OPTIMAL TIME-VARYING FLOWS
ON CONGESTED NETWORKS

MALACHY CAREY
Carnegie-Mellon University, Pittsburgh, Pennsylvania
(Received August 1982; revisions received September 1983, December 1984, February 1986; accepted March 1986)

This paper develops a well-behaved convex programming model for least-cost flows on a general congested network on which flows vary over time, as for example during peak/off-peak demand cycles. The model differs from static network models and from most work on multiperiod network models because it treats the time taken to traverse each arc as varying with the flow rate on the arc. We develop extensions of the model to handle multiple destinations and multiple commodities, though not all of these extensions yield convex programs. As part of its solution, the model yields a set of nonnegative time-varying optimal flow controls for each arc. We determine and discuss sufficient conditions under which some or all of these optimal flow controls will be zero-valued. These conditions are consistent with computational experience. Finally, we indicate directions for further research.

Flow variation over time is important in many types of networks, including networks for communication and information flow, computers, public utilities and services, commodity and financial flows, social and political interaction, and traffic flows. For the sake of concreteness, most examples and illustrations in this paper will refer to one class of networks, namely road traffic networks.

Dynamic network flow modeling is one of the most important, and most underdeveloped, areas of network modeling. The prediction or assignment of flows when demands are changing over time is especially relevant to the modeling of peak period flows, and it is peak period flows that impose the most pressure on capacity, and that are therefore most in need of modeling.

Almost all work to date on network flows over time has considered only one or two arcs or intersections, has been heuristic, has assumed that the travel time on each arc of the network is independent of the flow rate or volume on the arc, or has used simulation rather than optimization. See, for example, Yagar (1976), Robillard (1974), D'Ans and Gazis (1976), Gerlough and Huber, Chapter 7, (1975), Hendrickson and Kocur (1981) and Hurdle (1981), all of whom are concerned mainly with traffic networks. While this work is certainly important and interesting, it is not at all clear how its limitations can be relaxed, particularly since they were in most cases introduced to make the problems tractable. These authors, and others, have remarked on the lack, and the importance, of network models in which arc or path travel times depend on the flow rates. Bookbinder and Sethi (1980), in their survey of research on the dynamic transportation problem, suggest important future extensions and conclude (p. 84) that: "Perhaps the most important extension is the case in which the time required for a shipment to reach from source i to sink j is a function of the amount shipped."

Merchant and Nemhauser (1978a, b) made a promising start on the task of modeling dynamic network flow: they provide a single-destination, single-commodity, system optimizing traffic assignment model in which travel costs, route choices and flow rates are interdependent. (We will refer to Merchant and Nemhauser throughout as M-N.) Their model can be viewed as a partial generalization of the well-known static, system optimizing, assignment model, for which some basic references are Beckmann, McGuire and Winsten (1956), Dafermos and Sparrow (1969), Florian (1984), and Magnanti (1984). The M-N model is formulated as a discrete time nonlinear program, in which congestion is represented explicitly in the constraints. Unfortunately, this program is nonconvex, and this nonconvexity causes the analytical and computational problems they encounter with the model.

In this paper we reformulate the dynamic least-cost network flow problem as a convex nonlinear program. This new formulation has immediate analytical, computational and interpretational advantages, which the M-N model does not. In particular:

(i) It is a convex program, and hence can be solved using any of a variety of well-known algorithms al-

Subject classification: 481 optimal dynamic flows with congestion, 691 networks, 831 dynamic traffic assignment.
ready developed for convex programs; for example, a piecewise linearized version of the model is a standard linear program.

(ii) The Kuhn-Tucker conditions are both necessary and sufficient to characterize an optimal solution, whereas in the M-N model they are not sufficient and have not previously been shown to be necessary.

(iii) The new formulation allows us to determine whether the aggregate network costs or benefits can be improved by reducing the flow rates on arcs below their maximum or natural level, and determines an optimal pattern of such flow controls.

(iv) Various extensions of the model, introduced in Sections 3, 4 and 7, have the same desirable features (i) and (ii) above. These extensions include flexible departure times and elastic demands (Section 7) and some methods of introducing multiple destinations and multiple commodities (Sections 3 and 4).

On the other hand, it is worth highlighting some limitations of the present model that are associated with applying the model to congested road traffic networks.

(i) In the case of congested flows on road traffic networks, there is an important distinction between system-optimal flow patterns and “user-equilibrium” flow patterns (see the references cited earlier). The present paper considers only the former, and it is not obvious how our approach can be adapted to model the latter.

(ii) Section 4 introduces various approaches to modeling multiple commodity flows. Some of these approaches yield convex programs and some yield nonconvex programs. In the case of road traffic, the most appropriate multicommodity model appears to be nonconvex and there is as yet no computational evidence or experience with solving this model. The other (convex) multiple destination and multiple commodity models introduced in Sections 3 and 4 can be solved using existing general nonlinear convex programming methods. We have used the MINOS program package of Murtagh and Saunders (1977, 1980) to solve small-scale problems (Carey and Srinivasan 1982, and Section 8).

This paper is organized as follows. Section 1 sets out a convex, dynamic, single-destination network flow model, which represents congestion effects and flow controls in the constraints, and states some properties of this model. Section 2 makes some comparisons between this model and that of M-N. In Sections 3 and 4, we present several methods for extending the model to multiple destinations and multiple commodities. Section 5 introduces and interprets a set of assumptions which, as shown in Section 6, are sufficient to ensure that the optimal values of some or all of the flow controls are zero; this property is of interest for reasons stated in Section 5. Section 7 extends results from earlier sections to include flexible departure times and elastic travel demands. Section 8 presents some computational experience, and Section 9 indicates some directions for further research.

1. The Basic Model

Let the network be represented by a set of nodes $N$ joined by a set of directed arcs $A$. In this section, we assume that there is only one destination node. Let $N'$ be the set of nodes excluding the destination node. Let the time span to be considered (e.g., the morning rush period) be subdivided into $t = 0, \ldots, T$ time periods of equal length. Let

\[ x_{ij} = \text{the volume on arc } j \text{ at the beginning of period } t. \]

\[ d_{ij} = \text{the inflow into arc } j \text{ during period } t. \]

\[ b_{ij} = \text{the “actual” outflow from arc } j \text{ in period } t: \text{achieving this outflow may require introducing flow controls on arc } j \text{ to restrict the potential outflow, which is } g_i(x_{ij}). \]

The “maximum” or “capacity” or “natural” outflow associated with arc $j$ in period $t$, when the volume on the arc is $x_{ij}$ and when arc characteristics, which may include existing flow controls, are taken as given.

