Transpn Res.-B, Vol. 13B, pp. 5–17. Pergamon Press 1979. Printed in Great Britain

FORMULATING AND SOLVING THE NETWORK DESIGN PROBLEM BY DECOMPOSITION

GEORGE B. DANTZIG
Stanford University, Stanford, California, U.S.A.

ROY P. HARVEY, ZACHARY F. LANSDOWNE and DAVID W. ROBINSON
Control Analysis Corporation, 800 Welch Road, Palo Alto, California, U.S.A.

and

STEVEN F. MAIER
Duke University, Durham, North Carolina, U.S.A.

(*Received* 11 *May* 1978)

Abstract—The optimal transportation network design problem is formulated as a convex nonlinear programming problem and a solution method based on standard traffic assignment algorithms is presented. The technique can deal with network improvements which introduce new links, which increase the capacity of existing links, or which decrease the free-flow (uncongested) travel time on existing links (with or without simultaneously increasing link capacity). Preliminary computational experience with the method demonstrates that it is capable of solving very large problems with reasonable amounts of computer time.

1. INTRODUCTION

The network design problem is concerned with the modification of a transportation infrastructure by adding new links or improving existing ones. Perhaps the problem's name is somewhat unfortunate, since in most applications one is interested in selecting from among a relatively small set of improvements to an existing network rather than in designing an entirely new network from scratch. The terminology is well-established, however, and we conform to accepted usage herein.

Network design problems arise in many transportation modes: urban mass transit, highway, rail, etc. Most applications have been in highway improvement, however. Approaches to solving the optimal network design problem may be classified in several ways:

(a) According to the objective function. This may be to assign trips to the improved network according to either Wardrop's first principle (user equilibrium) or Wardrop's second principle (system optimal).

(b) According to how the cost of making the improvements is reflected. The investments may appear in a budget constraint or be included (with an appropriate weight) directly in the objective function.

(c) According to the type of link cost function employed. Linear, piecewise linear, quadratic, polynomial and nonlinear differentiable cost functions have all been used by researchers.

(d) Whether the improvement decisions are discrete or continuous. In most applications discrete investments are most appropriate; it makes little sense to add, say, two-thirds of a lane to a freeway. Continuous investments are far more tractable mathematically, however, and so are often used.

A review of the many previous formulations of and solution algorithms for the network design problem is presented in Dantzig *et al.* (1976), Chapter 4. The present work grew from the requirement to apply the existing methodology to problems of practical size, e.g. 1500 nodes and 4000 arcs. When such large problems are to be solved certain assumptions must be made in order to obtain a tractable problem formulation. Among these are:

(a) The traffic assignment must be carried out according to the system optimal criterion (minimize total travel time for all users) even when this may not be completely appropriate (as in highway applications). In order to solve the user equilibrium traffic assignment problem as a convex mathematical program, it is necessary that an objective function which is not total travel time be used. This means that the network design problem must be solved by a branch-and-bound approach (see LeBlanc, 1975) which is impractical for large problems.

(b) The investment decisions must be continuous rather than discrete. Once again, this assumption is required to obviate the necessity for branch-and-bound solution methods. Furthermore, the cost of making the improvements must be represented by a convex function even though nonconvex components (e.g. fixed charges) are often encountered in practice.

These two assumptions are clearly restrictive; in Section 5, however, we discuss some ways in which they can be relaxed.

The remainder of this paper is organized as follows: we first discuss ways to reflect the impact of investment on a given link in the transportation infrastructure (Section 2). We then demonstrate how a standard traffic assignment algorithm can be used to solve optimal network design problems when investment dollars are included in the objective function (Section 3) or when they are limited by a budget constraint (Section 4). We then describe the application of our method to actual problems (Section 5) and then summarize our results (Section 6).

2. FORMULATION OF A COMBINED INVESTMENT AND CONGESTION TRAVEL TIME FUNCTION

An important step in obtaining a mathematical formulation of the network design problem is the proper treatment of the interaction between travel time, flow, and investment. Let $D_j(f_j, z_j)$ be the total travel time on link j as a function of the link flow f_j and investment decision z_j for that link. This section will discuss three ways of representing the function $D_j(f_j, z_j)$, each based on a different total travel time function.

A common approach to modelling the total travel time on a link as a function of flow on that link is the so-called BPR curve:

$$T_j(f_j) = t_jf_j\left[1 + r\left(\frac{f_j}{c_j}\right)^k\right], \tag{1}$$

where

$T_j(f_j)$ = total travel time for all users on link j
f_j = flow on link j
t_j = free-flow travel time parameter for link j
c_j = capacity parameter for link j
r = constant
k = constant.

For example, the Federal Highway Administration uses $r = 0.15$ and $k = 4$ for modelling highway congestion (see COMSIS, 1973). Steenbrink (1974a, b) used a similar function in his study of road investments in the Netherlands with $k = 5$ and 7.

We note that the "capacity parameter" c_j in (1) does not actually give the link capacity. When the flow increases beyond c_j, however, the congestion increases rapidly and so this value represents a limit of sorts on flow. We therefore refer to c_j as the "capacity" of link j in what follows. Note that a link which does not exist in the current infrastructure can be introduced with a very small capacity.

If we assume that the effect of investment on link j is to increase the capacity, we obtain for the $D_j(\cdot, \cdot)$ function

$$D_j(f_j, z_j) = t_jf_j\left[1 + r\left(\frac{f_j}{c_j + z_j}\right)^k\right], \tag{2}$$

where the investment variable z_j is measured in units of capacity. This is the type of function which Steenbrink used in his study. By computing partial derivatives, the reader can show that $D_j(\cdot, \cdot)$ is a convex function whenever $k \geq 0$, $r \geq 0$, $t_j \geq 0$ and $c_j \geq 0$.

