COMPUTATION OF DYNAMIC USER EQUILIBRIA IN A MODEL OF PEAK PERIOD TRAFFIC CONGESTION WITH HETEROGENEOUS COMMUTERS

Donald K. Richter, John Griffin, and Richard Arnott

(April, 1990)

Working Paper No. 198

Preliminary Draft: Please do not cite or quote without permission of the authors.

1. INTRODUCTION

This paper has two legacies: the literature on fixed point algorithms (for example, SCARF[17] and TODD[19]) and the literature on dynamic user equilibrium in models of congestible facilities, in particular models of peak period traffic congestion (for example, ARNOTT et al.[2], BEN-AKIVA et al.[6], HENDERSON[10], HENDRICKSON AND KOCUR[11], MAHMASSANI AND HERMAN[12], SMITH[18], and VICKREY[21]). We draw on the first legacy to generalize some of the models from the second legacy. We study dynamic user equilibria in a deterministic model of peak period traffic congestion with heterogeneous commuters. Commuters may choose from different routes and modes having different performance characteristics, and may also trade off travel time and schedule delay by choosing different departure times. Computation of user equilibria with a fixed point algorithm permits treatment of models involving a larger number of routes, modes, and user types, and more general user cost functions, than much of the literature referenced above.

Fixed point algorithms, while widely used in economics and other areas of operations research (see, eg., SCARF[17] and TODD[19]), have had a negligible impact on the transportation literature.[1] Computational time of fixed point algorithms increases at least with the cube of the dimension of the set over which search is being conducted, and equilibrium models of transportation networks tend to have a large number of variables. We show in this paper how some natural special structure of user cost functions can be exploited to avoid these dimensionality problems.

Section 2 describes the base model, Section 3 provides a constructive existence proof of a dynamic user equilibrium in the base model, and Section 4 gives a numerical example of the algorithm involving bottleneck congestion. Section 5 views the computational routine as a decomposition procedure for nonlinear complementar-

---

[1]See DAFERMOS[8], MAGNANTI[13], and NAGURNEY[14] for a review of algorithms that have been used to compute traffic equilibria. Most of these models do not endogenize the time of departure decision.

ity problems. This view illuminates how we have been able to avoid the dimensionality problem inherent in fixed point algorithms. Section 6 generalizes the base model to include multiple origins and destinations, and the possibility that a commuter chooses not to make a trip. We also discuss the feasibility of generalizing the structure of the cost function to permit treatment of more realistic networks.

TABLE 1
Summary of Notation
(vectors are in boldface)

R	index set for routes
G	index set for groups
T	index set for time intervals
$\lvert R\rvert, \lvert G\rvert, \lvert T\rvert$	number of elements in the sets R, G, T respectively
N_g	number of commuters in group g
N	total number of commuters
d_{grt}	number of departures from group g on route r during interval t
d_{rt}	$\sum_{g\in G} d_{grt}$, the total departures across all groups on route r during interval t
$\mathbf{d}$	$(d_{grt})_{g\in G, r\in R, t\in T}$
C_{grt}	average cost incurred by a type g commuter who departs on route r during interval t
c_g	minimum average cost for a type g commuter
$\mathbf{c}$	$(c_g)_{g\in G}$
$E(\mathbf{c})$	image set of the excess commuter correspondence
$\delta(\mathbf{c})$	image set of the subproblem solution correspondence

2. THE BASE MODEL

There are a finite number of *routes*, indexed by $r\in R$, connecting a single origin (a residence location) to a single destination (the Central Business District (CBD)). Different routes may correspond to different modes of transportation. There are a finite number of homogeneous *groups* of commuters, indexed by $g\in G$. N_g is the number of commuters of group g, with $N\equiv \sum_{g\in G} N_g$. Each commuter makes 1 trip between the origin and destination. There are a finite number of *departure time intervals*, indexed by

$t \in T \equiv \{1,\ldots,|T|\}$.[2] $|G|$, $|R|$, and $|T|$ denote the number of elements in G, R, and T respectively.

Let d_{ij} denote the *total* number of commuters departing on route i (across all groups) during interval j. For any 3-tuple (g,r,t) we define the cost function $C_{grt}: \Re_+^t \rightarrow \Re_+$, where $C_{grt}(d_{r1},\ldots,d_{rt})$ denotes the average trip cost incurred by a commuter from group g who departs on route r during interval t.[3] C_{grt} summarizes the cost of travel time, schedule delay, and the like. The nature of traffic congestion is assumed to be embedded in these cost functions. The dependence of C_{grt} on $(d_{r1},\ldots,d_{rt})$ implies that congestion is anonymous, dependent on the number of users, but not their type. The lack of dependence of C_{grt} on the number of departures on routes other than r in any time interval makes it convenient to interpret the base model as one involving parallel roads. The lack of dependence of C_{grt} on $d_{i\tau}$ (for all $i \in R$ and $\tau > t$) is crucially exploited in the algorithm.[4] The indexing of the cost functions affords considerable flexibility in interpretation. The presence of the g index permits treatment of commuters with different costs of travel time, desired arrival times at the CBD, and costs of early or late arrival. The presence of the r index allows for different route

---

[2]For example, assumptions can be made which guarantee that the morning rush hour starts no earlier than 6 AM and ends no later than 10 AM. We can then partition this 4 hour period into a finite number of uniformly sized intervals. The larger the number of intervals, the more closely the model will approximate a continuous time model. Each commuter then chooses to depart for work during one of these intervals.

[3]In a model involving congestion and discrete time intervals, it may not be true that all commuters of a particular group that depart in the same time interval incur the same actual cost. C_{grt} represents the average of these actual costs. See Section 4 for an example. The number of departures d_{ij} is treated as a continuous variable. Under the assumption that the N_g are large, this causes no difficulty with the notion of approximate equilibrium discussed in Section 4.

[4]In Section 6 we consider generalizations involving an expansion of the list of arguments of C_{grt}.

(including mode) attractiveness, congestion characteristics and the like. The presence of the t index permits treatment of the important dynamic aspects of the problem.

Assumptions

(A.1) C_{grt} is continuous.

(A.2) For fixed values of $(d_{r1},\ldots,d_{rt-1})$, $C_{grt}(d_{r1},\ldots,d_{rt-1},d'_{rt}) > C_{grt}(d_{r1},\ldots,d_{rt-1},d''_{rt})$ if $d'_{rt} > d''_{rt}$.[5] (We say C_{grt} is *increasing* in d_{rt}.)

Each commuter is assumed to pick a route and a departure time interval, conditional on assumed departures of other commuters, so as to minimize cost. Let d_{grt} denote the number of commuters of group g who depart on route r during interval t.

DEFINITION. $(\mathbf{d}^*,\mathbf{c}^*) \geq \mathbf{0}$*, where* $\mathbf{d}^* \equiv (d^*_{grt})_{g\in G, r\in R, t\in T}$ *(with* $d^*_{rt} \equiv \sum_{g\in G} d^*_{grt}$*) and* $\mathbf{c}^* \equiv (c^*_g)_{g\in G}$ *is a dynamic user equilibrium if for each* $(g,r,t) \in G\times R\times T$:

(i) $\sum_{i,j} d^*_{gij} = N_g$;

(ii) $C_{grt}(d^*_{r1},\ldots,d^*_{rt-1},d^*_{rt}) = c^*_g$ *if* $d^*_{grt} > 0$;

(iii) $C_{grt}(d^*_{r1},\ldots,d^*_{rt-1},d^*_{rt}) \geq c^*_g$ *if* $d^*_{grt} = 0$.

Condition (i) simply requires that all commuters make the trip from the residence location to the CBD. c^*_g represents the minimum average cost that a commuter of group g incurs, given the departure pattern summarized by $\mathbf{d}^*$. Conditions (ii) and (iii) insure that each commuter's choice of route and departure time interval is consistent with cost minimization, given the choices of all other commuters. This equilibrium notion embodies WARDROP'S[22] principle—all routes utilized by a given group will

---

[5] While it would be more plausible to either restrict the domain of C_{grt} to a subset of $\Re^t_+$, or to assume that (A.1) and (A.2) hold only on a subet of $\Re^t_+$, for simplicity we assume (A.1) and (A.2) hold on all of $\Re^t_+$. We could restrict these assumptions to the subset $\{\mathbf{s} \in \Re^t_+ \mid \mathbf{s} \leq (N+1,\ldots,N+1)\}$ without affecting any of the results in this paper.

have the same travel cost, which must be minimal among travel costs on all routes for that group.

3. COMPUTING AN APPROXIMATION TO A USER EQUILIBRIUM

The basic idea behind the computational routine is similar to an idea exploited by ARNOTT AND MACKINNON[4] in a different context. Arnott and MacKinnon were concerned with allocating consumers to spatial locations in an urban area. Conditional on a certain vector of parameters, they allocated consumers to the various locations so as to build in certain equilibrium requirements (such as utility maximization and the equality of supply and demand for land in each location). At the end of their allocation process they compared the number of consumers actually allocated with the number that had to be allocated. If these numbers differed significantly, they systematically adjusted the vector of parameters and repeated the allocation process. Our algorithm in the present context embodies the same spirit. Conditional on a certain vector of parameters, instead of allocating consumers to spatial locations the algorithm allocates commuters to departure time intervals and routes so as to essentially build in the cost minimization requirements of equilibrium (conditions (ii) and (iii)). Then it compares the number of commuters (of each group) that are actually allocated with the number that must be allocated. If these numbers differ significantly (i.e., if condition (i) is not approximately satisfied), the vector of parameters (which corresponds to a |G|-vector of minimized costs) is systematically varied and the process is repeated. The algorithm stops when condition (i) is approximately satisfied.

The formal algorithm is explained by:

(1) Describing the creation of the *excess commuter correspondence*, whose domain is a set of candidate equilibrium minimized cost vectors $\mathbf{c} = (c_g)_{g \in G}$. Satisfaction of conditions (ii) and (iii) of equilibrium is built into this correspondence.

(2) Showing a zero $\mathbf{c}^*$ of this correspondence corresponds to a dynamic user equilibrium.