For example, for a road network, $g_i(x)$ is the flow rate when road characteristics, speed limits, traffic signals, and so forth are taken as given. $g_i(x)$ will be referred to as the congestion function or the exit function. Clearly, $g_i(0) = 0$ and $g_i(x) \geq 0$, for all $x \geq 0$, and for the purpose of our analysis, we can assume that $g_i(x)$ is continuous for $x \geq 0$. Any further properties assumed for $g_i(x)$ in the following theorems or lemmas will be explicitly stated. Thus, we can normally assume that $g_i(x)$ is concave, but we do not impose this assumption in the two main theorems (1 and 2) in Section 6. Also, it is convenient to assume in some of our lemmas that time periods $t = 0, \ldots, T$ are each sufficiently short as to ensure that the outflow cannot exceed the volume on the arc (i.e., $x \geq g_i(x)$), but in general, of course, this condition need not be met, since if periods are longer than the time taken to traverse the arc then both the inflow and the outflow per period from an
arc can exceed the volume on the arc.

\( h_{ij}(x_{ij}) = \) the travel cost incurred by the volume \( x_{ij} \) on arc \( j \) in period \( t \).

The function \( h_{ij}(\cdot) \) can usually be assumed to be continuous, convex, nondecreasing and nonnegative. For example, in the perhaps most likely scenario, the only cost to be considered is time, or is proportional to time, in which case \( h_{ij}(x_{ij}) = klx_{ij} \), with \( l \) the length of each time period and \( k \) the cost or value of unit time. However, in some applications some of these properties might not hold, hence when we assume any of these properties in the following theorems, they will be specified.

\( F_{ik} \) = the exogenous demand (inflow) at node \( k \) in period \( t \).

\( E_{ij} \) = the initial volume on arc \( j \), i.e., \( x_{0ij} = E_{ij} \geq 0 \).

We now introduce flow control or congestion control constraints

\[
0 \leq b_{ij} \leq g_{ij}(x_{ij}) \quad \text{for all} \quad t, j, \tag{1}
\]

which are implicit in the definitions of \( b_{ij} \) and \( g_{ij}(x_{ij}) \), and which can be interpreted as follows. Flow controls can be used to keep the actual outflow \( b_{ij} \) below the natural or unrestricted capacity level \( g_{ij}(x_{ij}) \). On the other hand, it will not normally be possible to increase the outflows \( b_{ij} \) above the level \( g_{ij}(x) \) without additional investment in arc capacity—and such investment would in any case change the form or parameters of the function \( g_{ij}(x) \). We will refer to the difference \( s_{ij} = (g_{ij}(x_{ij}) - b_{ij}) \) as the “flow control” for arc \( j \) in period \( t \). Clearly \( s_{ij} \geq 0 \). Though this flow control relates to the flow exiting from an arc, it is more interesting to think of it as controlling the flow entering the node, and the arcs into which the given arc leads.

It should be emphasized here that even if we do not wish to consider traffic controls (i.e., we want \( b_{ij} = g_{ij}(x_{ij}) \)), we will find that there are still substantial advantages, for analytical and programming reasons, in writing (1) as an inequality rather than as an equality. We will show in Section 6 that when inequality (1) is used in the dynamic network model C, to be specified, the solution will in fact turn out to satisfy (1) as an equality for a wide range of cases of interest. Thus, even for these cases, nothing is lost, and much is gained, by writing (1) as an inequality.

The usual nodal flow conservation equations can be stated as

\[
\sum_{j \in A(k)} d_{ij} = F_{ik} + \sum_{j \in B(k)} b_{ij} \tag{2}
\]

for each node \( k \in N^* \) and each period \( t = 0, \ldots, T - 1 \); in this expression \( A(k) \) is set of arcs pointing out of node \( k \) and \( B(k) \) is the set of arcs pointing into node \( k \). The volume on an arc in any period is equal to the volume on the arc in the previous period plus the net inflow during the period, thus

\[
x_{t+1,j} = x_{t,j} - b_{t,j} + d_{t,j} \tag{3}
\]

for all periods \( t \) and arcs \( j \).

The system cost minimizing, dynamic flow problem for a congested network can now be stated as follows.

**Model C**

Minimize

\[
\sum_{i=1}^{T} \sum_{j \in A} h_{ij}(x_{ij}) \tag{4a}
\]

subject to, for periods \( t = 0, \ldots, T - 1 \),

\[
g_{ij}(x_{ij}) \geq b_{ij} \quad \text{for all} \quad j \in A, \tag{4b}
\]

\[
b_{ij} = x_{t,j} - x_{t+1,j} + d_{t,j} \quad \text{for all} \quad j \in A, \tag{4c}
\]

\[
\sum_{j \in A(k)} d_{ij} = F_{ik} + \sum_{j \in B(k)} b_{ij} \quad \text{for all} \quad k \in N^*, \tag{4d}
\]

\[
x_{t,j} = E_{ij} \quad \text{for all} \quad j \in A, \tag{4e}
\]

\[
(b_{ij}, d_{ij}, x_{ij}) \geq 0 \quad \text{for all} \quad j \in A. \tag{4f}
\]

If \( g_{ij}(x) \) has the property \( g_{ij}(x) \leq x \) for all \( j \) and all \( x \geq 0 \), then the constraints \( x_{t,j} \geq 0 \) are redundant for all \( t \) and \( j \), since the constraints \( g_{ij}(x) \leq x \), (4b, 4c, 4d) and the rest of (4e) ensure that \( x_{t,j} \geq 0 \) for all \( t, j \).