In some network design applications, it may be desirable to have the investment effect the free-flow travel time on a link, rather than the link capacity, while in other applications it may be desirable to have the investment modify the free-flow travel time and capacity parameters simultaneously. For example, the effect of adding lanes to a road link will be to increase the capacity; but the additional lanes may also allow the speed limit to increase which effects the free-flow travel time. Unfortunately, it is not possible when using a differentiable travel time curve of the form (1) to allow the investment decision to lower the free-flow travel time. This is possible, however, when using a piecewise linear approximation for the travel time curve. Morlok *et al.* (1973) assumed that $T_j(f_j)$ could be approximated with a piecewise linear function with M_j breakpoints. Define the components x_j^m such that

$$T_j(f_j) = \min \sum_{m=1}^{m=M_j} C_j^m x_j^m,$$

with respect to x_j^m, $m = 1, \ldots, M_j$ subject to

$$f_j = \sum_{m=1}^{m=M_j} x_j^m,$$

and

$$0 \le x_j^m \le K_j^m \qquad (m = 1, \ldots, M_j),$$

where K_j^m are the segment lengths in the piecewise linear approximation and C_j^m are the slopes of the linear curves in the approximation.

To incorporate investment into the analysis, Morlok assumed that the only effect of investment on a link is to shift the locations of the breakpoints. Let z_j be the total capacity added to link j (measured in units of capacity), and let F_j^m be the proportion that is assigned to the mth increment. This means that the total travel time on link j as a function of flow f_j and investment z_j can be represented as

$$D_j(f_j, z_j) = \min \sum_{m=1}^{m=M_j} C_j^m x_j^m, \tag{3}$$

with respect to x_j^m, $(m = 1, \ldots, M_j)$, subject to

$$f_j = \sum_{m=1}^{m=M_j} x_j^m, \tag{4}$$

and

$$0 \le x_j^m \le K_j^m + F_j^m z_j, \qquad (m = 1, \ldots, M_j). \tag{5}$$

By using an argument similar to that given in Appendix C, it can be shown that $D_j(\cdot, \cdot)$ is a convex function.

The appropriate values for the multipliers F_j^m are determined by the lengths K_j^m. We will illustrate this procedure by considering the following example. Suppose that the original travel time curve is in the form (1) and that the breakpoints occur at flow values equal to $\alpha_1 c_j, \alpha_2 c_j, \ldots, \alpha_M c_j$, where c_j is the link capacity parameter and $\alpha_1 < \alpha_2 < \cdots < \alpha_M$. Thus

$$K_j^m = (\alpha_m - \alpha_{m-1})c_j \qquad \text{for } m = 1, \ldots, M,$$

where $\alpha_0 = 0$. If we take breakpoint values on the curve in the piecewise linear approximation, then the slopes C_j^m are independent of c_j and are multiples of the free-flow travel time t_j:

$$C_j^m = t_j \frac{\{\alpha_m[1 + r(\alpha_m)^k] - \alpha_{m-1}[1 + r(\alpha_{m-1})^k]\}}{\alpha_m - \alpha_{m-1}}$$

for $m = 1, \ldots, M$. The proper value for the multipliers F_j^m are given by

$$F_j^m = \alpha_m - \alpha_{m-1} \qquad \text{for } m = 1, \ldots, M.$$

Moreover, if we were to define new breakpoints $K_j^{m'} = K_j^m + F_j^m z_j$, then the breakpoints $K_j^{m'}$

and slopes C_j^m would provide the correct piecewise linear approximation to the travel time curve after c_j is replaced with $c_j + z_j$.

In the Morlok formulation, the effect of the investment decision is to change the breakpoints, but not the slopes of the piecewise linear approximation. If the original travel time curve were in the form (1), then this would correspond to increasing the capacity parameter c_j, but not the free-flow travel time t_j. If it were desired to affect either or both the capacity and free-flow travel time, then a simple extension is possible.

Suppose that at some maximum investment it is desired to have both a new higher capacity c_j' and a new lower freeflow travel time t_j'. As before, the first step is to represent the current travel time function (1), incorporating the current values for t_j and c_j, with segments having positive lengths K_j^m. The next step is to add additional segments whose slopes correspond to the new lower free-flow travel time t_j', but having initial lengths $K_j^m = 0$. These additional segments would have positive values for F_j^m, while the segments corresponding to the current t_j would have negative values for F_j^m. The constraint (5) then guarantees that the segments corresponding to the new t_j' will expand, while the segments corresponding to the current t_j will contract. The multipliers F_j^m are chosen so that the segments corresponding to the current t_j just vanish at the maximum investment, while the segments corresponding to the new t_j' have expanded to be the proper multiples of the new capacity c_j'.

This approach thus allows one to deal with improvements which lower the free-flow travel time, with or without a simultaneous increase in link capacity. Since the resulting $D(\cdot, \cdot)$ function is still convex, this avoids the difficulties that Steenbrink and others have encountered in trying to deal with free-flow travel time improvements.

The final functional form we will consider for $D_j(f_j, z_j)$ was proposed by Dafermos (1968). She examined functions of the form

$$D_j(f_j, z_j) = A_j(z_j)f_j^2 + B_j(z_j)f_j,$$

where the function $A_j(\cdot)$ was designed to incorporate interaction between vehicles and loosely corresponds to the capacity parameter c_j in eqn (1), while $B_j(\cdot)$ was designed to capture improvements in the free-flow travel characteristics of the network. In attempting to restrict $A_j(\cdot)$ and $B_j(\cdot)$ so that the resulting $D_j(\cdot, \cdot)$ function would be convex, she chose a function of the form

$$D_j(f_j, z_j) = a_j\left(\frac{l_j}{l_j + z_j}\right)^{\gamma_j} f_j^2 + B_j \cdot f_j \tag{6}$$

where $a_j > 0$, $l_j > 0$ and $0 \le \gamma_j \le 1$. Since B_j is independent of z_j, eqn (6) restricts the effect of investments to only link capacity.

In summary, the three functional forms for $D_j(f_j, z_j)$ are each convex, but only the piecewise linear form permits changes to both link capacity and free-flow travel time characteristics. The disadvantage of the piecewise linear function is that it does not possess derivatives for all values of f_j and z_j.