(3) Describing how a fixed point algorithm can be used to approximate a zero of this correspondence. (The dimensionality of the domain of the correspondence equals |G|, and is independent of |R| and |T|.)

The excess commuter correspondence

The creation of the excess commuter correspondence $E:\Re_+^{|G|} \to \Re^{|G|}$ is illustrated in Figure 1. Given an arbitrary vector $\mathbf{c} \in \Re_+^{|G|}$, we first define d_{r1}, the total number of commuters that are assigned to depart on route r in interval 1.[6] Making this assignment for each $r \in R$ yields the unique vector $(d_{r1})_{r \in R}$ of total departures in interval 1. Armed with this vector we then determine uniquely d_{r2}, the total number of commuters assigned to depart on route r during interval 2. After calculating d_{r2} for each $r \in R$, we will have in hand a unique vector $(d_{rj})_{r \in R, j \in \{1,2\}}$ representing the total departures on the various routes during intervals 1 and 2. Continuing this sequential process enables us to determine a unique vector $(d_{rj})_{r \in R, j \in T}$ which summarizes the number of commuters which are assigned to depart on route r during interval j for all routes $r \in R$ and all intervals $j \in T$. Note that this sequential procedure exploits crucially the property that C_{grt} does not depend on the number of departures on any route in

---

[6] If the first branch of d_{r1} is operative, it follows (since C_{gr1} is increasing in v) that $C_{gr1}(v) < c_g$ for all $v \in [0,N+1]$, verifying that the 2 branches are mutually exclusive. The continuity of C_{gr1} insures that d_{r1} is well defined if the second branch of the definition is operative. If the first branch is operative, there is some group g whose cost on route r cannot be driven up to the proposed equilibrium level c_g, even if more than the entire population of commuters uses that route. Obviously such a level c_g is not a viable equilibrium candidate, and so we assign more than the total population of commuters to that route, which insures that the algorithm cannot possibly converge to such a cost vector **c**. If the second branch is operative, then the total number of commuters placed on route r insures that no group incurs lower cost on that route than its proposed equilibrium cost, which at least allows for the possibility that **c** could be a vector of equilibrium costs.

$$\mathbf{c} = (c_1, \ldots, c_{|G|})$$

$$\downarrow$$

$$d_{r1} \equiv \begin{cases} N+1 & \text{if } \exists\, g \text{ with } C_{gr1}(N+1) < c_g \\ \min \{v \in [0, N+1] \mid C_{gr1}(v) \geq c_g \text{ for all } g\} & \text{otherwise} \end{cases}$$

$$\downarrow$$

$(d_{r1})_{r \in R}$ total departures in interval 1

$$\downarrow$$

$$d_{r2} \equiv \begin{cases} N+1 & \text{if } \exists\, g \text{ with } C_{gr2}(d_{r1}, N+1) < c_g \\ \min \{v \in [0, N+1] \mid C_{gr2}(d_{r1}, v) \geq c_g \text{ for all } g\} & \text{otherwise} \end{cases}$$

$$\downarrow$$

$(d_{rj})_{r \in R, j \in \{1,2\}}$ total departures in intervals 1 and 2

$$\downarrow$$

$$\vdots$$

$$\downarrow$$

$(d_{rj})_{r \in R, j \in \{1, \ldots, t-1\}}$ total departures in intervals 1,…,t–1

$$\downarrow$$

$$d_{rt} \equiv \begin{cases} N+1 & \text{if } \exists\, g \text{ with } C_{grt}((d_{rj})_{j \in \{1, \ldots, t-1\}}, N+1) < c_g \\ \min \{v \in [0, N+1] \mid C_{grt}((d_{rj})_{j \in \{1, \ldots, t-1\}}, v) \geq c_g \text{ for all } g\} & \text{otherwise} \end{cases}$$

$$\downarrow$$

$(d_{rj})_{r \in R, j \in \{1, \ldots, t\}}$ total departures in intervals 1,…,t

$$\downarrow$$

$$\vdots$$

$(d_{rj})_{r \in R, j \in T}$ total departures in all intervals

$$\downarrow$$

$$D(\mathbf{c}) \equiv \{(d_{grt})_{g \in G, r \in R, t \in T} \in \Re_+^{|G||R||T|} \mid \sum_{g \in G} d_{grt} = d_{rt}, \text{ with } d_{grt} > 0 \text{ only if } C_{grt}(d_{r1}, \ldots, d_{rt}) \leq c_g\}$$

$$\downarrow$$

$$E(\mathbf{c}) \equiv \{(N_1 - \sum_{r,t} d_{1rt}, \ldots, N_{|G|} - \sum_{r,t} d_{|G|rt}) \in \Re^{|G|} \mid (d_{grt})_{g \in G, r \in R, t \in T} \in D(\mathbf{c})\}$$

Fig. 1. Creation of the excess commuter correspondence

any interval subsequent to interval t.

Having determined the unique total departure vector $(d_{rj})_{r\in R, j\in T}$ we then define a collection of departure patterns D(**c**) which are consistent with this total departure vector, and which satisfy a certain cost constraint. It is straightforward to argue that D(**c**) is nonempty.[7] Each element of D(**c**) specifies the number of commuters from each group that depart in each interval on each route. A commuter from group g is allowed to depart on route r in interval t only if his cost would be no greater than the coordinate c_g of the given vector **c**. Each departure pattern vector $(d_{grt})_{g\in G, r\in R, t\in T}$ generates an element of E(**c**). The gth coordinate of such an element represents the difference between the size of group g and the number of commuters from group g that depart on some route during some time interval in the underlying departure pattern vector.

The algorithm finds an approximate zero of E. The definition of E is designed to insure (see the proof of Theorem 3.1) that at a zero of E, conditions (ii) and (iii) of equilibrium are satisfied.

THEOREM 3.1. *Let* $\mathbf{0} \in E(\mathbf{c}^*)$, *where* $\mathbf{0} = \left(N_g - \sum_{r,t} d^*_{grt}\right)_{g\in G}$. *Then* $(\mathbf{d}^*, \mathbf{c}^*)$, *where* $\mathbf{d}^* = (d^*_{grt})_{g\in G, r\in R, t\in T}$, *is a dynamic user equilibrium.*

---

[7]We can always construct an element of D(**c**) in the following manner. For an arbitrary pair $(r,t) \in R\times T$, if the first branch of the definition of d_{rt} in Figure 1 is operative, then there exists g' such that $C_{g'rt}(d_{r1},\ldots,d_{rt-1},N+1) < c_{g'}$. Set $d_{g'rt} = N+1$ and $d_{grt} = 0$ for all $g \neq g'$. If the second branch determines d_{rt}, and $d_{rt} = 0$, then we set $d_{grt} = 0$ for all $g \in G$. Finally, suppose the second branch determines a positive d_{rt}. Since $C_{grt}(d_{r1},\ldots,d_{rt-1},v)$ is continuous and increasing in v, it follows that $C_{grt}(d_{r1},\ldots,d_{rt-1},d_{rt}) \geq c_g$ for all g, with = holding for at least one g, say g = g'. Set $d_{grt} = 0$ for all $g \neq g'$, and set $d_{g'rt} = d_{rt}$. $d_{g'rt}$ is permitted by the definition of D(**c**) to be positive since $C_{g'rt}(d_{r1},\ldots,d_{rt-1},d_{rt}) = c_{g'}$.

Proof. Condition (i) of equilibrium is obviously satisfied. Since $\sum_{g,r,t} d^*_{grt} = \sum_g N_g = N$, $d^*_{rt} \equiv \sum_g d^*_{grt} \leq N$ for each pair (r,t). Thus d^*_{rt} is determined by the second branch of its definition in Figure 1, and hence $C_{grt}(d^*_{r1},\ldots,d^*_{rt}) \geq c^*_g$ for all $g \in G$. Condition (iii) of equilibrium follows immediately. If $d^*_{grt} > 0$, then it follows from the definition of $D(\mathbf{c}^*)$ that $C_{grt}(d^*_{r1},\ldots,d^*_{rt}) \leq c^*_g$, and thus the equality must hold, verifying condition (ii) of equilibrium. ∎

The link to fixed point algorithms

We have shown that a zero of the excess commuter correspondence corresponds to an equilibrium. Fixed point algorithms can be used to constructively demonstrate the existence of a zero of a well-behaved correspondence and to approximate such a zero numerically. In particular, the proof of Theorem 3.2 below implies that a fixed point algorithm can be used to demonstrate the existence of a zero of the excess commuter correspondence, and hence by implication can be used to approximate a dynamic user equilibrium. The proof of Theorem 3.2 is an application of Lemma 3.1, which is a slight variation of a theorem which is proved constructively in RICHTER[16] using a fixed point algorithm. Lemma 3.1 essentially presents a set of sufficient conditions for a correspondence to have a zero.[8]

LEMMA 3.1. *Let* $\mu > 0$, $X \equiv \left\{ \mathbf{x} \in \Re^n_+ \,\middle|\, \sum_{i=1}^n x_i \leq \mu \right\}$ *and* $\xi : X \to \Re^n$ *be an upper hemicontinuous, nonempty, convex, compact-valued correspondence. Suppose that:*

(a.1) For each $\mathbf{x} \in X$ *with* $\sum_{i=1}^n x_i = \mu$ *and each* $\boldsymbol{\varepsilon} \in \xi(\mathbf{x})$, $\exists i \in \{1,\ldots,n\}$ *such that* $x_i > 0$ *and* $\varepsilon_i \leq 0$.

Then there exists $\mathbf{x}^* \in X$ *and* $\boldsymbol{\varepsilon}^* \in \xi(\mathbf{x}^*)$ *such that* $\boldsymbol{\varepsilon}^* \leq \mathbf{0}$, *with* $\varepsilon^*_j = 0$ *if* $x^*_j > 0$.

The following theorem follows directly from Lemma 3.1 and Theorem 3.1.

---

[8]Strictly speaking, the $\mathbf{x}^*$ determined by Lemma 3.1 may not be a zero of ξ unless $\mathbf{x}^* >> \mathbf{0}$.

