REPORT NO. DOT-TSC-OST-76-26

THE APPLICATION OF DECOMPOSITION TO TRANSPORTATION NETWORK ANALYSIS

G. B. Dantzig
S. F. Maier
Z. F. Lansdowne

Control Analysis Corporation
Palo Alto CA

OCTOBER 1976
INTERIM REPORT

Prepared for

U.S. DEPARTMENT OF TRANSPORTATION
OFFICE OF THE SECRETARY
Office of the Assistant Secretary for
Systems Development and Technology
Office of Systems Engineering
Washington DC 20590

Technical Report Documentation Page

1. Report No. DOT-TSC-OST-76-26	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle THE APPLICATION OF DECOMPOSITION TO TRANSPORTATION NETWORK ANALYSIS		5. Report Date October 1976 6. Performing Organization Code
7. Author(s) G.B. Dantzig, S.F. Maier, Z.F. Lansdowne		8. Performing Organization Report No. DOT-TSC-OST-76-26
9. Performing Organization Name and Address Control Analysis Corporation* Palo Alto CA 94304		10. Work Unit No. (TRAIS) OS650/R6511 11. Contract or Grant No. DOT-TSC-1059
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Secretary Office of the Asst. Sec. for Sys. Dev. and Tech. Office of Systems Engineering Washington DC 20590		13. Type of Report and Period Covered Interim Report July 1975 - January 1976 14. Sponsoring Agency Code
15. Supplementary Notes *Under contract to: U.S. Department of Transportation, Transportation Systems Center, Kendall Square, Cambridge MA 02142		

16. Abstract

This document reports preliminary results of five potential applications of the decomposition techniques from mathematical programming to transportation network problems. The five application areas are 1) the traffic assignment problem with fixed demands, 2) the traffic assignment problem with elastic demand, 3) the transportation network improvement problem, 4) the optimal staging of transportation investments over time, and 5) the geographic decomposition of the traffic assignment problem. For all five, proposed solution techniques are presented and compared with previous work.

17. Key Words Decomposition Network Design Network Analysis Traffic Assignment		18. Distribution Statement DOCUMENT IS AVAILABLE TO THE U.S. PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 126	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

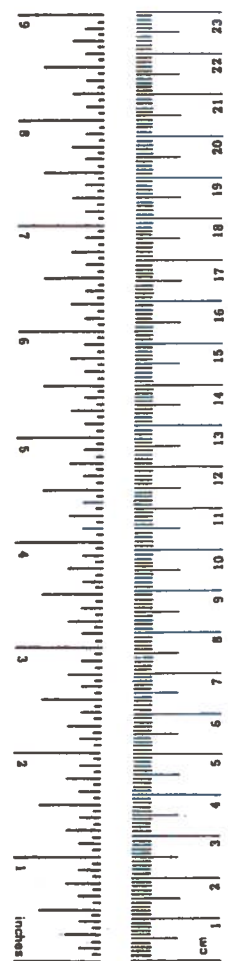
PREFACE

The research herein reported was funded by the Transportation Advanced Research Program (TARP) under the auspices of the Office of the Secretary, Department of Transportation. Technical review is the responsibility of the Special Studies Branch, Research Division, Transportation Systems Center. The objective of the TARP program is to stimulate basic scientific research in areas that are of major importance to the Department of Transportation. This particular project is intended to develop computerizable algorithms that will permit network analysis techniques to be applied to very large networks.