3. DECOMPOSITION SOLUTION OF NETWORK DESIGN PROBLEMS

This section will analyze the following problem: determine the optimal network design in order to minimize the sum of users' and investment costs, using a system optimal traffic assignment criterion. A budget constraint will not be incorporated into the formulation until Section 4.

In Section 2 we defined $D_j(f_j, z_j)$ to be the total travel time on link j as a function of link flow f_j and the investment decision z_j on the link. Let $G_j(z_j)$ be the investment cost (measured in dollars) associated with the investment decision z_j. We assume that $G_j(\cdot)$ is convex. Thus the total social transportation cost (user travel cost plus investment costs) for link j is equal to

$$D_j(f_j, z_j) + \lambda G_j(z_j),$$

where λ expresses the conversion between investment dollars and travel time.

The number of distinguishable flow commodities in this problem can be set equal to either the number of origin nodes or to the number of destination nodes. Suppose that commodities are distinguished by origin nodes. Let f_j^r be the total flow on link j that originates from node r.

The network design problem can then be stated as follows: determine the investment decisions z_j and flows f_j^r in order to minimize the total social transportation cost

$$\min \sum_{j \in A} D_j(f_j, z_j) + \lambda G_j(z_j), \tag{7}$$

with respect to f_j, f_j^r and z_j, subject to the conservation of flow equations defined for each node i and origin r

$$\sum_{j \in W_i} f_j^r - \sum_{j \in V_i} f_j^r = h_i^r \qquad (i \in N; r = 1, \ldots, R), \tag{8}$$

total flow in each link being equal to the sum of the flows from the sources

$$f_j = \sum_{r=1}^{r=R} f_j^r \qquad (j \in A), \tag{9}$$

upper and lower bounds on the maximum improvement possible

$$L_j \leq z_j \leq P_j \qquad (j \in A), \tag{10}$$

and non-negativity restrictions

$$f_j^r \geq 0 \qquad (j \in A; r = 1, \ldots, R), \tag{11}$$

where

A = set of links in the network
N = set of nodes in the network
R = number of origin nodes
f_j^r = flow on link j from origin r
f_j = total flow on link j
O_{ij} = number of trips from node i to node j

$$h_i^r = \begin{cases} -O_{ri}, \text{ if } i \text{ is a destination node} \\ \Sigma_j O_{rj}, \text{ if } i = r \\ 0, \text{ otherwise} \end{cases}$$

z_j = investment decision for link j
L_j = minimum value for the investment decision for link j
P_j = maximum value for the investment decision for link j
V_i = set of links terminating at node i
W_i = set of links originating at node i.

Any continuous convex functions $D_j(\cdot,\cdot)$ and $G_j(\cdot)$ can be used. We assume that the conversion factor λ is non-negative. For a given value of z_j, the objective function (7) minimizes the total travel time subject to the conservation of flow conditions; thus, the traffic is assigned according to the system optimal criterion. The upper bound P_j on investment could be set by a physical, technical, environmental, or financial constraint. The lower bound L_j may be useful to indicate projects that are already begun and for multistage applications. If $L_j = P_j = 0$, then no improvement is possible on link j.

In order to solve this formulation, we next consider a decomposition procedure devised by Steenbrink (1974a, b). Let

$I_j(f_j)$ = the optimal investment decision for link j as a function of link flow f_j,
$H_j(f_j)$ = the minimum travel and investment costs for link j as a function of the link flow f_j.

The approach begins by solving a separate subproblem for each link, which determines the functions $I_j(\cdot)$ and $H_j(\cdot)$. The function $H_j(\cdot)$ satisfies:

$$H_j(f_j) = \min_{z_j} [D_i(f_i, z_i) + \lambda G_j(z_j)] \tag{12}$$

subject to

$$L_j \le z_j \le P_j. \tag{13}$$

The function $I_j(f_j)$ is defined to be the value of z_j at which $H_j(f_j)$ attains its minimum. It might be thought that it is necessary to solve (12)–(13) for each value of f_i in order to construct the functions $H_j(\cdot)$ and $I_j(\cdot)$; however, we show in Appendices A and B that in many cases it is possible to derive closed-form expressions for $H_j(\cdot)$ and $I_j(\cdot)$ in terms of the input parameters.

The solutions of the subproblems (12)–(13) form the objective function for the master problem, and the purpose of the master problem is to compute the total flow in each link. Thus this problem becomes: determine f_j and f_j^r to

$$\min \sum_{j\in A} H_j(f_j) \tag{14}$$

subject to

$$\sum_{j\in W_i} f_j^r - \sum_{j\in V_i} f_j^r = h_i^r \qquad (i\in N; r = 1,\ldots, R), \tag{15}$$

$$\sum_{r=1}^{r=R} f_j^r = f_j \qquad (j\in A), \tag{16}$$

$$f_j^r \ge 0 \qquad (j\in A; r = 1,\ldots, R). \tag{17}$$

We show in Appendix C that $H_j(\cdot)$ will be convex whenever $D_j(\cdot,\cdot)$ and $G_j(\cdot)$ are continuous convex functions. If $H_j(\cdot)$ is convex, then (14)–(17) are equivalent to the traffic assignment problem, and thus it can be solved by any of a number of approaches; see Dantzig *et al.* (1976) and Harvey and Robinson (1977). It is clear that solving both the sub-problems (12)–(13) and the master problem (14)–(17) is equivalent to solving the original problem (7)–(11). Once the optimal value f_j is determined for the jth link, then the optimal investment for that link is given by $I_j(f_j)$.

The success of this particular approach to solving the network design problem depends on how easily the functions $H_j(\cdot)$ and $I_j(\cdot)$ can be obtained for each link, and whether $H_j(\cdot)$ is a convex function. Note that most traffic assignment algorithms can be modified to handle non-differentiable objective functions, although they are best suited for work with differentiable functions (see Dantzig *et al.*, 1976). However, all traffic assignment algorithms require $H_j(\cdot)$ to be convex in order to ensure that a global optimal solution will be obtained.