THEOREM 3.2. *Suppose that for each $(g,r,t) \in G \times R \times T$:*

(A.1) C_{grt} *is continuous.*

(A.2) For fixed values of $(d_{r1},\ldots,d_{rt-1})$, C_{grt} *is increasing in* d_{rt}.

Then there exists a dynamic user equilibrium $(\mathbf{d}^*,\mathbf{c}^*) = ((d^*_{grt})_{g\in G, r\in R, t\in T}, (c^*_g)_{g\in G})$.

Proof. (A.1) implies there exists $\mu > 0$ such that for all $(g,r,t) \in G \times R \times T$,

$$C_{grt}(\mathbf{s}) < \mu/|G| \text{ for all } \mathbf{s} \in S_t \equiv \{\mathbf{s} \in \mathfrak{R}^t_+ \mid \mathbf{s} \leq (N+1,\ldots,N+1)\}. \tag{1}$$

Use this μ in Lemma 3.1, set $n = |G|$, $\mathbf{x} = \mathbf{c}$, $X = \left\{ \mathbf{c} \in \mathfrak{R}^{|G|}_+ \,\middle|\, \sum_{g\in G} c_g \leq \mu \right\}$,[9] and let ξ be the restriction of E to X.[10] We first show that all of the hypotheses of Lemma 3.1 will be satisfied if (A.1) and (A.2) of Theorem 3.2 hold. We then show that $(\mathbf{d}^*,\mathbf{c}^*)$, where $\mathbf{c}^* \equiv \mathbf{x}^*$ and $\left(N_g - \sum_{r,t} d^*_{grt}\right)_{g\in G} \equiv \boldsymbol{\varepsilon}^*$, constitute an equilibrium. It is straightforward to show that the restriction of E to X is upper hemicontinuous, nonempty, convex, and compact valued.[11] To establish (a.1) of Lemma 3.1, suppose $\mathbf{c} \in X$ with $\sum_{g\in G} c_g = \mu$, and let $\left(N_g - \sum_{r,t} d_{grt}\right)_{g\in G}$ be an arbitrary element of $E(\mathbf{c})$. Then for some $k \in G$, $c_k \geq \mu/|G|$. Since $C_{k11}(N+1) < \mu/|G| \leq c_k$ (the first inequality following from (1)), it follows from Figure 1 that $d_{11} = N+1$. Then from the definition of $D(\mathbf{c})$, $\sum_{g\in G} d_{g11} = N+1$. Hence there exists $i \in G$ such that $N_i - d_{i11} < 0$ and thus $N_i - \sum_{r,t} d_{irt} < 0$. Furthermore, $d_{i11} > 0$ implies

---

[9]Any equilibrium $\mathbf{c}^*$ must lie in this set. Suppose $\mathbf{c}^*$ and $\mathbf{d}^* = (d^*_{grt})_{g\in G, r\in R, t\in T}$ describe an equilibrium. Then $d^*_{rt} \equiv \sum_{g\in G} d^*_{grt} \leq N$ for all $(r,t) \in R\times T$. Consequently, $(d^*_{r1},\ldots,d^*_{rt}) \in S_t$ for all $(r,t) \in R\times T$, and hence $c^*_g \leq C_{grt}(d^*_{r1},\ldots,d^*_{rt}) < \mu/|G|$ for all $g \in G$. Thus $\sum_{g\in G} c^*_g < \mu$.

[10]Note that (A.1) and (A.2) have already been used in insuring that E is well defined.

[11]The convexity of the set $E(\mathbf{c})$ follows from the convexity of the set $D(\mathbf{c})$. In establishing that $D(\mathbf{c})$ is a convex set, it is critical that the total generated departure vector $(d_{rt})_{r\in R, t\in T}$ corresponding to $\mathbf{c}$ is unique. In establishing the upper hemicontinuity of E, it is critical that this total departure vector is a continuous function of $\mathbf{c}$.

from the definition of $D(\mathbf{c})$ that $C_{i11}(d_{11}) \leq c_i$. Since $C_{i11}(0) \geq 0$, $d_{11} = N+1$, and C_{i11} is an increasing function, we conclude that $c_i > 0$. In summary, given $\mathbf{c} \in X$ with $\sum_{g \in G} c_g = \mu$, and an arbitrary element $\left(N_g - \sum_{r,t} d_{grt}\right)_{g \in G}$ of $E(\mathbf{c})$, we have shown there exists an i such that the ith coordinate of this element is negative and $c_i > 0$. Thus (a.1) of Lemma 3.1 is satisfied. Applying the conclusion of Lemma 3.1 in the present context, there exists $\mathbf{c}^* \in X$ and $\left(N_g - \sum_{r,t} d^*_{grt}\right)_{g \in G} \in E(\mathbf{c}^*)$ such that $\left(N_g - \sum_{r,t} d^*_{grt}\right)_{g \in G} \leq \mathbf{0}$, with $\left(N_g - \sum_{r,t} d^*_{grt}\right) = 0$ if $c^*_g > 0$. Suppose for some g, $\left(N_g - \sum_{r,t} d^*_{grt}\right) < 0$. Then $c^*_g = 0$, and there exists a route i and a time period j such that $d^*_{gij} > 0$ and hence $d^*_{ij} > 0$. Since $C_{gij}(d^*_{i1}, \ldots, d^*_{ij-1}, 0) \geq 0$ and C_{gij} is increasing in its last coordinate, it follows that $C_{gij}(d^*_{i1}, \ldots, d^*_{ij}) > 0 = c^*_g$. Then it follows from the definition of $D(\mathbf{c})$ that $d^*_{gij} = 0$, which is a contradiction. Thus $(N_g - \sum_{r,t} d^*_{grt})$ must be 0 for all $g \in G$. Thus $\left(N_g - \sum_{r,t} d^*_{grt}\right)_{g \in G}$ is a zero of $E(\mathbf{c}^*)$, and by Theorem 3.1, $(\mathbf{d}^*, \mathbf{c}^*)$ is a dynamic user equilibrium. ∎

Since a fixed point algorithm can be used to approximate the $\mathbf{x}^*$ and ε^* of Lemma 3.1, it can be used to approximate the user equilibrium summarized by $(\mathbf{d}^*, \mathbf{c}^*)$.

4. A NUMERICAL EXAMPLE

In this section we show how our algorithm can be used to approximate the equilibrium of a continuous time model of the morning rush hour involving heterogeneous commuters, schedule delay costs, bottleneck congestion, and a Vickrey type queueing model.[12] There are 3 parallel routes from a suburb to a CBD, with 3 commuter groups. Each of the N_g commuters in group g has the user cost function $U_g = \alpha_g(\text{travel time}) + \beta_g(\text{time early}) + \gamma_g(\text{time late})$. The characteristics of the groups are summarized in Table 2, where α_g, β_g, and γ_g are measured in dollars per hour.

[12] ARNOTT et al.[2] employ such a model with one route.

TABLE 2
Group Characteristics

Group	N_g	α_g	β_g	γ_g	Desired arrival time
1	7500	4.8	2.4	3.6	8:00
2	7500	15.0	1.0	2.0	8:30
3	7500	15.0	2.0	4.0	9:00

Each route contains a single bottleneck with a fixed capacity. We assume the bottleneck is entered as soon as a commuter leaves home, and is followed by a stretch of uncongested roadway before the CBD is reached. Travel time is the sum of waiting time at the bottleneck and time to traverse the uncongested portion of the route. The characteristics of the routes are summarized in Table 3.

TABLE 3
Route Characteristics

Route	Bottleneck capacity (commuters per minute)	Travel time on uncongested portion (minutes)
1	58.33	6.6
2	66.67	6
3	75	5.4

An equilibrium in such a model can be summarized by a function $r_g(t)$ for each group g, where t is a continuous variable representing departure time, and $r_g(t)$ is the departure rate of group g at time t. Such a collection of functions represents an equilibrium if no commuter can lower his costs by unilaterally altering his departure time.

In translating this model into our discrete formulation we chopped up time into 1 minute intervals and assumed that departures during any 1 minute interval were distributed uniformly over the interval.[13] Then we defined our cost functions C_{grt} as the average of the actual costs incurred by all commuters from group g departing on route r during interval t.[14]

---

[13]Alternative assumptions were tried. In one variant, all departures during an interval were assumed to be made at the beginning of the interval. Very little difference in the results occurred. Of course, as the length of the intervals approaches zero, these variants should converge to the same result.

[14]For example, suppose the interval t corresponds to the period 7:30–7:31, and that given the vector of total departures $(d_{r1},\ldots,d_{rt})$ on route r up through interval t, a commuter from group g leaving at 7:31 will arrive early to work. Let Q_{rt} denote the length of the queue at 7:30 on route r. Q_{rt} will be a function of $(d_{r1},\ldots,d_{rt-1})$ and can be determined iteratively from the equation $Q_{rt} = Q_{rt-1} - cap_r + d_{rt-1}$, where cap_r is the number of commuters the bottleneck on route r can process per minute. Let τ_r denote the time to travel the uncongested part of route r, t_g^* denote the desired arrival time of group g commuters, and 7:30:30 denote the midpoint of interval t (i.e., when the clock is half past the hour plus 30 seconds). Then we define

$$C_{grt}(d_{11},\ldots,d_{1t}) \equiv \frac{\alpha_g}{60}\left\{\max\left(\frac{Q_{rt}+0.5(d_{rt}-cap_r)}{cap_r},0\right)+\tau_r\right\}+$$
$$\frac{\beta_g}{60}\left\{t_g^* - 7{:}30{:}30 - \max\left(\frac{Q_{rt}+0.5(d_{rt}-cap_r)}{cap_r},0\right)-\tau_r\right\}.$$

α_g and β_g are divided by 60 to translate their units into dollars per minute. The term in braces multiplying α_g approximates average travel time for a commuter departing on route r during interval t, under the assumption that departures of these commuters are distributed uniformly in the time interval. The numerator of the first term under the max operator represents the total number of commuters in line at the bottleneck on route r at 7:30 plus one half the difference between the total departures on route r between 7:30 and 7:31 and the total number of commuters the bottleneck can process during this minute. If the numerator is positive, it represents the average queue size during the interval. Dividing by cap_r then yields the average waiting time in the queue for commuters who depart during the interval. If the numerator is nonpositive, then it can be shown that $Q_{rt} = 0$, and hence no queue will appear during the interval, implying average waiting time is 0. (One feature of this particular model is that the queue is

Tables 4 and 5 and Figure 2 are examples of the type of output that can be obtained from our computations.[15]

---

nondecreasing over time as long as early arrival is feasible for at least one group. Without this feature a more complex representation of C_{grt} would be necessary.) The term in braces multiplying β_g measures the average number of minutes early (assumed positive) that a commuter from group g arrives who departs during this interval. The desired arrival time is subtracted from the average arrival time, where average arrival time is determined by adding to the average time of departure (7:30:30) the average travel time. Analogous expressions can be derived for the case where everyone from group g who departs during interval t arrives late, and the case where there is a mix of early and late arrivals. C_{grt} may only be nondecreasing (rather than increasing as required by assumption (A.2)) in d_{rt} if there is no queue on route r in interval t, but this has caused no particular difficulties in the computations. We have chosen $\alpha_g > \beta_g$ since otherwise C_{grt} may be decreasing in d_{rt}, which can cause existence problems, as has been noted by HENDRICKSON AND KOCUR[11].