METRIC CONVERSION FACTORS

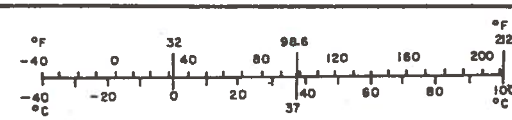
Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



CONTENTS

CONTENTS (Cont'd)

CONTENTS (Cont'd)

ILLUSTRATIONS

TABLES

1. INTRODUCTION AND SUMMARY

1.1 BACKGROUND

Many of the network problems of current practical interest to the Department of Transportation (DOT) are too large to be solved directly: they must be solved by either abstraction, extraction, or decomposition. Abstraction refers to combining groups of nodes and arcs in order to obtain a smaller network. Extraction refers to selecting a subnetwork and analyzing it in isolation. And decomposition, which is the subject of this report, refers to dividing the original network into subnetworks, which are then analyzed partly in isolation and partly in combination. In addition to being decomposed by subnetworks, a problem may also be decomposed by mode, commodity, or time period. Note that decomposition provides the actual solution to the original problem, whereas abstraction and extraction only provide approximate solutions. A variety of decomposition techniques have been devised for various types of problems. This report will show how decomposition can be applied to five specific network problems in transportation:

a) traffic assignment with fixed demands;

b) traffic assignment with elastic demands;

c) network design;

d) optimal staging of investments over time;

e) sub-area focusing.

For each of these network problems, Table 1.1 briefly describes the problem, the decomposition solution technique, and some of the potential applications. The remainder of this introductory section summarizes our results for each of these problem areas. In all cases, we limited ourselves to problem formulations and solution techniques which could handle large network problems: such as 1,000

TABLE 1.1

SUMMARY OF TRANSPORTATION PROBLEMS STUDIED

Problem	Description	Decomposition Solution Method	Potential Application
Traffic Assignment-Fixed Demand	Distribute given interzonal trip demands to alternate routes between these zones. The objective is to obtain either a user equilibrium or a systems optimal assignment. The demands are fixed and do not respond to changes in travel costs.	Solve a subproblem for each source using a spanning tree or shortest path algorithm. This method can be further decomposed by geographic regions using Generalized Benders decomposition. Convex GUB may be required to speed up convergence.	a. Mass transit planning. b. Intercity highway planning. c. Rail network planning and analysis of inter-company competition. d. Real-time highway traffic routing and freight car management. e. Commodity movements. f. Air traffic flow analysis.
Traffic Assignment-Elastic Demand	This is the same as above, except that the origin-destination demands do respond to changes in travel costs.	Expand dimensions of network to obtain an equivalent fixed demand traffic assignment problem, which is then solved as described above.	Same as above.
Network Design	Determine the optimal investments with which to expand a transportation network to minimize user travel costs subject to a budget constraint. Investments may be used to expand the capacity of existing links or to introduce new links.	Use Lagrange multipler technique to handle budget constraint. Solve a sub-problem for each link to construct objective function for master, which is solved by a traffic assignment algorithm. This method can be further decomposed by geographic regions using Generalized Benders decomposition.	a. Intercity highway planning; improvement and new construction. b. Rail line rehabilitation; analysis of loan guarantee and subsidy process. c. Multi-modal competition and interactions; analysis of legislation, investments, incentives; planning for transportation for energy commodities.
Optimal Staging of Investments over Time	Determine the optimal sequence or staging of network investment decisions over a planning horizon to minimize user travel costs subject to a budget constraint for each stage of the horizon. The final network configuration at the end of horizon may or may not be specified.	Use dynamic programing to decompose problem into a series of network design problems. This method can be further decomposed by geographic regions using Generalized Benders decomposition.	Same as above.
Geographic Decomposition and Sub-Area Focusing	In the sub-area focusing problem, a sub-area is extracted from a very large network so that detailed changes to this sub-area can be examined. Assign traffic within this sub-area by approximating how demand external to the sub-area will change in response to changes in the sub-area network.	Use Generalized Benders decomposition to decompose traffic assignment problem into separate geographic areas. The focusing problem requires repeated solution of the Benders' master and the sub-area's problem.	a. Single mode planning and analysis: decomposition of nation by region or state; decomposition by operating rail company; analysis of inter-company competition. b. Multi-mode planning and analysis: decomposition by modes or authorities; analysis of multi-mode composition. c. Mass transit planning by sub-area focusing.

nodes and 3,000 links. In the Appendix is a brief discussion of the decomposition methods that were used in our analysis of the foregoing network problems.