Corresponding to the three $D(\cdot.\cdot)$ functions developed in Section 2 we have investigated the following cases:

(a) *BPR curve with change in link capacity*

The non-linear differentiable function (1) is used to represent the total travel time on a link as a function of the capacity and free-flow travel time of that link. If we assume that the only effect of the investment decision is to change the link capacity, then $D_j(\cdot,\cdot)$ is given by (2). If the investment cost $G_j(\cdot)$ is a convex function, then $H_j(\cdot)$ is convex. Furthermore, if $G_j(\cdot)$ is linear, then $H_j(\cdot)$ is differentiable. In Appendix A we derive the formulas for $H_j(\cdot)$ and $I_j(\cdot)$ for the case in which $G_j(\cdot)$ is linear.

(b) *Piecewise linear travel time function with changes in either or both link capacity and free-flow travel time*

If a piecewise linear function is used to represent the total travel time on a link and if we assume that the only effect of investment is to shift the locations of the breakpoints, then $D_j(\cdot,\cdot)$ is given by (3)–(5). If $G_j(\cdot)$ is convex, then $H_j(\cdot)$ will be convex. In Appendix B we give the formulas for $H_j(\cdot)$ and $I_j(\cdot)$ for the case in which there are two segments in the approximation ($M_j = 2$) and $G_j(\cdot)$ is linear; however, $H_j(\cdot)$ will not be differentiable in this case. If the original travel time curve is of the form (1), we have shown how the piecewise linear representation can be used to model improvements in the free-flow travel time.

(c) *Quadratic travel time curve with change in link capacity*

The total travel time on a link is represented as a quadratic function of the flow on that link. If we assume that the only effect of investment is to change link capacity, then (6) can be used for $D_j(\cdot, \cdot)$. If $G_j(\cdot)$ is convex, then $H_j(\cdot)$ will be convex. The analysis for this case is similar to that given in Appendix A and therefore will be omitted.

4. INCLUSION OF A BUDGET CONSTRAINT

The network design formulation described in Section 3 minimized travel plus investment costs, but without a budget restriction. In contrast, the objective of the model in this section is to minimize users' travel costs subject to a budget constraint. Thus the formulation becomes: determine the investment decisions z_j and flows f_j^r in order to minimize total travel costs

$$\min \sum_{j \in A} D_j(f_j, z_j), \tag{18}$$

subject to the conservation of flow equations defined for each node i and origin r

$$\sum_{j \in W_i} f_j^r - \sum_{j \in V_i} f_j^r = h_i^r \qquad (i \in N; r = 1, \ldots, R), \tag{19}$$

total flow on each link being equal to the sum of the flows from the sources

$$f_j = \sum_{r=1}^{r=R} f_j^r \qquad (j \in A), \tag{20}$$

upper and lower bounds on the maximum improvement possible

$$L_j \le z_j \le P_j \qquad (j \in A), \tag{21}$$

non-negativity restrictions

$$f_j^r \ge 0 \qquad (j \in A; r = 1, \ldots, R), \tag{22}$$

and total investment costs limited by the available budget B

$$\sum_{j \in A} G_j(z_j) \le B, \tag{23}$$

where the input parameters are defined in Section 3.

Any continuous convex functions $D_j(\cdot, \cdot)$ and $G_j(\cdot)$ can be used. For example, if $D_j(\cdot, \cdot)$ is specified by the piecewise linear approximation (3)–(5) and $G_j(\cdot)$ is linear, then the model (18)–(23) is basically the one formulated by Morlok *et al.* (1973). If $D_j(\cdot, \cdot)$ is given by (6) and $G_j(\cdot)$ is linear, then (18)–(23) is basically the model proposed by Dafermos (1968).

Our approach for solving the design problem (18)–(23) is illustrated in Fig. 1 and involves using a Lagrange multiplier technique to handle the budget constraint (23). As Everett (1963) has shown, this approach will yield a good approximate answer. Implicit in Everett's approach is the need to update trial values of the Lagrange multipliers until the budget constraint is approximately satisfied. Brooks and Geoffrion (1966) show how to do this systematically with linear programming. As seen in Zangwill (1969), this can be thought of as the Dantzig–Wolfe decomposition method (Dantzig, 1963; Dantzig and Wolfe, 1960). However, we suggest using the method of Fox and Landi (1970), which is to employ a binary sequential search procedure, sometimes called Bolzano's method (see Wilde and Beightler, 1967). The problem of finding the appropriate Lagrange multiplier is equivalent to finding the zero-crossing of a monotone

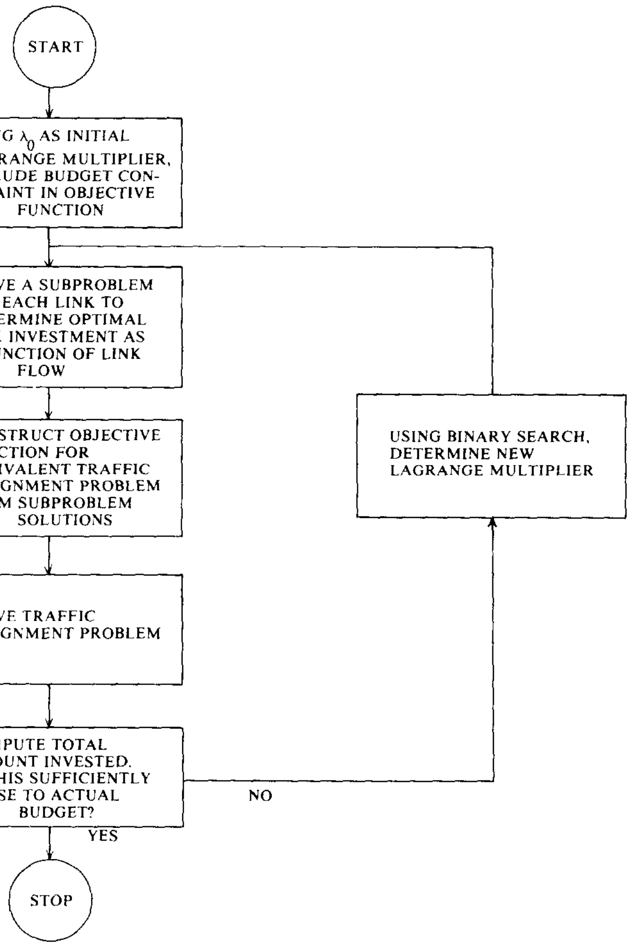
START
USING λ_0 AS INITIAL LAGRANGE MULTIPLIER, INCLUDE BUDGET CONSTRAINT IN OBJECTIVE FUNCTION
SOLVE A SUBPROBLEM FOR EACH LINK TO DETERMINE OPTIMAL LINK INVESTMENT AS A FUNCTION OF LINK FLOW
CONSTRUCT OBJECTIVE FUNCTION FOR EQUIVALENT TRAFFIC ASSIGNMENT PROBLEM FROM SUBPROBLEM SOLUTIONS
SOLVE TRAFFIC ASSIGNMENT PROBLEM
COMPUTE TOTAL AMOUNT INVESTED. IS THIS SUFFICIENTLY CLOSE TO ACTUAL BUDGET?
NO
YES
USING BINARY SEARCH, DETERMINE NEW LAGRANGE MULTIPLIER
STOP