[15]An approximate equilibrium was found using a Fortran implementation of a fixed point algorithm due to TODD[20]. The equilibrium was found in about 17 seconds of cpu time on a Vax 8700, and involved the calculation of 121 elements from the image sets of the excess commuter correspondence. Increasing the number of routes would have relatively little effect on computational time. The equilibrium is approximate in the sense that a commuter may be able to find an alternative route that is slightly cheaper. In the equilibrium presented, no commuter could reduce his cost by more than 0.006% by switching routes, and hence, practically speaking, would have no incentive to do so. By appropriate rounding and reallocation of a small number of commuters to alternative routes, one can easily obtain an approximate equilibrium involving an integral number of departures during every time interval.

TABLE 4
Equilibrium Departure Patterns

	Departure time	Group departing	State of arrival	Departure rate (commuters/min)
ROUTE 1	7:29–7:41	1	early	116.67
	7:41–8:03	1	late	33.33
	8:03–8:19	2	early	62.50
	8:19–8:32	2	late	51.47
	8:32–8:49	3	early	67.30
	8:49–9:09	3	late	46.05
ROUTE 2	7:21–7:29	2	early	71.43
	7:29–7:41	1	early	133.33
	7:41–8:03	1	late	38.10
	8:03–8:19	2	early	71.43
	8:19–8:32	2	late	58.82
	8:32–8:49	3	early	76.92
	8:49–9:12	3	late	52.63
ROUTE 3	7:13–7:29	2	early	80.36
	7:29–7:41	1	early	150.00
	7:41–8:03	1	late	42.86
	8:03–8:19	2	early	80.36
	8:19–8:32	2	late	66.18
	8:32–8:49	3	early	86.54
	8:49–9:15	3	late	59.21

TABLE 5
Departure Times for On Time Arrival

Group	Desired arrival time	Departure time to arrive on time
1	8:00	7:41
2	8:30	8:19
3	9:00	8:48

In Figure 2 the equilibrium size of the queue at the bottleneck on route 3 during the rush hour is plotted. From 7:13 to 7:29 commuters from group 2 arrive at a rate (80.36 commuters per minute) exceeding the capacity (75 commuters per minute) of the bottleneck, thus causing the queue to rise. (Note that the arrival rate of commuters from group 2 at the bottleneck from 7:13 to 7:29 is the same as the departure rate in Table 4 for group 2 on route 3 from 7:13 to 7:29, reflecting our assumption that commuters reach the bottleneck as soon as they leave home.) The queue continues to rise as commuters from group 1 arrive from 7:29 to 7:41 at a rate exceeding the capacity. From 7:41 to 8:03 the queue lessens, since the arrival rate of commuters from group 1 is less than the capacity of the bottleneck. By 9:15, after several local peaks, the queue disappears.

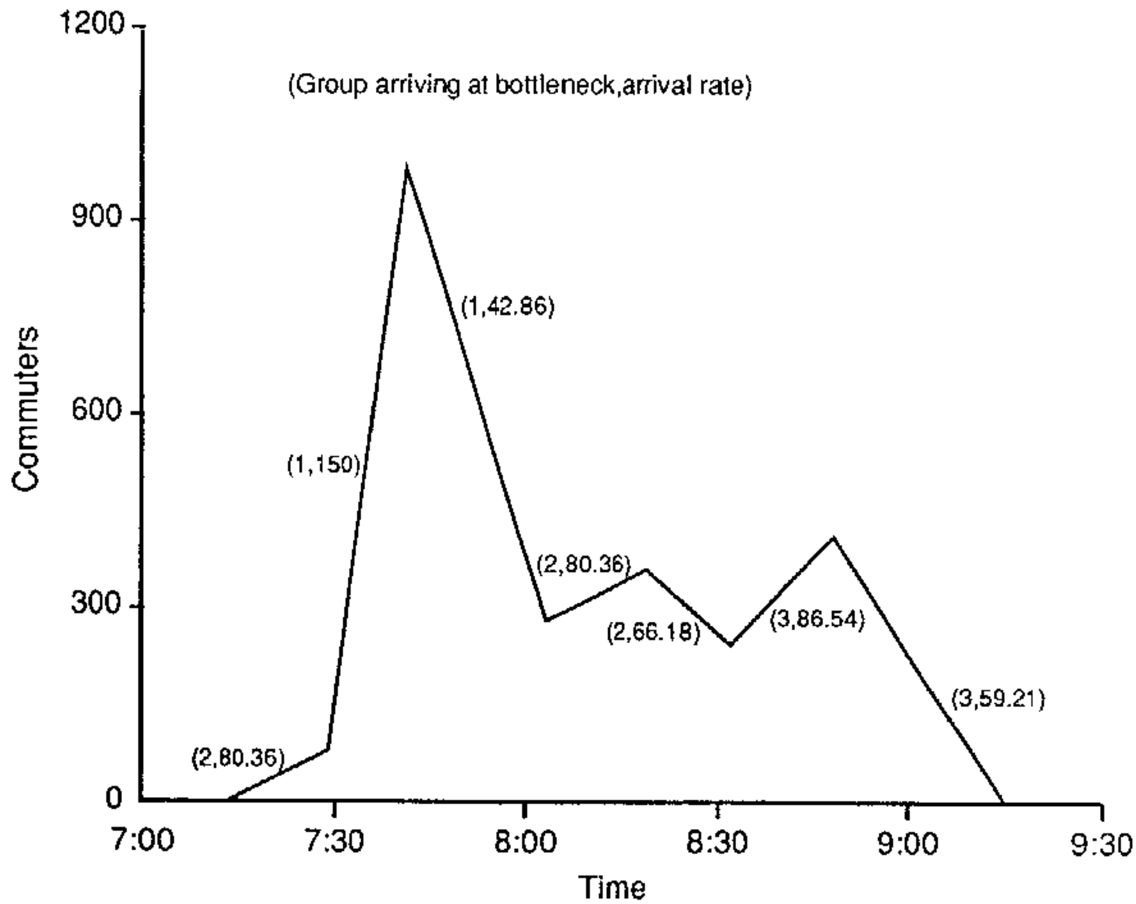

Fig. 2. The queue on route 3

5. AN ALTERNATIVE VIEW OF THE COMPUTATIONAL PROCEDURE

In this section we view our computational procedure as a decomposition procedure for a nonlinear complementarity problem. This view makes transparent why our technique alleviates to a considerable degree, in the present context, the dimensionality problems inherent in the use of fixed point algorithms.

AASHTIANI AND MAGNANTI[1] have formulated the static user equilibrium problem as a nonlinear complementarity problem. In the spirit of their approach, we can formulate our dynamic user equilibrium problem as a nonlinear complementarity problem.

THEOREM 5.1. *$(\mathbf{d}^*,\mathbf{c}^*)$, where $\mathbf{d}^* = (d^*_{grt})_{g\in G, r\in R, t\in T}$, is a dynamic user equilibrium if and only if $(\mathbf{d}^*,\mathbf{c}^*)$ is a solution of:*

Find $(d_{grt})_{g\in G, r\in R, t\in T} \geq \mathbf{0}$ and $(c_g)_{g\in G} \geq \mathbf{0}$ such that

$$C_{grt}\left(\sum_{k\in G} d_{kr1},\ldots,\sum_{k\in G} d_{krt}\right) - c_g \geq 0 \quad \forall\,(g,r,t)\in G\times R\times T$$

$$\left(C_{grt}\left(\sum_{k\in G} d_{kr1},\ldots,\sum_{k\in G} d_{krt}\right) - c_g\right) d_{grt} = 0 \quad \forall\,(g,r,t)\in G\times R\times T \qquad (2)$$

$$\sum_{r\in R, t\in T} d_{grt} - N_g \geq 0 \quad \forall\, g\in G$$

$$\left(\sum_{r\in R, t\in T} d_{grt} - N_g\right) c_g = 0 \quad \forall\, g\in G.$$ [16,17]

$(\mathbf{d}^*,\mathbf{c}^*)$ is a solution of (2) if and only if it is a zero of the function $\Phi: \Re_+^{|G||R||T|+|G|} \to \Re^{|G||R||T|+|G|}$, where

---

[16]The general nonlinear complementarity problem is: Let $\zeta: \Re_+^n \to \Re^n$ be a continuous function. Find $\mathbf{z} \geq \mathbf{0}$ with $\zeta(\mathbf{z}) \geq \mathbf{0}$ and $\mathbf{z}\cdot\zeta(\mathbf{z}) = 0$. Problem (2) corresponds to choosing $n = |G||R||T|+|G|$, $\mathbf{z} = ((d_{grt})_{g\in G, r\in R, t\in T}, (c_g)_{g\in G})$, and $\zeta(\mathbf{z}) = \left(\left(C_{grt}(\cdot) - c_g\right)_{g\in G, r\in R, t\in T}, \left(\sum_{r\in R, t\in T} d_{grt} - N_g\right)_{g\in G}\right)$.

