \right)$$

5. $D(x)$ is continuously differentiable with respect to x .

11. A Heuristic Treatment of State-Dependent Time Shifts in DTA Problems

A DTA problem is a special case of continuous time optimal control problem. However, the presence of state-dependent time shifts in the proper flow progression constraints makes most numerical approach inapplicable. The state-dependent time shifts must and can be accommodated using an implicit fixed point perspective, as innovated for the dynamic user equilibrium in the

previous section. More specifically, in such an approach, one employs control and state information from a previous iteration to approximate current time shifted functions. This idea, requires very strong assumptions if one wishes to give a rigorous proof of convergence. Generally, for DUE computations, the implicit fixed point approach is a heuristic for treating flow propagation constraints that may be summarized as follows:

1. Articulate the current approximate states (volumes) and controls (arc exit rates) by spline or other curve fitting techniques as continuous functions of time.
2. Using the aforementioned continuous functions of time, express time shifted controls as pure functions of time, while leaving unshifted controls as decision functions to be updated within the current iteration.
3. Update the states and controls, then repeat Step 2 and Step 3 until the control controls converge to a suitable approximate solution.

References

Adamo, V., V. Astarita, M. Florian, M. Mahut, and J.H. Wu (1998) "A framework for introducing spillback in link based dynamic network loading models", presented at: TRISTAN III, San Juan.

Astarita, V. (1995) "Flow propagation description in dynamic network loading models", in: *Proceedings of the International Conference on Applications of Advanced Technologies in Transportation Engineering*, Capri.

Astarita, V. (1996) "A continuous time link model for dynamic network loading based on travel time functions", in: *13th International Symposium on Theory of Traffic Flow*. New York: Elsevier.

Carey, M. (1986) "A constraint qualification for a dynamic traffic assignment problem", *Transportation Science*, 20(1).

Carey, M. (1992) "Nonconvexity of the dynamic traffic assignment problem", *Transportation Research B*, 26:127-133.

Daganzo, C.F. (1994) "Cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory", *Transportation Research B*, 28:269-287.

Daganzo, C.F. (1995) "Finite difference approximation of the kinematic wave model of traffic flow", *Transportation Research B*, 28:269-287.

Friesz, T.L. (1992) "Letter", *Operations Research*.

Friesz, T.L., J. Luque, R.L. Tobin and B.W. Wie (1989) "Dynamic network traffic assignment considered as a continuous time optimal control problem", *Operations Research*, 37:893-901.

Friesz, T.L., D. Bernstein, T.E. Smith, R.L. Tobin and B.W. Wie (1993) "A variational inequality formulation of the dynamic network user equilibrium problem", *Operations Research*, 41:179-191.

Friesz, T.L. and D. Bernstein (1993) "Variational inequalities, dynamical systems and control theoretic models for predicting time-varying urban network flows", *Transportation Research*, 30:14-31.

- Friesz, T.L., R.L. Tobin, D. Bernstein and Z. Suo (1995) "Proper flow propagation constraints which obviate exit functions in dynamic traffic assignment", presented at: INFORMS Spring National Meeting, Los Angeles.
- Friesz, T.L., D. Bernstein and R. Stough (1996) "Dynamic systems, variational inequalities and control theoretic models for predicting time-varying urban network flows", *Transportation Science*, 30:14-31.
- Friesz, T.L., D. Bernstein, Z. Suo and R.L. Tobin (1999) "A new formulation of the dynamic network user equilibrium problem", Network Analysis Laboratory, George Mason University, Working Paper 99-05.
- Friesz, T.L. and R. Mookherjee (2006) "Solving the dynamic user equilibrium problem with state-dependent time shifts", *Transportation Research Part B*, 40(3):207-229.
- Merchant, D.K. and G.C. Nemhauser (1978a) "A model and an algorithm for the dynamic traffic assignment problems", *Transportation Science*, 12:183-199.
- Merchant, D.K. and G.C. Nemhauser (1978b) "Optimality conditions for a dynamic traffic assignment model", *Transportation Science*, 12:200-217.
- Lo, H.K. (1999) "A dynamic traffic assignment formulation that encapsulates the cell-transmission model", in: A. Ceder, ed., *Transportation and traffic theory*, Oxford: Pergamon Press, pp. 327-350.
- Ran, B. and D.E. Boyce (1994) *Dynamic urban transportation network models: theory and implications for intelligent vehicle-highway systems*. Berlin: Springer-Verlag.
- Ran, B., D.E. Boyce and L.J. LeBlanc (1993) "A new class of instantaneous dynamic user-optimal traffic assignment models", *Operations Research*, 41:192-202.
- Tobin, R.L. (1993) "Notes on flow propagation constraints", Network Analysis Laboratory, George Mason University, Working Paper 93-10.
- Wie, B.-W., R. Tobin, D. Bernstein and T.L. Friesz (1995) "Comparison of system optimum and user equilibrium dynamic traffic assignments with schedule delays", *Transportation Research D*, 3:389-411.
- Wu, J.H., Y. Cen and M. Florian (1998a) "The continuous dynamic network loading problem: A mathematical formulation and solution method", *Transportation Research B*, 32:173-187.
- Wu, J.H., M. Florian and J.M. Rubio-Ardannaz (1998b) "The continuous dynamic network loading problem: recent computational results", presented at: TRISTAN III, San Juan.
- Xu, Y., J.H. Wu, M. Florian, P. Marcotte and D.L. Zhu (1999) "Advances in continuous dynamic network loading problem", *Transportation Science*, 33:341-353.
- Zhu, D.L. and P. Marcotte (1999) "On the existence of solutions to the dynamic user equilibrium problem", unpublished.