Fig. 1. Solution of network design model with budget constraint using decomposition.

function, although in our problem this function may be discontinuous; thus there may be no multiplier which yields a total investment equal to the specified budget. A sequential search is a procedure that evaluates the monotone function at a succession of points which are determined by the outcomes of the preceding evaluations; at each step, the interval of uncertainty is reduced, and the process either repeats or terminates. A minimax sequential search is a scheme that minimizes the maximum length of the interval remaining after a fixed number of steps. The binary method proceeds by successively halving the interval of uncertainty. Fox and Landi (1970) showed that the binary method was the unique minimax sequential search procedure for finding the zero-crossing of a monotone function known to lie in a given interval.

Next, we will outline this Lagrange multiplier approach in more detail. Let w represent the vector (f_j^r, z_j) of decision variables, S be the set of values for w satisfying the constraints (19)–(22), $Q(w)$ represent the objective function (18), and $R(w) \le B$ represent the budget constraint (23). Thus (19)–(23) can be rewritten as:

$$\min_{w \in S} Q(w), \qquad \text{subject to } R(w) \le B. \tag{24}$$

Using Everett's method (1963), problem (24) can be (approximately) solved by finding a multiplier λ and the corresponding solution $w(\lambda)$ to

$$\min_{w \in S} [Q(w) + \lambda R(w)]. \tag{25}$$

It can be shown that $R[w(\lambda)]$ is monotone with respect to λ, although possibly discontinuous.

The binary method proceeds by generating a series of values for λ and solving (25) for each value, until the total investment $R[w(\lambda)]$ is approximately equal to the available budget B. Because $R[w(\lambda)]$ may be discontinuous, there may be no multiplier λ for which $R[w(\lambda)] = B$; thus several criteria should be used for stopping the search procedure: whenever $R[w(\lambda)]$ is sufficiently close to B; or whenever the interval of uncertainty for λ is sufficiently small.†

What makes this approach efficient is that problem (25) is identical to the model (7)–(11) formulated in Section 3. Thus, the decomposition approach discussed in Section 3 shows that problem (25) can be solved by using a traffic assignment algorithm, assuming that the objective function for the master problem has the appropriate convexity. In other words, the network design model with budget constraint can be solved by solving a sequence of traffic assignment problems, one for each value of the multiplier.

5. APPLICATION TO PRACTICAL PROBLEMS

The decomposition procedure discussed in Sections 2–4 has been embedded in a traffic assignment code which uses the Frank–Wolfe algorithm to carry out the system optimal traffic assignment. The code is called CATNAP (for Control Analysis Network Analysis Program) and is extensively described in Harvey and Robinson (1977). In this Section we briefly summarize some preliminary experience with the CATNAP program and we also discuss some means provided in the code for dealing with two restrictive assumptions in Section 1.

A network design problem having 394 nodes, 1042 arcs, 84 origin nodes and 103 possible investments has been used to test the performance of CATNAP's network design capabilities. While this is not a particularly large problem in practical applications, it is far bigger than most previously reported work in this area. The test problem included a budget constraint which restricted the final solution to about 30 investments. (CATNAP has the ability to deal with the simpler formulation of Section 3, as well.)

The test problem was solved in roughly 5 min of CPU time on an IBM 370 Model 168 computer at a cost of about $90. By way of comparison, the linear programming formulation of Morlok *et al.* (1973) required over 40 min of CPU time for a problem with just 24 nodes and 76 arcs. The costs for a CATNAP network design solution are thus substantially lower than for previously suggested approaches and in fact are comparable to those for a simple traffic assignment problem solution.

In order to solve actual problems in many applications it is necessary to consider methods for dealing with networks where traffic is distributed according to a user equilibrium criterion. The difficulty in using the system optimal network design solution in these cases is that there is no way to determine how much better the user equilibrium solution would be (if we could obtain it).

For those instances in which user equilibrium assignment would be preferable in the design application (such as for highway networks), however, it can be shown that the minimum total cost (total travel time plus possibly the investment costs) for the optimal network design with user equilibrium assignment will be bounded by two numbers: from below, by the total cost for the optimal network design with system optimal assignment; and from above, by the total cost when applying user equilibrium traffic assignment to the optimal network design that was determined with systems optimal assignment. In other words, bounds on the user equilibrium network design problem can be computed from the solution of the systems optimal network design problem. When these bounds are close (perhaps within 3%), this would justify the use of system optimal traffic solution as a valid approximation to user equilibrium solution. The CATNAP program is capable of carrying out the final user equilibrium traffic assignment in the modified network and can thus be used to generate the required bounds. For the 394 node test problem the bounds were within 7% which does not provide as clear cut a justification as might be wished for using the system optimal solution. More work remains to be done in this area.