[17]Proof: The only part which is not obvious is that if $(\mathbf{d}^*,\mathbf{c}^*)$ is a solution of (2), then $\sum_{r\in R, t\in T} d^*_{grt} - N_g = 0$ $\forall\, g\in G$. Suppose for some g' that $\sum_{r\in R, t\in T} d^*_{g'rt} - N_{g'} > 0$. Then $c_{g'} = 0$ and for some $(i,j)\in R\times T$, $d^*_{g'ij} > 0$. But $d^*_{g'ij} > 0$ implies $C_{g'ij}\left(\sum_{k\in G} d^*_{ki1},\ldots,\sum_{k\in G} d^*_{kij}\right) - c_{g'} = 0$. Since $c_{g'} = 0$, we conclude that $C_{g'ij}\left(\sum_{k\in G} d^*_{ki1},\ldots,\sum_{k\in G} d^*_{kij}\right) = 0$. Since $\sum_{k\in G} d^*_{kij} > 0$, this equality is inconsistent with (A.2).

$$\Phi(\mathbf{d},\mathbf{c})=\left(\left(\min\left[d_{grt},C_{grt}(\cdot)-c_g\right]\right)_{g\in G,r\in R,t\in T},\left(\min\left[c_g,\sum_{r\in R,t\in T}d_{grt}-N_g\right]\right)_{g\in G}\right).$$

One way of solving (2) is to use a fixed point algorithm to find a zero of Φ (see, eg., TODD[19], p. 22). However, computational time of a fixed point algorithm generally varies with the number of arguments of Φ raised to a power of 3 or more. Since $|T|$ and possibly $|R|$ are likely to be large in the present context, use of a fixed point algorithm to compute a dynamic user equilibrium by finding a zero of Φ does not appear promising.[18] However, we now show that our earlier algorithm amounts to decomposing (2) into a number of interdependent smaller nonlinear complementarity problems (which can be solved easily because of the special structure of the cost functions), with a fixed point algorithm applied to a "master" coordinating problem involving only $|G|$ variables. In this fashion we are able to reduce the dimensionality of the problem to which the fixed point algorithm is applied from an unmanageable $|G||R||T|+|G|$ to a much more manageable $|G|$.

Problem (2) generates $|R||T|$ nonlinear complementarity subproblems. Subproblem (r,t), for $(r,t)\in R\times T$, is defined for predetermined nonnegative vectors $(c_g)_{g\in G}$ and $(d_{r1},\ldots,d_{rt-1})$:

Find $(d_{grt})_{g\in G}\geq\mathbf{0}$ such that

$$C_{grt}\left(d_{r1},\ldots,d_{rt-1},\sum_{k\in G}d_{krt}\right)-c_g\geq 0\quad\forall g\in G \tag{3}$$

$$\left(C_{grt}\left(d_{r1},\ldots,d_{rt-1},\sum_{k\in G}d_{krt}\right)-c_g\right)d_{grt}=0\quad\forall g\in G.$$

Figure 3 illustrates the creation of the *subproblem solution correspondence* $\delta:\Re_+^{|G|}\to\Re_+^{|G||R||T|}$. Given the nonnegative vector $\mathbf{c}=(c_1,\ldots,c_{|G|})$, we define $\delta_{r1}(\mathbf{c})$, for

---

[18]An optimistic upper bound on the number of arguments that a fixed point algorithm can be expected to handle lies between 50 and 100.

$$\mathbf{c} = (c_1, \ldots, c_{|G|})$$

$$\downarrow$$

$$\delta_{r1}(\mathbf{c}) \equiv \begin{cases} \left\{ (\hat{d}_{gr1})_{g\in G} \geq \mathbf{0} \;\middle|\; \sum_{g\in G} \hat{d}_{gr1} = N+1, \text{ with } \hat{d}_{gr1} > 0 \text{ only if } C_{gr1}(N+1) \leq c_g \right\} \\ \qquad \text{if subproblem } (r,1) \text{ does not have a solution } (\hat{d}_{gr1})_{g\in G} \text{ with } \sum_{g\in G} \hat{d}_{gr1} \leq N+1 \\ \left\{ (\hat{d}_{gr1})_{g\in G} \;\middle|\; \sum_{g\in G} \hat{d}_{gr1} \leq N+1 \text{ and } (\hat{d}_{gr1})_{g\in G} \text{ is a solution of subproblem } (r,1) \right\} \quad \text{otherwise} \end{cases}$$

$$\downarrow$$

$$(d_{r1})_{r\in R}$$

$$\downarrow$$

$$\delta_{r2}(\mathbf{c}) \equiv \begin{cases} \left\{ (\hat{d}_{gr2})_{g\in G} \geq \mathbf{0} \;\middle|\; \sum_{g\in G} \hat{d}_{gr2} = N+1, \text{ with } \hat{d}_{gr2} > 0 \text{ only if } C_{gr2}(d_{r1}, N+1) \leq c_g \right\} \\ \qquad \text{if subproblem } (r,2) \text{ does not have a solution } (\hat{d}_{gr2})_{g\in G} \text{ with } \sum_{g\in G} \hat{d}_{gr2} \leq N+1 \\ \left\{ (\hat{d}_{gr2})_{g\in G} \;\middle|\; \sum_{g\in G} \hat{d}_{gr2} \leq N+1 \text{ and } (\hat{d}_{gr2})_{g\in G} \text{ is a solution of subproblem } (r,2) \right\} \quad \text{otherwise} \end{cases}$$

$$\downarrow$$

$$(d_{rj})_{r\in R, j\in\{1,2\}}$$

$$\vdots$$

$$\downarrow$$

$$(d_{rj})_{r\in R, j\in\{1,\ldots t-1\}}$$

$$\downarrow$$

$$\delta_{rt}(\mathbf{c}) \equiv \begin{cases} \left\{ (\hat{d}_{grt})_{g\in G} \geq \mathbf{0} \;\middle|\; \sum_{g\in G} \hat{d}_{grt} = N+1, \text{ with } \hat{d}_{grt} > 0 \text{ only if } C_{grt}((d_{rj})_{j\in\{1,\ldots,t-1\}}, N+1) \leq c_g \right\} \\ \qquad \text{if subproblem } (r,t) \text{ does not have a solution } (\hat{d}_{grt})_{g\in G} \text{ with } \sum_{g\in G} \hat{d}_{grt} \leq N+1 \\ \left\{ (\hat{d}_{grt})_{g\in G} \;\middle|\; \sum_{g\in G} \hat{d}_{grt} \leq N+1 \text{ and } (\hat{d}_{grt})_{g\in G} \text{ is a solution of subproblem } (r,t) \right\} \quad \text{otherwise} \end{cases}$$

$$\downarrow$$

$$\vdots$$

$$\delta(\mathbf{c}) \equiv \left\{ (\hat{d}_{grt})_{g\in G, r\in R, t\in T} \;\middle|\; (\hat{d}_{grt})_{g\in G} \in \delta_{rt}(\mathbf{c}) \right\}$$

Fig. 3. Creation of the subproblem solution correspondence

each $r \in R$, as indicated.[19] $\delta_{r1}(\mathbf{c})$ is essentially the set of solutions of subproblem (r,1), appropriately bounded. If $(d'_{kr1})_{k\in G}$ and $(d''_{kr1})_{k\in G}$ belong to $\delta_{r1}(\mathbf{c})$, it follows from monotonicity of the cost functions (assumption (A.2)) that $\sum_{k\in G} d'_{kr1} = \sum_{k\in G} d''_{kr1}$.[20] Thus $\delta_{r1}(\mathbf{c})$ determines the unique nonnegative number d_{r1}, where $d_{r1} = \sum_{k\in G} \hat{d}_{kr1}$ for any $(\hat{d}_{kr1})_{k\in G} \in \delta_{r1}(\mathbf{c})$, and collectively $\delta_{r1}(\mathbf{c})$ (for $r = 1,\ldots,|R|$) determine the unique nonnegative vector $(d_{r1})_{r\in R}$.

Armed with the vector $(d_{r1})_{r\in R}$, we then consider subproblem (r,2) for all $r \in R$ by setting $t = 2$ in (3). $\delta_{r2}(\mathbf{c})$ is defined analogously to $\delta_{r1}(\mathbf{c})$, and the collection of sets $\{\delta_{r2}(\mathbf{c})\}_{r\in R}$ generates the unique vector $(d_{r2})_{r\in R} \geq \mathbf{0}$, where $d_{r2} = \sum_{k\in G} \hat{d}_{kr2}$ for any $(\hat{d}_{kr2})_{k\in G} \in \delta_{r2}(\mathbf{c})$. Continuing in this fashion we end up with the sets $\delta_{rt}(\mathbf{c})$ for all

---

[19]The set determined by the first line of $\delta_{r1}(\mathbf{c})$ is nonempty. To verify this it is sufficient to show that if subproblem (r,1) does not have a solution $(\hat{d}_{gr1})_{g\in G}$ with $\sum_{g\in G} \hat{d}_{gr1} \leq N+1$, then there exists $g' \in G$ with $C_{g'r1}(N+1) < c_{g'}$. Suppose not. Then $C_{gr1}(N+1) \geq c_g \; \forall g \in G$. Choose the smallest $\lambda \in [0,1]$ such that $C_{gr1}(\lambda(N+1)) \geq c_g \; \forall g \in G$. (The continuity of C_{gr1} insures that λ is well defined.) If $\lambda = 0$, then $(\hat{d}_{gr1})_{g\in G} = \mathbf{0}$ contradicts our hypothesis that subproblem (r,1) does not have a solution $(\hat{d}_{gr1})_{g\in G}$ with $\sum_{g\in G} \hat{d}_{gr1} \leq N+1$. If $\lambda > 0$, $\exists k \in G$ with $C_{kr1}(\lambda(N+1)) = c_k$. (This assertion follows from the continuity and monotonicity of the cost functions.) Let $\hat{d}_{gr1} = 0 \; \forall g \neq k$ and let $\hat{d}_{kr1} = \lambda(N+1)$. Then $(\hat{d}_{gr1})_{g\in G}$ is a solution of subproblem (r,1) — contradiction.