1.2 TRAFFIC ASSIGNMENT PROBLEM -- FIXED DEMAND

The traffic assignment problem with fixed demands distribute given interzonal trip demands to alternate route between these zones. The demands are fixed and do not respond to changes in travel costs. The objective is to cbtain either a user equilibrium or a systems optimal assignment. This problem is not only interesting in itself, but, as we will show in Sections 3-6, it also provides the basis for our solution techniques for several other problems: traffic assignment with elastic demands, network design, investment staging, and subarea focusing. In Section 2, we discuss three decomposition methods for solving the traffic assignment problem with fixed demands. The first method is based upon the Frank-Wolfe algorithm [1]. The second method is based upon Zangwill's convex simplex method [2]. Both of these first two methods have been proposed and implemented by other authors for the solution of the traffic assignment problem. The third method, called Convex-Generalized Upper Bounding (Convex-GUB), is a new proposal, and it is based upon the generalized upper bounding technique of Dantzig and Van Slyke [3]. In Section 2, the strengths and weaknesses of these three decomposition methods are compared.

1.3 TRAFFIC ASSIGNMENT PROBLEM -- ELASTIC DEMAND

In the traffic assignment problem with elastic demands, the number of trips between a particular origin and destination is allowed to depend upon the cost of travel between that pair of zones (or nodes), and this relationship is specified by a demand function. The problem is to determine the user equilibrium traffic assignment subject to these elastic demand functions. Section 3 discusses three decomposition methods for solving this problem. All of these techniques solve the elastic demand problem by converting it into a fixed demand problem. The first method was originally introduced by Florian and Nguyen [4], and it uses Benders decomposition [5] to reduce the elastic demand problem into a fixed demand traffic assignment problem. Unfortunately, it also requires the solution of a master problem and restricts the solution of the fixed demand traffic assignment problem to a special class of algorithms. The second and third methods are new proposals in which we show how the Frank-Wolfe or convex simplex algorithms (both introduced in Section 2) can be used to solve the elastic demand problem, by solving a fixed demand traffic assignment problem over an expanded network.

1.4 NETWORK DESIGN

The network design problem is concerned with the addition or modification of links within a transportation infrastructure so that social costs of transportation are minimized subject to a budget constraint. The network could refer to either rail, highway, or mass transit applications. A general convex network design model is formulated in Section 4 that is related to models developed by Steenbrink [6, 7]; Morlok, Schofer, et al. at Northwestern University [8]; and Dafermos [9]. We will show that the basic decomposition procedure devised by Steenbrink can be used to obtain the global optimal solution to this general model by solving a series of traffic assignment problems. Our network design formulation has the following features:

a) Investment Decision Variables. The algorithm determines the optimal solution with respect to continuous investment decision variables.

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, although user equilibrium assignment would be preferred in highway applications. Section 4.1 gives bounds on the user equilibrium design problem using only the solution to the systems optimal design problem; if these bounds were close, then this would justify using the network design based on systems optimal assignment in an application in which user equilibrium assignment was preferred.

c) Travel Time as a Function of Flow and Investment. The model assumes that the total travel time on a link is a continuous convex function of the flow and investment decision for that link. This includes as special cases the non-linear differentiable curve, similar to the FHWA (the U.S. Federal Highway Administration) travel time function, used by Steenbrink [6, 7]; the piecewise linear curve used by the Northwestern group [8]; and the quadratic curve used by Dafermos [9].

d) Investment Cost Function. The model assumes that the cost for making an investment on a link is a continuous convex function of the investment decision.

e) Investment Alternatives. If the only effect of investment is to increase the capacity on existing links, then Section 4.2.1 shows how this could be implemented with a differentiable travel time curve; and if the effect of investment is to change either or both the free-flow travel time and capacity on existing links, then we show in Section 4.2.2 how this could be done with a piecewise linear travel time curve. The piecewise linear approach can also handle the introduction of entirely new links by specifying their initial capacity as being zero.

f) Solution Algorithm. For the case in which there is no budget constraint but the investment cost is included in the objective function, Section 4.3 shows how the solution to the network design problem could be obtained by solving a traffic assignment problem. For the case in which a budget constraint is used, then Section 4.4 gives a Lagrange multiplier technique that obtains a solution to the design problem by solving a series of traffic assignment problems, one for each value of the multiplier.