A second requirement for solving practical problems is to reflect the discrete nature of some investment possibilities. Of course, some improvements are actually continuous; one can repave as much or as little of a road link as desired to obtain an increase in travel speed, for

†However, by taking the convex combination of two solutions $w(\lambda)$ whose investments $R[w(\lambda)]$ straddle the desired budget B, it is always possible to obtain a feasible (but not necessarily optimal) investment schedule whose total cost is equal to B.

example. However, most improvements are by their nature "all-or-nothing". To deal with this situation, CATNAP employs a heuristic rounding technique (which is fully described in Harvey and Robinson (1977)) to select a feasible set of discrete investments. Computational experience with the heuristic has thus far been limited by the absence of an actual test problem but the preliminary results are reasonable.

The assumption of a convex-cost investment function is essential to guarantee convergence for the decomposition methodology of Section 3. CATNAP therefore uses the convex hull of the input investment functions to solve the network design problem, replacing them with the actual functions for a final solution evaluation.

6. CONCLUSIONS

In summary, the network design approach described in this paper has the following characteristics:

(a) *Continuous investment decision variables.* The algorithm determines the optimal solution based upon continuous decision variables. If a discrete solution is needed for a particular application, the approach given in Section 5 may be adopted.

(b) *Systems optimal traffic assignment.* The formulation is based upon systems optimal traffic assignment, which is the preferred assignment in rail or mass transit applications; but user equilibrium assignment would be preferred in highway applications. The discussion in Section 5 shows how to obtain bounds on the user equilibrium design problem using only the solution to the systems optimal design problem; if these bounds are close, there is justification for using the network design based on systems optimal assignment in an application in which user equilibrium assignment is preferred.

(c) *Travel time as a function of flow and investment.* The model assumes that $D_j(f_j, z_j)$, which is the total travel time on link j as a function of the link flow f_j and investment decision z_j for that link, is a continuous convex function. This includes as special cases the non-linear differentiable curve, similar to the BPR travel time function, used by Steenbrink (1974a); the piecewise linear curve used by Morlok (1973); and the quadratic curve used by Dafermos (1968).

(d) *Investment cost function.* The model assumes that $G_j(z_j)$, which is the cost for making investment z_j on link j, is a continuous convex function.

(e) *Investment alternatives.* If the only effect of investment is to increase the capacity of existing links, this can be modeled with a differentiable travel time curve; if the effect of investment is to change the free-flow travel time (with or without a capacity change) this can be done with a piecewise linear travel time curve. Either approach can also handle the introduction of entirely new links by specifying their initial capacity as being effectively zero, i.e. some very small value.

(f) *Solution algorithm.* For the case in which there is no budget constraint but the investment cost is included in the objective function, Section 3 shows how the solution to the network design problem can be obtained by solving a traffic assignment problem. For the case in which a budget constraint is used, Section 4 demonstrates how a Lagrange multiplier technique can be used to obtain a solution to the network design problem by solving a series of traffic assignment problems, one for each value of the multiplier.

The computational experience gained thus far with this method indicates that it is far more efficient than previous formulations and that it is capable of solving a broad class of practical transportation planning problems. Further work in this area is currently in progress and will be reported elsewhere.

Acknowledgements—This research was sponsored by the Transportation Advanced Research Project (TARP) under the auspices of the Office of the Secretary, U.S. Department of Transportation under contract No. DOT-TSC-1059.

REFERENCES

Brooks R. B. S. and Geoffrion A. M. (1966) Finding Everett's Lagrange multipliers by linear programming. *Operations Res.* **14**, 1149–1153.

Comsis Corporation (1973) *Traffic Assignment*, prepared for Urban Planning Division, Office of Highway Planning, U.S. Federal Highway Administration, Washington, D.C.

Dafermos S. C. (1968) *Traffic Assignment and Resource Allocation In Transportation Networks*, Ph.D. Thesis, The John Hopkins University, Baltimore, MD.

Dantzig G. B. (1963) *Linear Programming and Extensions.* Princeton University Press, Princeton, NJ.

Dantzig G. B., Maier S. F. and Landsdowne Z. F. (1976) *The Application of Decomposition to Transportation Network Analysis,* Control Analysis Corporation Tech. Rep. No. DOT-TSC-OST-76-26, Palo Alto, CA.

Dantzig G. B. and Wolfe P. (1960) Decomposition principle for linear programs. *Operations Res.* **8**, 101–111.

Everett H. (1963) Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Res.* **11**, 399–417.

Fox B. L. and Landi D. M. (1970) Searching for the multiplier in one-constraint optimization problems. *Operations Res.* **18**, 253–262.

Harvey R. P. and Robinson D. W. (1977) *Computer Code for Transportation Network Design and Analysis,* Control Analysis Corporation Tech. Rep. No. DOT-TSC-OST-77-39, Palo Alto, CA.

LeBlanc L. J. (1975) An algorithm for the discrete network design problem. *Transpn Sci.* **9**, 183–199.

Morlok E. K., Schofer J. L., Pierskalla W. P., Marsten R. E., Agarwal S. K., Edwards J. L., LeBlanc L. J., Spacek D. T. and Stoner J. W. (1973) *Development and Application of a Highway Network Design Model,* (Final Report prepared for Federal Highway Administration, Environmental Planning Branch), Department of Civil Engineering, Northwestern University, Evanston, IL.

Newel G. F. (1974) Optimal network geometry, *Proc. 6th Int. Sym. Transpn Traffic Flow Theory.* American Elsevier, New York, 561–580.

Steenbrink P. A. (1974a) Transport network optimization in the Dutch integral transportation study. *Transpn Res.* **8**, 11–27.

Steenbrink P. A. (1974b) *Optimization of Transport Networks.* Wiley, New York.

Wilde D. J. and Beightler C. S. (1967) *Foundation of Optimization.* Prentice-Hall, Englewood Cliffs, NJ.

Zangwill W. I. (1969) *Nonlinear Programming, A Unified Approach.* Prentice-Hall, Englewood Cliffs, NJ.