[20]This assertion trivially follows if the first line of δ_{r1} is operative. Thus suppose $(d'_{kr1})_{k\in G}$ and $(d''_{kr1})_{k\in G}$, with $\sum_{k\in G} d'_{kr1} > \sum_{k\in G} d''_{kr1}$, are determined by the second line of $\delta_{r1}(\mathbf{c})$. Then there exists $g \in G$ with $d'_{gr1} > 0$. Since $(d'_{kr1})_{k\in G}$ is a solution of subproblem (r,1), $C_{gr1}(\sum_{k\in G} d'_{kr1}) - c_g = 0$. But then monotonicity of C_{gr1} implies $C_{gr1}(\sum_{k\in G} d''_{kr1}) - c_g < 0$, which contradicts the supposition that $(d''_{kr1})_{k\in G}$ is a solution of subproblem (r,1).

$(r,t) \in R{\times}T$. We then define $\delta(\mathbf{c}) \equiv \{(\hat{d}_{grt})_{g\in G, r\in R, t\in T} \mid (\hat{d}_{grt})_{g\in G} \in \delta_{rt}(\mathbf{c})\}$. It is easily verified that $\delta(\mathbf{c})$ is identical to the set $D(\mathbf{c})$ of Figure 1.[21]

Given the correspondence δ, we then define the *master problem:*

Find $\mathbf{c} = (c_g)_{g\in G} \geq \mathbf{0}$ such that for some $(d_{grt})_{g\in G, r\in R, t\in T} \in \delta(\mathbf{c})$:

$$\sum_{r\in R, t\in T} d_{grt} - N_g \geq 0 \quad \forall g \in G \tag{4}$$

$$\left(\sum_{r\in R, t\in T} d_{grt} - N_g\right) c_g = 0 \quad \forall g \in G.$$[22]

If $\mathbf{c}^*$ is a solution of the master problem, and $\mathbf{d}^*$ is the associated element of $\delta(\mathbf{c}^*)$, then $(\mathbf{d}^*,\mathbf{c}^*)$ is a solution of problem (2), and hence (by Theorem 5.1), a dynamic user

---

[21]Suppose $(d_{grt})_{g\in G, r\in R, t\in T} \in D(\mathbf{c})$. If $d_{rt} = \sum_{g\in G} d_{grt}$ is determined by the first line of its definition in Figure 1, then $\sum_{g\in G} d_{grt} = N+1$, and the existence of a g with $C_{grt}(d_{r1},\ldots,d_{rt-1},N+1) < c_g$ implies (using monotonicity of the cost functions) that subproblem (r,t) does not have a solution whose coordinate sum is less than or equal to N+1. Then the first line of the definition of $\delta_{rt}(\mathbf{c})$ implies $(d_{grt})_{g\in G} \in \delta_{rt}(\mathbf{c})$. If d_{rt} is determined by the second line of its definition in Figure 1, then $\sum_{g\in G} d_{grt} \leq N+1$. Furthermore, the requirement that $d_{grt} > 0$ only if $C_{grt}(d_{r1},\ldots,d_{rt}) \leq c_g$ (from the definition of $D(\mathbf{c})$), coupled with the restrictions from the second line of the definition of d_{rt} in Figure 1, implies $(d_{grt})_{g\in G}$ is a solution of subproblem (r,t), and hence $(d_{grt})_{g\in G} \in \delta_{rt}(\mathbf{c})$. Thus $(d_{grt})_{g\in G, r\in R, t\in T} \in \delta(\mathbf{c})$. The proof of the converse is left to the reader.

[22]Problem (4) can be viewed as a correspondence version of the standard nonlinear complementarity problem. The standard version is: Let $\zeta: \Re^n_+ \to \Re^n$ be a continuous function. Find $\mathbf{z} \geq \mathbf{0}$ with $\zeta(\mathbf{z}) \geq \mathbf{0}$ and $\mathbf{z}\cdot\zeta(\mathbf{z}) = 0$. The correspondence version is: Let $\zeta: \Re^n_+ \to \Re^n$ be an upper hemicontinuous correspondence. Find $\mathbf{z} \geq \mathbf{0}$ such that for some $\mathbf{e} \in \zeta(\mathbf{z})$ with $\mathbf{e} \geq \mathbf{0}$, $\mathbf{ze} = 0$. Problem (4) is a special case of the correspondence version in which $n = |G|$, $\mathbf{z} = \mathbf{c}$, and $\zeta(\mathbf{z}) =$
$\left\{\left(\sum_{r\in R, t\in T} d_{grt} - N_g\right)_{g\in G} \,\middle|\, (d_{grt})_{g\in G, r\in R, t\in T} \in \delta(\mathbf{c})\right\}$.

equilibrium. To see this note that $\sum_{r\in R,t\in T} d^*_{grt} - N_g = 0 \ \ \forall g \in G$,[23] and hence $\sum_{g\in G} d^*_{grt} < N+1$ for all $(r,t)\in R{\times}T$. Since $(d^*_{grt})_{g\in G} \in \delta_{rt}(\mathbf{c}^*)$, it must be the case that the last line of the definition of $\delta_{rt}(\mathbf{c}^*)$ is operative, and hence $(d^*_{grt})_{g\in G}$ is a solution of subproblem (r,t). It follows immediately that $(\mathbf{d}^*,\mathbf{c}^*)$ is a solution of problem (2).

As noted above, if $\mathbf{c}^*$ is a solution of the master problem, and $\mathbf{d}^*$ is the associated element of $\delta(\mathbf{c}^*)$, then $\sum_{r\in R,t\in T} d^*_{grt} - N_g = 0 \ \ \forall g \in G$. Thus solving the master problem amounts to finding a zero of the correspondence $\mathrm{M}: \Re_+^{|G|} \to \Re^{|G|}$, where $\mathrm{M}(\mathbf{c}) = \left\{ \left(N_g - \sum_{r\in R,t\in T} d_{grt} \right)_{g\in G} \,\middle|\, (d_{grt})_{g\in G,r\in R,t\in T} \in \delta(\mathbf{c}) \right\}$. Since $\delta(\mathbf{c})$ is identical to $D(\mathbf{c})$, M is identical to the excess commuter correspondence. Thus the decomposition procedure of the present section reduces to our earlier excess commuter approach.

6. GENERALIZATIONS

Generalization of our base model to include treatment of multiple origins and/or destinations, with each route carrying commuters only for one origin-destination pair, is straightforward. For example, suppose different groups are exogenously located at different origins, with all groups facing a choice of multiple destinations. Then each route will have an origin-destination pair associated with it, and the cost functions for the various routes can reflect the relative attractiveness of the various destinations.[24]

---

[23]Suppose $\sum_{r\in R,t\in T} d^*_{g'rt} > N_{g'}$ for some $g' \in G$. Then $c^*_{g'} = 0$, and there exists a route i and a time period j such that $d^*_{g'ij} > 0$. Since $(d^*_{gij})_{g\in G} \in \delta_{ij}(\mathbf{c}^*)$, it follows that $C_{g'ij}(\sum_{k\in G} d^*_{ki1},\ldots,\sum_{k\in G} d^*_{kij}) - c_{g'} = 0$, which implies that $C_{g'ij}(\sum_{k\in G} d^*_{ki1},\ldots,\sum_{k\in G} d^*_{kij}) = 0$. This equality contradicts the monotonicity of $C_{g'ij}$.

[24]For example, in addition to travel time and schedule delay costs, negative terms reflecting attractiveness of the origin or destination can be added to the cost function. (The cost functions can be appropriately translated to preserve their nonnegativity.) If group g does not reside at the location from which route r originates, or the destination of route r does not correspond to a possible workplace for group g, then C_{grt} can be chosen to be a suitably large number for all t.

Note that this generalization does not cause an increase in the number of arguments of the excess commuter correspondence (or equivalently, the number of variables in the master problem). The choice of origin can be made endogenous, with the cost function reflecting relative attractiveness of the origins. Scarcity of land at the various origins can be handled by expanding the number of variables of the excess commuter correspondence to include a land rental at each origin, and expanding the number of coordinates of an image point to include an excess demand for land at each origin.[25] An artificial route representing the option of staying at home rather than making a trip can also be introduced.[26] Adding policy instruments such as congestion tolls is also straightforward.

Relaxing the structure of the cost functions would permit the treatment of more complicated networks, but is not nearly as straightforward. Our computational procedure has crucially exploited the following properties of C_{grt}: (1) its lack of dependence on total departures in intervals subsequent to t; (2) its lack of dependence on total departures on routes other than r in interval t. From the excess commuter

---

[25]Suppose for simplicity that there is a fixed supply of land at each origin, and each commuter's demand for land at a particular origin is totally inelastic. The cost functions would have as additional arguments land rentals at the various origins. Conditional on a vector of minimized costs **c** and land rentals, a departure pattern $(d_{grt})_{g\in G, r\in R, t\in T}$ would be generated essentially as in the base model. Since each route would have an origin associated with it, this departure pattern would generate a demand for land at each origin, and hence an excess demand for land at each origin. We would then define an augmented excess commuter correspondence, whose image points would contain additional coordinates corresponding to the excess demand for land at the various origins. The fixed point algorithm would then be used to find a zero of this augmented excess commuter correspondence.

[26]For example, suppose there is a fixed cost $\bar{c}$ for an individual of type g staying at home (i.e., using route $\bar{r}$). In order to preserve assumption (A.2), we can define $C_{g\bar{r}t}(d_{\bar{r}1},\ldots,d_{\bar{r}t}) = \bar{c} + \varepsilon d_{\bar{r}t}$ for all $t \in T$, where ε is an arbitrarily small positive number.

viewpoint, property (1) allowed us to determine the total departures in each interval sequentially, first determining $(d_{r1})_{r\in R}$, then $(d_{r2})_{r\in R}$ etc. Within an interval, property (2) allowed us to determine the total departures on each route independently of one another, thus permitting the determination of d_{rt} via a one-dimensional search. From the decomposition perspective of Section 5, property (1) allowed us to solve the subproblems sequentially, while (2) allowed us to index the subproblems by r as well as t, making the solution of the subproblems particularly simple.