In this program, the constraint set (4b) is convex, since \( g_{ij}(x) \) is concave. All the other constraint functions are linear, hence the constraint set of C is convex. Also, the objective function is convex, since \( h_{ij}(x) \) is convex. Thus C is a convex program, and has the desirable properties of such programs. For example, (a) any local optimum of C is a global optimum, (b) if \( h_{ij}(x) \) is strictly convex then any local optimum of C will be the unique optimum, and (c) the Kuhn-Tucker conditions for C are sufficient to ensure a global optimum of C. Also, the following lemma will enable us to derive properties of the solution of C.

**Lemma 1.** Let \( E_{ij} > 0 \) for all \( j \). Let \( g_{ij}(x) \) be concave and \( x \geq g_{ij}(x) > 0 \) for all \( x > 0 \) and \( j \in A \). Then program C will satisfy a constraint qualification, and the Kuhn-Tucker conditions (A.3a–A.3f in the Appendix) will hold for any optimal solution of C.

**Proof.** Lemma A1 in the Appendix ensures that C satisfies a constraint qualification, hence the lemma follows (see, for example, Bazarra and Shetty 1979).

Since an explicit statement of the Kuhn-Tucker conditions is not needed until Lemma A3 in the
Appendix, we will defer stating them until then. The assumption that the initial arc volumes \( E_j \) are all positive is a trivial restriction since it can always be ensured by arbitrarily small positive values of \( E_j \).

Since \( C \) is a convex program, any of several well-known nonlinear programming algorithms can be used to solve it. Or, we can take a piecewise linear approximation to \( C \), by piecewise linearizing the functions \( g_i \) and \( h_{ij} \). This formulation can then be solved by the simplex algorithm: the ordered set property (OSP) associated with separable programming will be automatically satisfied since \( C \) is a convex program (Bazaraa and Shetty, Chapter 11).

2. The Merchant-Nemhauser Model

If we change the inequalities (4b) to strict equalities, and use these to eliminate \( h_{ij} \) from (4c) and (4d), we obtain the dynamic assignment program set out in M-N (1978a,b), that is:

Model P

Minimize \[ \sum_{i=1}^{T} \sum_{j \in A} h_{ij}(x_{ij}) \]  

subject to, for periods \( t = 0, \ldots, T - 1 \),

\[ x_{i+1,j} = x_{ij} - g_i(x_{ij}) + d_{ij} \quad \text{for all } j \in A, \]  

\[ \sum_{j \in A(k)} d_{ij} = F_{ik} + \sum_{j \in B(k)} g_i(x_{ij}) \quad \text{for all } k \in N', \]  

\[ x_{ij} = E_j \quad \text{for all } j \in A, \]  

\[ d_{ij} \geq 0 \quad \text{for all } j \in A. \]

M-N show that their formulation is in general non-convex, and point out that it has not been shown to satisfy a "constraint qualification." Therefore, they encountered computational and analytical difficulties, none of which arise with formulation C. For example:

(a) They propose a solution technique based on piecewise linearization and linear programming. This approach requires that a certain ordered set property (OSP) be satisfied, and because of the nonconvexity of the problem, this requires a potentially expensive computational scheme, in addition to the simplex algorithm. They are able to show (in their main theorem) that the simplex algorithm will find the optimal objective function value for their original OSP-constrained linear program. But finding the optimal solution will almost always require an additional computational scheme: they recommend the one devised by Ho (1980).

(b) The lack of a constraint qualification means that it is not known whether the Kuhn-Tucker conditions hold at an optimal solution of program P. But the entire analysis of P in M-N (1978b) is based on Kuhn-Tucker conditions, and hence is conditional upon this unknown. (We are able to remedy this situation in Section 6.)

(c) M-N state (1978a, p. 193) that "we suspect that problem P cannot have multiple local optima when.... However, we have been unable to prove this conjecture." No such difficulty arises with program C, since it is convex.

Finally, suppose that we do not in fact wish to consider nonzero flow controls. Then it might seem obvious that in this case we should use model P rather than model C. However, this is not so. The very assumptions which M-N (1978a) introduced in order to devise a solution scheme for their model P turn out to be (almost) sufficient to ensure that the optimal flow controls in model C are zero—in which case model C provides a solution for model P, as shown in Sections 5 and 6.

3. Multiple Destinations

The single-destination, single-commodity model can easily be used to handle multiple destinations, since, as well-known, multiple destination networks can be reduced to a single destination network by introducing artificial arcs linking the given destinations to a single artificial destination.

However, there is a further complication in multi-period network models, since it may not be feasible for some of the volume on the network to get to the destination by the final period. In view of this potential problem, we now consider several different ways of introducing demand at destinations. In the first four of these approaches (i)-(iv)), we assume exogenous demands, whereas in (v) we introduce demands determined endogenously as a function of prices/costs. Let \( D \) be a set of artificial arcs, each of which links a given destination to the single artificial destination: the flows on these arcs represent the demands at the corresponding destinations.

(i) Introduce fixed aggregate demands \( \bar{d}_j \) at each destination,

\[ \sum_{i=0}^{T} d_{ij} = \bar{d}_j \quad \text{for all destination arcs } j \in D. \]  

A problem with this approach lies in choosing feasible \( \bar{d}_j \) values, since we may not know in advance what volumes can feasibly reach the destination by the terminal period.
(ii) The problem specified in (i) can be overcome by letting the \( \bar{d}_j \)'s be maximum demands, thus

\[
\sum_{t=0}^{T} d_{tj} \leq \bar{d}_j \quad \text{for all destination arcs } j \in D. \tag{9}
\]

If we chose the \( \bar{d}_j \)'s so that their sum (\( \sum_{j \in D} \bar{d}_j \)) equals the sum of the exogenous inflows \( F_{ik} \) plus the initial volumes \( E_{ij} \), then (9) will always be feasible, and indeed at least one of the constraints (9) will be a strict inequality if there is any volume still on the network in the terminal period, i.e., if some \( x_{tT} > 0 \).

(iii) To have more exogenous control over the actual outflows at destinations \( j \in D \) than is implied in (ii), one could solve the network model, subject to (9), and then parametrically adjust the \( \bar{d}_j \)'s so as to explore the set of feasible outflows. For example, if (9) turns out to be slack for some destinations \( j \in D \) and an equality for some other destinations, then the \( \bar{d}_j \)'s for the latter could be reduced so as to force more flow to be sent to satisfy the demands at the former (slack demand) destinations.