APPENDIX A

Derivation of the investment cost function-differentiable case

Consider the case in which the travel time curve is in the form of eqn (1) and that the only effect of the investment decision is to change the capacity parameter c_j. Section 2 showed that the total travel time on link j as a function of the flow f_j and investment decision z_j is given by

$$D_j(f_j, z_j) = t_jf_j\left[1 + r\left(\frac{f_j}{c_j + z_j}\right)^k\right],$$

where the variable z_j is measured in units of capacity.

For this case, subproblem (12)–(13) becomes:

$$H_j(f_j) = \min_{z_j}\left\{t_jf_j\left[1 + r\left(\frac{f_j}{c_j + z_j}\right)^k\right] + \lambda G_j(z_j)\right\} \tag{A1}$$

subject to

$$L_j \le z_j \le P_j. \tag{A2}$$

The function $I_j(f_j)$ is defined as the value of z_j at which $H_j(f_j)$ attains its minimum. In Appendix C we show that if the investment cost $G_j(\cdot)$ is a convex function, then $H_j(\cdot)$ will be convex also.

We now derive explicit formulas for $H_j(\cdot)$ and $I_j(\cdot)$ for the case of a linear investment cost function: $G_j(z) = g_jz_j$. Define

$$h_j(f_j, z_j) = t_jf_j\left[1 + r\left(\frac{f_j}{c_j + z_j}\right)^k\right] + \lambda g_jz_j. \tag{A3}$$

If the optimal investment decision z_j occurs at an interior point, then it will satisfy the equation

$$\frac{\partial h_j}{\partial z_j} = -\frac{krt_j(f_j)^{k+1}}{(c_j + z_j)^{k+1}} + \lambda g_j = 0. \tag{A4}$$

The parameters g_j, k, r and t_j are fixed for any given link j; thus to simplify notation, define

$$\phi_j(\lambda) = \left(\frac{\lambda g_j}{krt_j}\right)^{[1/(k+1)]}. \tag{A5}$$

It follows that the solutions to the subproblem (A1)–(A2) can be computed as:

$$I_j(f_j) = \begin{cases} L_j & 0 \le f_j \le (c_j + L_j)\phi_j(\lambda) \\ \phi_j(\lambda)^{-1}f_j - c_j & (c_j + L_j)\phi_j(\lambda) \le f_j \le (c_j + P_j)\phi_j(\lambda) \\ P_j & f_j \ge (c_j + P_j)\phi_j(\lambda) \end{cases} \tag{A6}$$

$$H_j(f_j) = \begin{cases} t_jf_j\left[1 + r\left(\dfrac{f_j}{c_j + L_j}\right)^k\right] + \lambda g_jL_j & 0 \le f_j \le (c_j + L_j)\phi_j(\lambda) \\ t_jf_j[1 + (k + 1)r\phi_j(\lambda)^k] - \lambda g_jc_j & (c_j + L_j)\phi_j(\lambda) \le f_j \le (c_j + P_j)\phi_j(\lambda) \\ t_jf_j\left[1 + r\left(\dfrac{f_j}{c_j + P_j}\right)^k\right] + \lambda g_jP_j & f_j \ge (c_j + P_j)\phi_j(\lambda). \end{cases} \tag{A7}$$

By computing the derivative, the reader can verify that $\mathrm{d}H_j/\mathrm{d}f_j$ is continuous and nondecreasing, which implies that $H_j(\cdot)$ is convex and differentiable. It is possible to extend these formulas for $H_j(\cdot)$ and $I_j(\cdot)$ to handle a piecewise linear convex investment cost function $G_j(\cdot)$, in which case $H_j(\cdot)$ will still be convex, but no longer differentiable.

The next step is to solve the master problem (14)–(17) using the foregoing expression for $H_j(\cdot)$. Because $H_j(\cdot)$ is convex, this master problem can be solved by any one of a number of traffic assignment algorithms. Once the optimal flow f_j is determined for the jth link, the optimal investment for that link is given by (A6).

APPENDIX B

Derivation of the investment cost function—piecewise linear case

In this appendix we consider the case in which the piecewise linear approximation is used for the travel time curve and the only effect of investment is to shift the locations of the breakpoints. Section 2 showed that the total travel time on link j as a function of the flow f_j and investment decision z_j (measured in units of capacity) is given by (3)–(5).

For this case, subproblem (12)–(13) becomes:

$$H_j(f_j) = \min_{x_j^m, z_j} \left\{ \sum_{m=1}^{m=M_j} \mathrm{C}_j{}^m \mathrm{x}_j{}^m + \lambda \mathrm{G}_j(\mathrm{z}_j) \right\} \tag{B1}$$

subject to

$$\sum_{m=1}^{m=M_j} x_j^m = f_j, \tag{B2}$$

$$x_j^m \le K_j^m + F_j^m z_j \qquad (m = 1, \ldots, M_j), \tag{B3}$$

$$L_j \le z_j \le P_j, \tag{B4}$$

$$z_j \ge 0, \qquad x_j^m \ge 0 \qquad (m = 1, \ldots, M_j). \tag{B5}$$

The function $I_j(f_j)$ is defined to be the value of z_j at which $H_j(f_j)$ attains its minimum. It can be shown that if $G_j(\cdot)$ is convex, then $H_j(\cdot)$ is convex; however, $I_j(\cdot)$ need not be convex. Refer to Appendix C for the proof of the convexity of $H_j(\cdot)$.