One can trivially extend our computational procedure to permit C_{grt} to depend on $((d_{ij})_{i\in R,\, j\in\{1,\ldots,t-1\}}, d_{rt})$, i.e., on total departures on all routes (not just route r) in periods 1,...,t−1, as well as total departures in period t. While this generalization would not significantly enhance the kinds of networks that could be treated, it would allow for intermodal congestion.[27] Permitting C_{grt} to also depend on $((d_{ij})_{i\in R\setminus\{r\}, j\in\{t,\ldots,|T|\}})$, i.e., on total departures on routes other than r in period t and beyond would require a more significant alteration. To see how this might be accomplished, we shall view the essential elements of the decomposition procedure of Section 5, for the case of 2 routes and 2 time periods and an aribitrary number of groups,[28] using a metaphor drawn from the literature on decentralized economic planning (see Figure 4). Suppose there is a central planner and 4 managers. The central planner sends cost signals $(c_g)_{g\in G}$ to each of the 4 managers. Conditional on these signals, manager (r,1) (for r = 1,2) decides, by solving subproblem (r,1), how

[27]For example, suppose there was only one road between a single residential location and the CBD, but 2 modes (bus and car). Then we can define 2 routes, with route 1 corresponding to traveling by car on the road, and route 2 corresponding to traveling by bus on the same road. Then including $(d_{2j})_{j\in\{1,\ldots,t-1\}}$ among the list of arguments of C_{g1t} allows us to treat the case where the cost incurred by a commuter from group g departing in period t by car depends upon the number of buses that have left in previous periods on the same road.

[28]The diagram trivially generalizes to arbitrary numbers of routes and time periods.

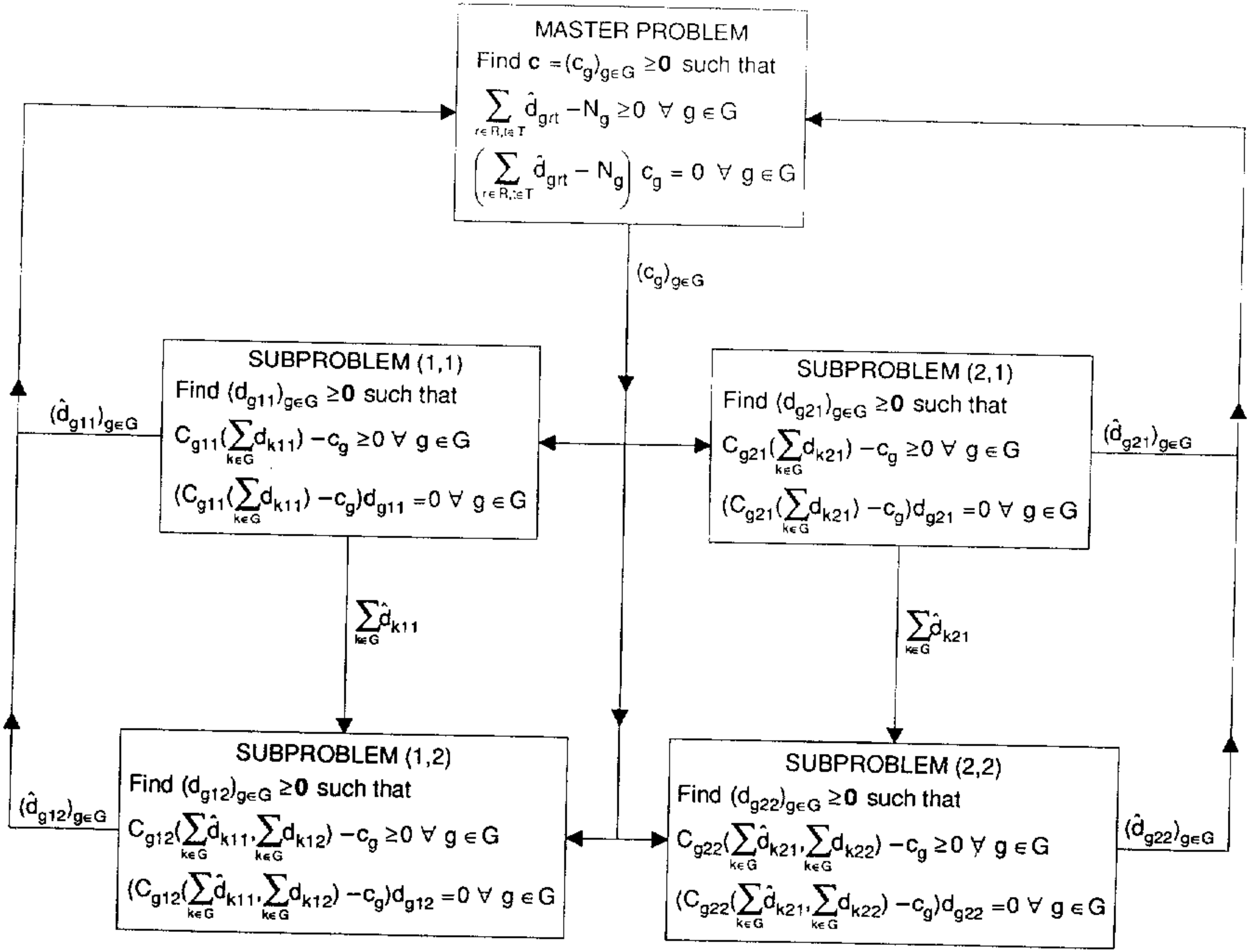

Fig. 4. An economic planning interpretation of the algorithm for the case $|R| = |T| = 2$

many commuters from each group must depart on route r during interval 1. This decision, $(\hat{d}_{gr1})_{g\in G}$, is reported to the central planner, while the total number of commuters that have been assigned to route r, $\sum_{k\in G}\hat{d}_{kr1}$, is reported to manager (r,2). Manager (r,2) (for r = 1,2) then uses this information together with the cost signals sent from the central planner to solve subproblem (r,2), reporting an optimal solution $(\hat{d}_{gr2})_{g\in G}$ to the central planner. The central planner then plugs the reports of the 4 managers into the master problem and checks whether the cost signals that had been sent solve the master problem. If not, the central planner uses a fixed point algorithm to adjust the cost signals and the whole process is repeated. Basically what we have shown in Section 5 is that there exists a vector of cost signals which the central planner can send, and a vector of corresponding assignments which the managers can report, such that the master problem will be solved. Furthermore, a fixed point algorithm can be used to find such a set of signals.

The lack of dependence of C_{grt} on departures subsequent to interval t permits the subproblems to be solved sequentially, with managers (1,1) and (2,1) solving their problems, and then managers (1,2) and (2,2) solving theirs.[29] The lack of dependence of C_{grt} on departures on other routes in the same interval then permits managers (1,1) and (2,1) to make their assignments independently of one another, and managers (1,2) and (2,2) (once they have received their reports from managers (1,1) and (2,1)) to solve their problems independently.

A tiny step toward treating more complicated networks can be made by allowing C_{grt} to depend on $(d_{ij})_{i\in R, j\in\{1,\ldots,t\}}$, i.e., C_{grt} depends on total departures on all routes in the current period as well as all previous periods. A network involving nonparallel roads, such as the one pictured in Figure 5, can be handled using this formulation.

---

[29]The lack of dependence of C_{grt} on past departures on routes other than r can be easily accomodated. For example, manager (1,2) could receive reports from both managers (1,1) and (2,1).

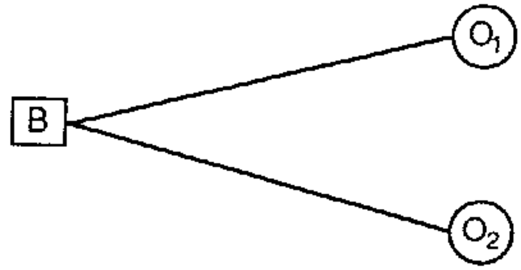

Fig 5. A simple network involving nonparallel roads

Commuters are exogenously located at the origins O_1 and O_2 Routes from origins O_1 and O_2 feed into a common bottleneck labeled B. Exiting from the bottleneck corresponds to arriving at the CBD. Travel time between each origin and the bottleneck is the same and independent of the number of commuters using the route. C_{grt} will depend on $(d_{ij})_{i\in\{1,2\},j\in\{1,\ldots,t\}}$ because waiting time in any queue at the bottleneck will depend on the total number of past departures as well as current departures on both routes. This case can be handled by amending the procedure of Figure 4 so that managers (1,1) and (2,1) jointly solve their problems, passing on the results to the managers controlling period 2's assignments, who then jointly solve their problems. While the derivation of the assignments for each time period will now be more difficult, other aspects of the procedure, in particular the dimensionality of the master problem, remain unchanged. One can also take advantage of the fact that the fixed point algorithm does not change the cost signals $(c_g)_{g\in G}$ very much from one iteration to the next. Consequently the subproblem for period t on successive iterations will not be very different, and a solution for it from the previous iteration can be used as a starting point in searching for a solution on the current iteration.