(iv) If the demands \( \bar{d}_j \) in (9) cannot be met, that is, if in an optimal solution there are significant slacks in some members of (9), and if this outcome is considered to be a problem, then a possible remedy would be to introduce more time periods \( t = T, T + 1, \ldots \), so as to allow more time for the volumes still on the network in the terminal period \( T \) to make their way to the destination(s). If the exogenous inflows during these additional periods are zero, or small, then the slacks in (9) will be reduced.

(v) Instead of introducing demand constraints such as (8) or (9), a rather different approach consists of associating "benefit" (or negative costs) with the outflows at each destination, and then introducing these benefits in the objective function so as to "attract" flows to the destination. Thus, let \( p_{ij} \) be the benefit or value of one unit of \( d_{ij} \) at destination \( j \in D \), and add the expression \( -\sum_{t=0}^{T} \sum_{j \in D} p_{ij} d_{tj} \) to the cost-minimizing objective function. The objective function now represents net cost, or (negative) net benefit. More generally, a convex nonlinear benefit function can be used. Such benefit functions are widely available, since in economics and cost-benefit analysis it is common to estimate demand functions \( d_i(p_{ij}) \), where \( p_{ij} \) is the price or marginal benefit associated with \( d_{ij} \), and use the area \( w(d_{ij}) = \int_{p_{ij}}^{p_{ij}} p_{ij}(d_{ij}) \, dd_{ij} \) under the inverse demand function \( p_{ij}(d_{ij}) \) as a measure of the benefit associated with \( d_{ij} \); the function \( p_{ij}(d_{ij}) \) is normally assumed to have a negative slope, so that the benefit integral is concave.

4. Multiple Commodities and Flow Controls

For brevity, we will restrict our discussion of introducing multiple commodity types: Carey (1984) discusses the topic at greater length. An obvious first step to introducing multiple commodities in the dynamic flow model C is as follows.

(a) Introduce a commodity subscript \( c = 1, \ldots, C \), on all variables and constants, and repeat constraints (4c)-(4f) for all \( c = 1, \ldots, C \).

(b) In (4a) and (4b) replace \( h_i(x_{ij}) \) and \( g_i(x_{ij}) \) respectively with \( h_i(x_{ij1}, \ldots, x_{ijC}) \) and \( g_i(x_{ij1}, \ldots, x_{ijC}) \).

Beyond this first step, there are several different ways to introduce multiple commodity types in model C, the particular formulation depending on the extent to which the various commodity types can in practice be separated, segregated or otherwise treated differentially by the flow controller. To illustrate, it is convenient to consider two extreme cases (Carey 1984 considers intermediate cases).

Case (i). Suppose that it is relatively easy, or costless, to hold some types of commodity back on an arc while allowing other types to exit from the arc. To model this case, we simply introduce (a) and (b) into program C and replace the flow control constraints (4b) with

\[
g_i \left( \sum_c x_{ijc} \right) \geq \sum_c b_{ijc}. \tag{10}
\]

This tactic ensures that the flow capacity \( g_i(\cdot) \) for arc \( j \) is satisfied, but imposes no restriction on the order in which the commodities on arc \( j \) exit from arc \( j \).

Case (ii). At the other extreme, suppose that it is impossible or very costly in practice to treat commodity types differently (holding some types back so as to allow other types priority to proceed). To model this case we again introduce (a)-(b) and (10), as in (i), but also introduce

\[
\frac{b_{ijc}}{b_{ij1}} = \frac{x_{ijc}}{x_{ij1}} \quad c = 1, \ldots, C, \tag{11}
\]

which ensures that the ratio (mixture) of commodities exiting from arc \( j \) is identical to the ratios (mixture) of commodity volumes present on arc \( j \). In other words, (11) states that the controller does not give any one commodity priority over any other.

The formulation for (i) has the advantage that, like program C, it is a convex optimization model and hence retains the desirable properties of C. The formulation for (ii) yields a nonconvex program, since
(11) are nonlinear equations and hence represent a nonconvex set. Consequently, the program for (ii) is more difficult to analyze and solve than the program for (i).

Formulations (i) and (ii) reflect the presence of different types of flow controls in the environment being modeled. For example, in the case of road traffic, flow controls can be implemented by installing electronic, mechanical, or human counters to measure traffic flows, and using these to activate traffic stop lights when the desired number of vehicles has passed by in each time period. In the case of air traffic, aircraft can be held in waiting areas or required to circle the airport. In the case of rail traffic, trains can be held in sidings or required to alter speed. In the case of communications networks, information can be held in buffers and released as desired.

If there are multiple commodities on the network, then the problem formulation should reflect the commodity flow interactions and the kind of control the operator has over these flows. For example, in a rail network or a packet-switched communications network, it costs relatively little to record the commodity type and/or destination, and to impose different flow rates or different routes on these different commodities. This situation suggests using formulation (i). On the other hand, in the case of road traffic, flows of different types of vehicles tend to move together at the same speed, and it is usually rather difficult and costly to treat vehicle types differently at flow control points. We may therefore wish to treat vehicles homogeneously, which can be achieved by using formulation (ii). This raises the question, why introduce vehicle types at all if they are then not differentiated at flow control points? The answer is of course that travel demands, travel costs and benefits, and routes chosen may differ between types: types may refer to physical type (cars, trucks, or busses); destination type (i.e., vehicles bound for particular destinations); or social type (i.e., vehicles having a driver/owner in a particular social or income group).

5. Zero versus Nonzero Flow Controls: Assumptions and Interpretations

Though nonzero flow controls are clearly of interest, we are also particularly interested in establishing conditions under which the optimal flow controls yielded by program C will be zero-valued. These conditions are of interest since:

(a) they should help flow controllers/planners decide when not to control flows,
(b) they reduce our dynamic model to a generalization of the well-known static model in which flow controls do not exist,
(c) they relate our model to that of M-N, in which flow controls are not considered, and
(d) they allow the model to be used as an approximation to a “user equilibrium” or “descriptive assignment” for road traffic. (A user equilibrium assumes that individual travelers are left to make their own routing decisions and are not directed by a system optimizing traffic controller (see, e.g., Florian).)