The role of the subproblems is to construct the appropriate objective function for the master problem. If $G_j(\cdot)$ is piecewise linear, then both $H_j(\cdot)$ and $I_j(\cdot)$ are piecewise linear, and furthermore the breakpoints and slopes for these functions can be determined directly in terms of the input parameters. We will illustrate this procedure for the following problem: two linear segments for the travel time function, $M_j = 2$; and linear investment cost function, $G_j(z_j) = g_j z_j$. We observe that because of the convexity of the travel time function, $C_j^1 \le C_j^2$. For the case

$$C_j^1 + \frac{\lambda g_j}{F_j^1} < C_j^2,$$

it can be shown that the following formulas for $H_j(\cdot)$ and $I_j(\cdot)$ give the solution to the subproblem (B1)–(B5):

$$H_j(f_j) = \begin{cases} C_j^1 f_j + \lambda g_j L_j & \text{for } f_j \le K_j^1 + F_j^1 L_J \\ C_j^1 f_j + \dfrac{(f_j - K_j^1)\lambda g_j}{F_j^1} & K_j^1 + F_j^1 L_j < f_j \le K_j^1 + F_j^1 P_j \\ C_j^1(K_j^1 + F_j^1 P_j) + C_j^2(f_j - K_j^1 P_j - F_j^1 P_j) + P_j \lambda g_j & K_j^1 + F_j^1 P_j < f_j \le K_j^1 + K_j^2 + (F_j^1 + F_j^2) P_j \end{cases} \tag{B6}$$

and

$$I_j(f_j) = \begin{cases} L_j & \text{for } f_j \le K_j^1 + F_j^1 L_j \\ \dfrac{f_j - K_j^1}{F_j^1} & K_j^1 + F_j^1 L_j < f_j \le K_j^1 + F_j^1 P_j \\ P_j & K_j^1 + F_j^1 P_j < f_j \le K_j^1 + K_j^2 + (F_j^1 + F_j^2) P_j. \end{cases} \tag{B7}$$

And for the case

$$C_j^1 + \frac{\lambda g_j}{F_j^1} \ge C_j^2,$$

it can be shown that

$$H_j(f_j) = \begin{cases} C_j^1 f_j + \lambda g_j L_j & \text{for } f_j \le K_j^1 + F_j^1 L_j \\ C_j^1(K_j^1 + F_j^1 L_j) + \lambda g_k L_j & f_j \le K_j^1 + K_j^2 + (F_j^1 + F_j^2) L_j \\ \quad + C_j^2(f_j - K_j^1 - F_j^1 L_j) & f_j > K_j^1 + F_j^1 L_j \\ C_j^1\left(K_j^1 + F_j^1 \dfrac{f_j - K_j^1 - K_j^2}{F_j^1 + F_j^2}\right) & \begin{array}{l} f_j \le K_j^1 + K_j^2 + (F_j^1 + F_j^2) P_j \\ f_j \ge K_j^1 + K_j^2 + (F_j^1 + F_j^2) L_j \end{array} \\ \quad + C_j^2\left(K_j^2 + F_j^2 \dfrac{f_j - K_j^1 - K_j^2}{F_j^1 + F_j^2}\right) & \\ \quad + \dfrac{f_j - K_j^1 - K_j^2}{F_j^1 + F_j^2} \lambda g_i & \end{cases} \tag{B8}$$

and

$$I_j(f_j) = \begin{cases} L_j & \text{for } f_j \leq K_j^1 + K_j^2 + (F_j^1 + F_j^2)L_j \\ \dfrac{f_j - K_j^1 - K_j^2}{F_j^1 + F_j^2} & f_j \leq K_j^1 + K_j^2 + (F_j^1 + F_j^2)P_j \\ & f_j \geq K_j^1 + K_j^2 + (F_j^1 + F_j^2)L_j. \end{cases} \tag{B9}$$

The next step is to solve the master problem (14)–(17), using the above expression for $H_j(\cdot)$, in order to determine the optimal flow values f_j. Because $H_j(\cdot)$ is convex, (14)–(17) is equivalent to the traffic assignment problem with a non-differentiable objective function, and thus it can be solved by the Frank–Wolfe method (see Dantzig *et al.* 1976). The optimal investment on link j is then given by $I_j(f_j)$.

APPENDIX C

Proof of the convexity of $\mathrm{H}_j(\cdot)$

In the decomposition technique given in Section 3, it is necessary for the objective function of the master problem, denoted as $H_j(\cdot)$, to be convex in order to solve the master problem by a standard traffic assignment algorithm. We have defined $D_j(\cdot, \cdot)$ to be the total travel time on link j as a function of investment and flow on that link and have defined $G_j(\cdot)$ to be the investment cost function for link j. We will show in this Appendix that if $D_j(\cdot, \cdot)$ and $G_j(\cdot)$ are both continuous convex functions, then $H_j(\cdot)$ will be convex also. Note that formulas (2), (3)–(5) and (6) given for $D_j(\cdot, \cdot)$ in Section 1 are convex.

The function $H_j(\cdot)$ is defined as follows:

$$H_j(f_j) = \min_{z_j} [D_j(f_j, z_j) + \lambda G_j(z_j)] \tag{C1}$$

subject to

$$L_j \leq z_j \leq P_j. \tag{C2}$$

Theorem

Assume that $D_j(f_j, z_j)$ and $G_j(z_j)$ are continuous convex functions defined for $f_j \geq 0$ and z_j satisfying (C2), that λ is nonnegative, and that L_j and P_j are finite; then $H_j(\cdot)$ is convex.

Proof. By the assumptions,

$$h_j(f_j, z_j) = D_j(f_j, z_j) + \lambda G_j(z_j) \tag{C3}$$

is a continuous convex function and is defined for $f_j \geq 0$ and z_j satisfying (C2). Suppose $f_j^1 \geq 0$, $f_j^2 \geq 0$ and $0 \leq \alpha \leq 1$. Because a continuous function on a compact set attains its minimum value, choose z_j^i satisfying (C2) so that $H_j(f_j^i) = h_j(f_j^i, z_j^i)$. Thus

$$\alpha H_j(f_j^1) + (1-\alpha)H_j(f_j^2) = \alpha h_j(f_j^1, z_j^1) + (1-\alpha)h_j(f_j^2, z_j^2) \tag{C4}$$

$$\geq h_j[\alpha(f_j^1, z_j^1) + (1-\alpha)(f_j^2, z_j^2)] \tag{C5}$$

$$\geq H_j[\alpha f_j^1 + (1-\alpha)f_j^2]. \tag{C6}$$

Equation (C4) follows from the definition of z_j^i, (C5) follows from the convexity of $h_j(\cdot, \cdot)$ and (C6) from the definition of $H_j(\cdot)$ given in (C1) and (C2). Thus $H_j(\cdot)$ is convex, which completes the proof.