A giant step toward treating more complicated networks can be made by allowing C_{grt} to depend on $(d_{ij})_{i\in R,j\in T}$, i.e., C_{grt} depends on total departures on all routes in all time periods. We can amend the procedure of Figure 4 by expanding the number of signals sent from the central planner (see Figure 6). In addition to sending $(c_g)_{g\in G}$, suppose the central planner sends a vector of proposed total departures $(\bar{d}_{ij})_{i\in R,j\in T}$. Manager (1,1) would now be asked to solve the subproblem:

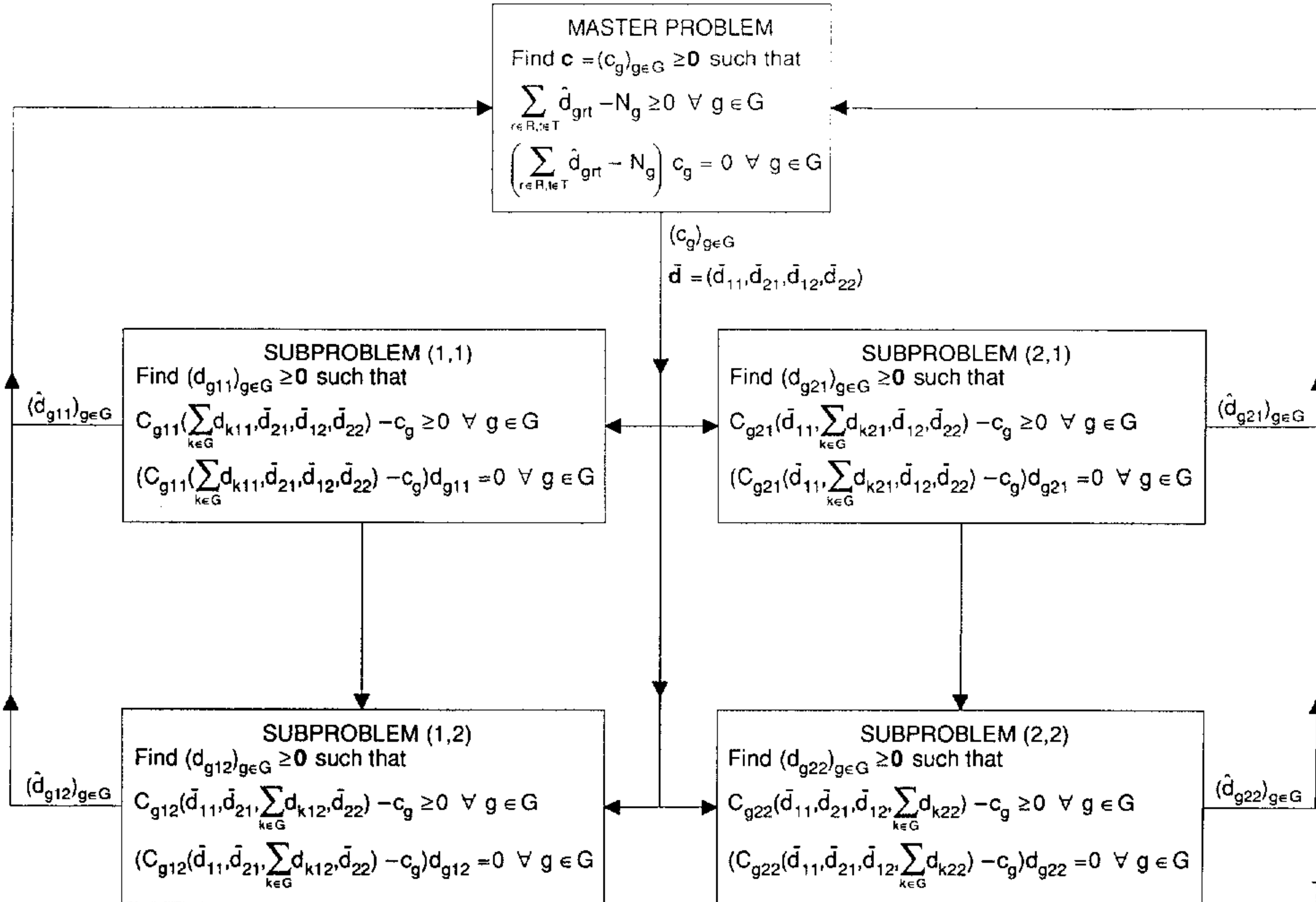

Fig. 6. A generalized procedure for handling more complicated networks

Find $(d_{g11})_{g\in G} \geq \mathbf{0}$ such that

$$C_{g11}(\sum_{k\in G} d_{k11}, \bar{d}_{21}, \bar{d}_{12}, \bar{d}_{22}) - c_g \geq 0 \quad \forall g \in G$$

$$(C_{g11}(\sum_{k\in G} d_{k11}, \bar{d}_{21}, \bar{d}_{12}, \bar{d}_{22}) - c_g) d_{g11} = 0 \quad \forall g \in G.$$

This manager, as in Figure 4, would only be responsible for assigning commuters to route 1 during interval 1. However, unlike Figure 4, this assignment would be conditional upon not only the cost signals sent from the central planner but also all of the coordinates of the total departure signal vector $\bar{\mathbf{d}}$ except the one pertaining to route 1 and time interval 1. Analogously, manager (r,t) would solve for the assignment vector $(d_{grt})_{g\in G}$ conditional on $(c_g)_{g\in G}$ and $\bar{\mathbf{d}}$, ignoring coordinate $\bar{d}_{rt}$. Another difference from Figure 4 is that the managers only report their decisions to the central planner, not to other managers.[30] As in Figure 4, the central planner receives the reports of the managers and then revises the cost signals with a fixed point algorithm. However, on the next iteration the central planner sends not only these revised cost signals but also a revised total departure vector $\bar{\mathbf{d}}$. The simplest procedure would be to set the new $\bar{\mathbf{d}}$ equal to the vector of total departures $(\sum_{k\in G} \hat{d}_{k11}, \sum_{k\in G} \hat{d}_{k21}, \sum_{k\in G} \hat{d}_{k12}, \sum_{k\in G} \hat{d}_{k22})$ corresponding to the managers' reports on the previous interation. The advantage of this procedure would be that the dimensionality of the master problem and the ease of solution of the subproblems would have been preserved from Figure 4. The disadvantage is that by not having the revision of the aggregate departure vector signal also under the control of the fixed point algorithm, the nice convergence properties of the procedure of Figure 4 may be lost.

[30]A variant of this procedure would have manager (1,1) report $\sum_{k\in G} \hat{d}_{k11}$ and manager (2,1) report $\sum_{k\in G} \hat{d}_{k21}$ to managers (1,2) and (2,2), and then have these managers use these reports in place of $\bar{d}_{12}$ and $\bar{d}_{22}$ in their subproblems.

The two alternative structures we have assumed for the cost functions are likely to bracket a broad range of interesting networks. Determining the best set of signals to send, and which ones to systematically revise, is the subject of ongoing research.

REFERENCES

1. H. Z. AASHTIANI AND T.L. MAGNANTI, "Equilibria on a Congested Transportation Network," *SIAM J. Alg. Disc. Meth.* **2**, 213-226 (1981).
2. R. ARNOTT, A. DE PALMA AND R. LINDSEY, "Schedule Delay and Departure Time Decisions with Heterogeneous Commuters," *Trans. Res. Record* **1197**, 56-67 (1989).
3. R. ARNOTT, A. DE PALMA AND R. LINDSEY, "Economics of a Bottleneck," *J. Urban Econ.* **27**, 111-130 (1990).
4. R. ARNOTT AND J. MACKINNON, "The Effects of the Property Tax: A General Equilibrium Simulation," *J. Urban Econ.* **4**, 389-407 (1977).
5. M. BEN-AKIVA, M. CYNA AND A. DE PALMA, "Dynamic Model of Peak Period Congestion," *Trans. Res.* **18B**, 339-355 (1984).
6. M. BEN-AKIVA, A. DE PALMA AND P. KANAROGLOU, "Dynamic Model of Peak Period Traffic Congestion with Elastic Arrival Rates," *Trans. Sci.* **20**, 164-181 (1986).
7. M. CAREY, "Optimal Time-Varying Flows on Congested Networks," *Operations Research* **35**, 56-69 (1987).
8. S. DAFERMOS, "Congested Transportation Networks and Variational Inequalities," in *Flow Control of Congested Networks* (Proceedings of the NATO Advanced Research Workshop on Flow Control of Congested Networks, Capri, Italy, 1986), pp. 1-11, A. Odoni, L. Bianco, G. Szego (eds.), Springer-Verlag, Berlin, 1987.
9. J.E. FERNANDEZ AND T. L. FRIESZ, "Equilibrium Predictions in Transportation Markets: The State of the Art," *Trans. Res.* **17B**, 155-173 (1983)
10. J.V. HENDERSON, *Economic Theory and the Cities,* Academic Press, New York, 1977.
11. C. HENDRICKSON AND G. KOCUR, "Schedule Delay and Departure Time Decisions in a Deterministic Model," *Trans. Sci.* **15**, 62-77 (1981).
12. H. MAHMASSANI AND R. HERMAN, "Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials," *Trans. Sci.* **18**, 362-384 (1984).

13. T.L. MAGNANTI, "Models and Algorithms for Predicting Urban Traffic Equilibrium," in *Transportation Planning Models*, M. Florian (ed.), North-Holland, Amsterdam, 1984.

14. A. NAGURNEY, "Computational Comparisons of Algorithms for General Asymmetric Traffic Equilibrium Problems with Fixed and Elastic Demands," *Trans. Res.* **20B**, 78-84 (1986).

15. G. NEWELL, "The Morning Commute for Nonidentical Travelers," *Trans. Sci.* **21**, 74-88 (1987).

16. D. RICHTER, "A Computational Approach to Resource Allocation in Spatial Urban Models," *Reg. Sci. Urban Econ.* **10**, 17-42 (1980).

17. H. SCARF (in collaboration with T. HANSEN), *The Computation of Economic Equilibria,* Yale University Press, New Haven, 1973.

18. M. SMITH, "The Existence of a Time-Dependent Equilibrium Distribution of Arrivals at a Single Bottleneck," *Trans. Sci.* **18**, 385-394 (1984).

19. M. TODD, *The Computation of Fixed Points and Applications* (Lecture Notes in Economics and Mathematical Systems 124), Springer-Verlag, Berlin, 1976.

20. M. TODD, "PLALGO: A Fortran Implementation of a Piecewise-Linear Homotopy Algorithm for Solving Systems of Nonlinear Equations," *Technical Report 452,* School of Operations Research and Industrial Engineeering, Cornell University, Apr. 1980, rev. July 1985.

21. W.S. VICKREY, "Congestion Theory and Transport Investment," *Amer. Econ. Rev.* **59**, 251-261 (1969).

22. J.G. WARDROP, "Some Theoretical Aspects of Raod Traffic Research," *Proceedings of the Institute of Civil Engineers*, Part II **1**, 325-378 (1952).