In order to establish conditions under which the optimal flow controls in program C are zero (i.e., $s = 0$) we must now introduce certain assumptions concerning the cost functions $h_{ij}(x)$ and the exit functions $g(x)$. These assumptions will be sufficient to ensure $s = 0$ (Theorems 1 and 2), though they are not all always necessary. With each assumption we give an intuitive interpretation of how it contributes to ensuring $s = 0$, or why its violation could cause $s \neq 0$. These interpretations should contribute to a better understanding of modeling dynamic network flows.

The Cost Function Assumptions

We now introduce

Cost function assumption CFA: For each path to the destination containing arcs $j_1$ and $j_2$ with arc $j_3$ nearer to the destination,

$$\frac{dh_{ij}(y)}{dy} \leq \frac{dh_{ij}(z)}{dz}$$

for all $y > 0$, $z > 0$ and all $t$. If arc $j_2$ points into the destination node, then the left-hand side of (12) is taken as zero, which implies, from (12), that $h_{ij} \geq 0$ for all $t$ and $j$.

CFA is satisfied if costs are proportional to time, as is common in static models. In that case we have $h_{ij}(x_{ij}) = klx_{ij}$ for all $t$ and $j$, where $l$ is the length of each time period and $k$ is a constant. Then $h'_{ij} = kl$, a constant, and (12) holds as a strict equality for all $j$ and $t$.

This CFA is a restatement of CFA(b) from M-N (1978a)—we do not include their assumption that $h_{ij}(x)$ is convex. If $h_{ij}(x)$ has right and left derivatives $(dh_{ij}/dx$ and $dh_{ij}/dx)$ which differ, then (12) holds for both derivatives; for example, $h_{ij}(x)$ may be piecewise linear.

Stated less formally, CFA requires that as flows move along any path to the destination, the marginal
cost per unit volume, per unit time should not increase on moving from one arc to the next. This condition provides a cost incentive to keep arc volumes moving towards the destination without delays (without flow controls), since at the destination the flows will exit from the network and will thereby cease incurring travel costs. This interpretation of CFA applies also at the destination node: if \( h_{ij} < 0 \) for an arc pointing into the destination node, then there would be a cost incentive to hold back traffic from exiting at the destination.

We now slightly strengthen the cost function assumption, as follows.

*Strict CFA for the terminal period, \( T \),* is defined by letting the inequality in CFA be a strict inequality for period \( T \).

This assumption is very reasonable. In the terminal period, an additional cost component is needed to account for the fact that the volume \( x_{ij} \), which is still on arc \( j \) in the terminal period, will incur further costs beyond the terminal period before it reaches the destination. An estimate of these residual costs should be included in the terminal period costs, and the costs will be higher the further one is from the destination—which is precisely what the strict CFA assumption states. Handling terminal period costs is a common problem in finite horizon economic models.

An intuitive explanation of why strict CFA is needed for period \( T \), if we are to ensure that the optimal flow controls will be zero, is as follows. First recall that, due to (4c), flows \( d_{ij} \) entering arc \( j \) in period \( t \) cannot exit from \( j \) in the same period, so that it takes at least one period to traverse each arc. Now consider a period \( t \) which is, say, \( m \) periods before the terminal period, and consider an arc \( j \) which is at least \( m \) arcs away from the destination, along all paths through \( j \) to the destination. Clearly, none of the volume \( x_{ij} \) on this arc will be able to reach the destination before the terminal period; hence we will refer to \( x_{ij} \) in this case as being “inaccessible” to the destination. If (12) happened to be a strict equality for all inaccessible arcs, then there would be no incentive for the volumes on these arcs to proceed toward the destination. This situation would cause slacks in the flow control constraint (4b), i.e., nonzero flow controls. The strict CFA assumption provides just sufficient incentive to prevent this outcome, and to keep flows on inaccessible arcs moving toward the destination.

**Exit Function Assumptions**

Turning now to the congestion functions \( g_j(x) \), we introduce an assumption EFA which, together with the cost function assumptions, will be sufficient to ensure that the optimal flow controls are zero valued.

**Exit function assumption (EFA):** We will say that \( g_j(x) \), for all \( j \in A \), satisfies EFA if, for all \( j \),

\[
\begin{align*}
(a) & \quad x > g_j(x) > 0 \quad \text{for} \quad x > 0, \\
(b) & \quad 1 > g_j'(x) > 0 \quad \text{for} \quad x > 0, \\
(c) & \quad g_j(0) = 0
\end{align*}
\]

The property \( x > g_j(x) \) simply states that the exit rate should not exceed the volume on the arc. To motivate assumption \( g_j'(x) > 0 \), note that this condition restricts \( x \) to the upward sloping part of the \( g(x) \) curve in Figure 1(iii), and thus rules out the “overcongestion” or “oversaturation” associated with a downward sloping \( g(x) \). (With a downward sloping \( g(x) \), flow controls could be used to increase outflows \( g(x) \) by reducing \( x \).)

Finally, consider the assumption \( g_j'(x) < 1 \). It is easy to show that this condition holds if \( g_j(x) \) is concave, given the other assumptions \((x > g_j(x)\) and \(g_j(0) = 0\)). But \( g_j'(x) < 1 \) does not require concavity of \( g_j(x) \). These properties are illustrated in 3 examples in Figure 1: examples (i) and (ii) satisfy EFA, and (iii) does not satisfy EFA but satisfies EFA, of the Appendix and Section 6. Example (ii) shows that EFA does not imply concavity of \( g_j(x) \).

![Figure 1. Illustrative flow capacity functions.](image-url)
6. Sufficient Conditions for Optimal Flow Controls to Be Zero

Theorems 1 and 2 state our main results concerning conditions under which the optimal flow controls are zero. In the corollaries following Theorem 1, we establish some properties of the equality-constrained non-convex model P of M-N. These results enable us to greatly strengthen the results of M-N.

**Theorem 1.** Suppose that the model C satisfies a constraint qualification (sufficient conditions are given in Lemmas 1 and A1) and the EFA and CFA conditions, with the inequality in CFA being strict for the terminal period T. Then in any optimal solution of program C, the optimal flow controls will be zero for all arcs and all periods.

**Remark.** Note that this theorem does not require convexity of any \( h_\alpha(x) \) or \( g_\alpha(x) \) term. Section 5 showed that CFA implies that \( h_\alpha(x) \) is nondecreasing, hence \( h_\alpha(x) \) is both quasiconvex and quasiconcave, but this situation does not ensure that the sum of \( h_\alpha(\cdot) \)'s in the objective function of P, is quasiconvex or quasiconcave.

**Proof.** Lemma 1 ensures that the Kuhn-Tucker conditions (A.3a–A.3f) hold at any optimum of C. Lemma A5 in the Appendix then applies and states that the constraints (4b) are all strict equalities, which completes the proof.

**Corollary 1.** Suppose that the assumptions of Theorem 1 apply. Then the (global optimal) solution (set) of program C is identical to the global optimal solution (set) of program P.

**Proof.** Since C is a convex program, all its solutions are global optimal solutions. From Theorem 1, (4b) is an equality in any optimal solution of C, hence imposing (4b) as a strict equality constraint in C cannot alter the global optimal solution set of C. But this outcome reduces C to P, hence the corollary follows.

**Corollary 2.** Suppose that the assumptions of Theorem 1 apply. Then the K-T conditions for program P hold at any global optimum of programs C or P.

**Proof (outline).** Let K-Tc and K-Tp denote the K-T conditions for programs C and P, respectively. By algebraic manipulation we can show that K-Tp, in M-N (1978b), holds when K-Tc holds: using the inequalities (A.3b) to eliminate \( \alpha_j \) from (A.3a) and (A.3d) then reduces K-Tc to K-Tp, except for the additional constraint (A.3b) in K-Tc; hence K-Tp holds when K-Tc holds. But K-Tc holds at any optimum of C (Lemma 1), hence so will K-Tp. Also, by Corollary 1, programs C and P have the same set of global optimal solutions, hence K-Tc holds at any global optimum of P.

An implication of Corollary 2 is that the results in M-N (1978b) are now strengthened. In that paper, the authors use the K-T conditions for program P to derive some interesting characterizations of the solution(s) of program P—see in particular their Lemma 1 and Theorems 1 and 2. However, their results depend on an assumption that the K-T conditions for P hold at an optimum of P: they did not prove this result due to the difficulty of obtaining a constraint qualification. But Corollary 2 now shows that, when the conditions specified in Theorem 1 are satisfied, the results in M-N (1978b) hold unequivocally, for any global optimal solution of the M-N program P, and for any optimal solution of our program C.

Theorem 1 gives sufficient conditions for the optimal flow controls to be zero for all arcs of the network. However, if the assumptions CFA or EFA are violated for even one arc of the network, then Theorem 1 as it stands tells us nothing. We will therefore state a much stronger theorem giving sufficient conditions for the optimal flow controls for a single arc to be zero, even if some or all of the optimal flow controls elsewhere on the network are nonzero. To obtain this result, we break assumptions CFA and EFA down into separate assumptions CFA_{ij} and EFA, for each arc: these assumptions are defined at the beginning of the Appendix. Then we have the following result.

**Theorem 2.** Assume program C satisfies a constraint qualification (Lemmas 1 and A1 give sufficient conditions for this result), and let EFA_{ij}, CFA_{ij}, \( i = 0, \ldots, T-1 \), and strict CFA_{Tj} hold for arc j. Then in any optimal solution of C, the flow controls will be zero for arc j for all periods \( t = 0, \ldots, T-1 \).

**Proof.** Lemma 1 ensures that the Kuhn-Tucker conditions (A.3a–A.3f) are satisfied at any optimum of C. The theorem then immediately follows from Lemma A5, which states that the constraints (4b) are all strict equalities for arc j.

If we think of \( g_j'(x) < 0 \) as indicating saturation or “overcongestion” on arc j, notice the EFA_{ij} allows arc j to be saturated but requires that arcs immediately following arc j should not be saturated. Also notice that Theorem 2, like Theorem 1, does not require convexity of any cost function or exit function, unless the latter is needed in order to ensure a constraint qualification. Finally, it is obvious that Theorem 1 is
equivalent to applying Theorem 2 simultaneously to all arcs.

7. Flexible Departure Times and Elastic Demands

One of the interesting ways to extend model C is to allow flexible departure times by letting the origin-destination demands be determined within the model rather than be given exogenously by $F_{ik}$, as in program C. This situation can be achieved in a number of ways. One is to replace the constants $F_{ik}$ in (4d) with variables $f_{ik}$, for some or all nodes, $\sum_{t=0}^{T} f_{ik} = F_{ik}$ for those nodes, where $F_{ik}$ is the exogenously given aggregate demand at node $k$ over the time span of the model. We will refer to this reformulation below as model $C^1$.

An alternative way to introduce variable demands is similar to the treatment of elastic demands in static network models. Let the demand for flows from origin $k$ in period $t$ to the destination be $f_{ik} = f_{ik}(p_{ik})$ where $p_{ik}$ is the unit cost of such flows. The inverse of these demand functions can be written as $p_{ik} = p_{ik}(f_{ik})$, for all $t$, $k$, and the areas under these inverse demand functions can be taken as a measure of the gross benefits derived from the flows $f_{ik}$. Thus the gross benefit from all trips $f = \{ f_{ik} \}$ is

$$ b(f) = \sum_{i=0}^{T} \sum_{k \in N} \int_{0}^{f_{ik}} p_{ik} \left( f_{ik} \right) df_{ik}. $$

By analogy with the static convex cost network model with elastic demands, the dynamic system optimizing network model with elastic demands is now,

Model $C^2$

Maximize $b(f) - \sum_{i=0}^{T} \sum_{j \in A} h_{ij}(x_{ij})$

subject to (4b)-(4f),

with the $F_{ik}$'s in (4d) replaced by variables $f_{ik}$, and $f_{ik} \geq 0$ for all $t$ and $k$.

This model $C^2$ can easily be further extended to allow for cross-elasticities of demand, by using $f_{ik} = f_{ik}(p_{0i}, \ldots, p_{Ri})$, $p_{ik} = p_{ik}(f_{0i}, \ldots, f_{Ri})$, and the line integral in place of the given forms. As in the case of static models, a meaningful solution requires that the line integral be path independent which requires in turn that the Jacobian matrix formed from the system of demand functions be symmetric. Concavity of the objective function further requires that the Jacobian matrix be negative (semi) definite.

A question which now arises is, do models $C^1$ and $C^2$ have all the desirable properties (mutatis mutandis) specified earlier for $C$? The answer is yes. To see this, note that:

(a) $C^1$ and $C^2$ are convex programs.
(b) The constraint sets of $C^1$ and $C^2$ still satisfy Lemmas A1 and 1.
(c) The proofs of Lemmas A2–A5 rely on (4b)–(4f) and on the Kuhn-Tucker conditions (A,3a–A,3f).

But these are included in the constraints and Kuhn-Tucker conditions for $C^1$ and $C^2$, hence Lemmas A2–A5 hold for $C^1$ and $C^2$.

Note that, from (a)–(c), Theorems 1 and 2 from the previous section also hold for $C^1$ and $C^2$; or more constructively, we can show this result as follows. Theorems 1 and 2 hold for any given finite nonnegative value of $\{ F_{ik} \}$, hence in particular they will hold when $\{ F_{ik} \} = \{ f_{ik} \}$, where the latter is taken from any optimal solution of $C^1$ or $C^2$. Thus Theorems 1 and 2 hold for programs $C^1$ and $C^2$.

8. Some Computational Experience

The programs set out in Sections 1, 3 and 4 are convex, with the exception of one of the formulations presented in Section 4; hence with this exception, they can be solved using any of a variety of existing algorithms for solving convex nonlinear programs. It is possible, and will be of interest, to develop faster algorithms that take advantage of the special structure of these problems, including the network type subset of constraints and the intertemporal staircase structure of the constraints. But this development is beyond the scope of this paper. In our experiments the algorithm we have used has consisted of piecewise linearizing the nonlinear constraints and objective function and solving using the linear programming option of the MINOS program (Murtagh and Saunders 1977) operating on a DEC-20 computer. In this way we have solved various of the formulations for small test networks using a variety of data sets. Some of this work is reported in Carey and Srinivasan (1982), which also reports (a) experiments comparing the solution of the programs with the solutions of existing static models applied to the same networks, and (b) experiments using the programs to estimate the benefits to be obtained from introducing "flextime" for workers in a congested urban area.

Here we report only one experiment, which was conducted in order to illustrate and confirm the theorems concerning zero valued optimal flow controls presented in Section 6. In this example we used a piecewise linear version of program C to solve a small network problem given in Ho. We adapted the prob-
lem slightly so as to satisfy the assumptions CFA and EFA of Section 5.

In this problem there are 7 nodes, 12 arcs, 10 time periods and each piecewise linear exit function and cost function has three pieces. Also, \( F_{ik} = 30, t = 1, \ldots, 9, k = 1, \ldots, 6; E_{j} = 0.001, j = 1, \ldots, 12; \) and \( h_{t}(x_{t}) = 1.00x_{t}, t = 1, \ldots, 10. \) This \( h_{t}(x_{t}) \) satisfies CFA, except for the terminal period \( T. \) To ensure the later, we let the marginal costs \( h'_{T} \) on successive arcs along each path differ by amounts of the order of 0.001. For example, the arcs 1, 2, 7 and 12 form a path to the destination, and for these we let the costs \( h_{T}(x_{T}) \) be 1.005\( x_{T1} \), 1.004\( x_{T2} \), 1.003\( x_{T7} \), and 1.00\( x_{T12} \), respectively.

Since the exit function is piecewise linear, EFA and concavity of all \( g_{j}(x) \) implies \( 1 > g_{j,p} \geq g_{j,p+1} \geq 0 \) for all \( j \in A \) and \( p = 1, \ldots, P, \) where \( g_{j,p} \) is the slope of the \( p \)th piece of the piecewise linear \( g_{j}(x). \) To satisfy this condition we let \( 0.999 > g_{j,p} > g_{j,p+1} > 0.001. \)

9. Further Research

This paper has considered a model, or models, of time-varying network flows in which the time taken to traverse each arc depends on the flow or volume on that arc. Research in formulating, analyzing and solving such models is still at a very early stage of development as compared with work on static network flows. As a contribution to this research we plan to extend the report work in several directions. Some of these are as follows.

First, the models need to be applied and further explored for the various types of networks mentioned in the introduction, including networks for road traffic, communications, commodity and financial flows, social interaction and public service.

Second, the models can be extended for use in investment planning and network design for congested networks like that presented in Dantzig et al. (1979) and elsewhere for static convex cost network models. Thus, by introducing arc capacities as variables, the model can be used to determine the optimal levels of investment for improving or expanding existing arcs and for constructing new arcs.

Third, the present paper has considered only models that represent a system optimum. However, in the case of road traffic there is also a need for dynamic network models that represent a “user equilibrium” or “descriptive assignment” (Florian 1984), since this should give a better description of how traffic behaves where there is no central traffic controller or toll setter. This extension requires that we develop an inter-temporal formulation of the static user equilibrium criterion (Wardrop’s (1952) first principle), in terms of path traversal times that vary over both space and time. In this connection it should be interesting to use the dynamic model to arrive at optimal tolls, that is the tolls that would just persuade road users to behave in a system optimal fashion. These tolls can be defined by taking the solution of the system optimizing model and comparing it with the criteria for a dynamic user equilibrium. Our preliminary theoretical results show that the dynamic tolls can differ dramatically from the traditional static tolls, and that they are quite different when traffic is building up to a peak than when it is falling off again.

Finally, the usefulness of the present and proposed research would be enhanced by the development of specialized algorithms to exploit the special structure of the problems. As mentioned in Section 8, this special structure includes the network flow conservation subset of constraints and the intertemporal staircase structure of the constraints.

Appendix

Lemmas A1 and A5 are used in proving theorems in Section 6, and Lemmas A2–A4 are used in proving Lemma A5. Proofs of Lemmas A1–A4 are somewhat lengthy and are omitted for brevity—they can be found in Carey (1984), which is available from the author.

In order to derive properties of flows on each arc \( j \) for each time period \( t \) for program \( C \) we subdivide the assumptions CFA and EFA, from Section 5, into the following separate assumptions for each arc \( j \) and each period \( t. \)

Cost function assumption, CFA\(_{ij} \): We will say that \( h_{ij}(x) \) satisfies CFA\(_{ij} \) if

\[
\frac{dh_{ij}(y)}{dy} \leq \frac{dh_{ij}(z)}{dz} \tag{A.1}
\]

for all \( y \geq 0, z > 0, \) and all arcs \( j \) pointing out of the node \( k \) into which arc \( j \) points. When \( k \) is the destination node, the left-hand side of (A.1) is taken as zero. We will say that strict CFA\(_{ij} \) holds if the inequality (A.1) is strict.

Exit function assumption, EFA\(_{ij} \): We will say that \( g_{j}(x) \) satisfies EFA\(_{ij} \) if \( x > g_{j}(x) > 0 \) and \( 1 > g_{j}(x) \) when \( x > 0, \) and \( g_{j}(x) > 0 \) for all arcs \( j \) pointing out of the node \( k \) into which arc \( j \) points.

Note that CFA = (CFA\(_{ij} \), for all \( j \) and \( t \)) and EFA = (EFA\(_{ij} \), for all \( j \)).
Lemma A1. Let $E_j > 0$, for all $j$. Let $g_j(x)$ be concave and $x \geq g_j(x) > 0$, for all $x > 0$ and for all $j \in A$. Then the constraint set of $C$ satisfies the generalized Slater constraint qualification. (See for example Mangasarian 1969.)

Lemma A2. Let $EFA_j$ hold. Then in any feasible solution of (4b)-(4f),

(a) if $x_{tj} > 0$ then $(x_{tj} > 0, t = t^*, \ldots, T - 1)$
(b) if $x_{tj} = 0$ then $(x_{tj} = 0, t = 0, \ldots, t^*)$.

In the following lemmas (A3–A5) we derive some properties of the Kuhn-Tucker conditions corresponding to program $C$ of Section 1. These Kuhn-Tucker conditions can be stated as in (A.3a)-(A.3f). Let $\alpha_{ij}, \beta_{ij}$ and $\mu_k$ be Lagrange multipliers corresponding to constraints (4b), (4c) and (4d) respectively. Then, for each period $t = 0, \ldots, T - 1$.

$$u_{tj} = \alpha_{ij}g_j - \beta_{ij} + \beta_{t-1,j} - h_{ij} \leq 0$$

for all $j \in A$, $t \neq 0$ (A.3a)

$$u_{tj} = \alpha_{ij}g_j - \beta_{t-1,j} - h_{ij} \leq 0$$

for all $j \in A$ (A.3b)

$$v_k = -\alpha_{ij} + \beta_{ij} - \mu_k \leq 0$$

for all $k \in N'$, for all $j \in B(k)$ (A.3c)

$$w_j = -\beta_{ij} - \mu_k \leq 0$$

for all $k \in N'$, for all $j \in A(k)$ (A.3d)

$$\alpha_{ij} \leq 0$$

for all $j \in A$ (A.3e)

and the complementarity conditions for the inequalities (4b, A.3a–A.3c), thus

$$\alpha_{ij}(b_{ij} - g_j(x_{ij})) = 0$$

$$x_{ij}u_{ij} = 0, \quad b_{ij}v_{ij} = 0, \quad d_{ij}w_{ij} = 0$$

for all $j \in A$ (A.3f)

where $g_j'$ denotes $dg_j(x)/dx$ when $x = x_{ij}$

Lemma A3. Let $EFA_j$ and strict $CFA_{ij}$ hold. Then $\alpha_{t-1,j} > 0$ and/or $x_{t-1,j} = 0$, for all $j \in B(k)$ in any solution of (A.3a–A.3f).

Lemma A4. Let $CFA_{ij}$ and $EFA_j$ hold, and let $\alpha_{ij} > 0$. Then $\alpha_{t-1,j} > 0$ and/or $x_{t-1,j} = 0$ in any solution of (A.3a–A.3f).

Lemma A5. Let $EFA_j$, $CFA_{ij}, t = 0, \ldots, T - 1$, and strong $CFA_{Tj}$ hold. Then in any solution of (A.3a–A.3f), there is a $t_j$ such that

(a) $b_{tj} = g_j(x_{tj}) > 0$ and $\alpha_{ij} > 0$, for all $t = t_j, \ldots, T - 1$.

(b) $b_{tj} = g_j(x_{tj} = 0$ and $\alpha_{ij} > 0$, for all $t = 0, \ldots, T - 1$.

Remark. (a) and (b) ensure that the constraints (4b) are all strict equalities for arc $j$. Thus the constraints (4b) will be strict equalities for all arcs $j \in A$ if EFA and CFA hold: to show this result, note that by definition $EFA = \{EFA_j, \text{ for all } j\}$ and $CFA = \{CFA_{ij}, \text{ for all } j, t = 0, \ldots, T - 1\}$ and apply Lemma A5 for all arcs $j \in A$.

Proof. Consider $t = t_j, \ldots, T - 1$. Lemma A2(a) and $x_{tj} > 0$ implies $x_{tj} > 0$, $t = t_j, \ldots, T$. Then Lemma A3 yields $\alpha_{t-1,j} > 0$, hence repeated application of Lemma A4 yields $\alpha_{t-1,j} > 0, t = t_j, \ldots, T - 1$. This result implies (from complementarity of $\alpha_{ij}$ and (4b)) that $b_{ij} = g_j(x_{ij})$, $t = t_j, \ldots, T - 1$. But $x_{ij} > 0$ implies $g_j(x_{ij}) > 0$, from EFA, and $x_{ij} = 0, t = 0, \ldots, t - 1$, imply $g_j(x_{ij}) = 0, t = 0, \ldots, t - 1$, hence (4b) yields (b).

Acknowledgment

I wish to thank Ashok Srinivasan, a Ph.D. student at Carnegie-Mellon University, who performed the computation for Section 8. I also wish to thank two anonymous referees for their helpful comments and suggestions.

This work was supported by the Urban Mass Transportation Administration under grant PA-06-0063, and is based on two earlier unpublished papers by the author (Carey 1980a,b).

References