1.5 OPTIMAL STAGING OF INVESTMENTS OVER TIME

The network design model only treats a static problem: a network was to be examined for possible link additions or modifications, without regard to the sequence or timing of the implementation of these changes. However, it may be desirable to perform long-range planning for a highway, mass transit, or rail network in which investments are to be planned during each stage (or year) in a multi-stage horizon, subject to a budget constraint on the total investment in each stage. The staging problem is based on the observation that all network improvements would not contribute equally towards the efficient functioning of the system: some should be added early and some can be delayed. The analysis is complicated by the fact that the user benefits derived from improvements on different links are interrelated in a complex way through the network structure. We consider two variations of the staging problem in Section 5:

a) The final configuration is not specified. In this case the staging model would determine both the final configuration and the order in which the investments would occur.

b) The final configuration is specified. In this case the staging model would determine only the order in which the recommended network investments should be constructed over the horizon.

The objective function used in two previous formulations of the staging problem, Bergendahl [10] and Schimpeler-Corradino Associates [11], was to minimize the weighted sum of costs for each individual stage. The difficulty with this approach is that it becomes intractable for large networks. An alternative objective function can be specified by the application of the lexicographic ordering rule of vector analysis, and this can be viewed as being a generalization of a technique that was suggested by Roberts [12]. The method we propose in Section 5 uses an objective function based upon the lexicographic rule, continuous investment decision variables, and systems optimal traffic assignment. Using a decomposition procedure based upon dynamic programming, we show in Section 5 that a T-stage staging problem can be decomposed into T single-stage network design models if the final configuration is not specified, and that it can be decomposed into T-1 single-stage network design models if the final configuration is specified. As demonstrated in Section 4, the network design model with a budget constraint, continuous investment decision variables, and systems optimal assignment can be decomposed into a series of traffic assignment problems using the Lagrange multiplier technique. We are, therefore, able to decompose the staging model into a series of traffic assignment problems.

1.6 SOLVING TRAFFIC ASSIGNMENT AND SUB-AREA FOCUSING PROBLEMS BY GEOGRAPHIC DECOMPOSITION

Sections 2 - 5 discuss several large network problems in transportation: user equilibrium traffic assignment with fixed demands; systems optimal traffic assignment with fixed demands; user equilibrium traffic assignment with elastic demands; network design; and staging. Section 6 presents a new method, based upon Generalized Benders decomposition [5], which is able to provide a geographic separation of any of the foregoing network problems into smaller, more manageable subproblems. Geographic decomposition is based upon the observation that large transportation networks are often only loosely connected: if a small set of links are deleted, then the original network will decompose into a series of disjointed subnetworks. Although this methodology is equally applicable to any of these network problems, for the sake of expositional simplicity, we limit our discussion in Section 6 to the user equilibrium traffic assignment problem with fixed demands and to a new application, which Dial [13] has called sub-area focusing.

The sub-area focusing problem involves the extraction from a very large network of a sub-area, often called a window, so that detailed changes to the window can be examined. The so-called "sub-area windowing problem" makes the simplistic assumption that there will be no changes in flows or demands external to the window's network; however, the so-called "sub-area focusing problem" attempts to approximate changes in the external demands that could affect flows within the window. Therefore, sub-area focusing provides an estimate of how the whole network will respond to the changes within the window. As we show in Section 6, the analysis of a network using geographic decomposition involves the use of dual variables. The key observation is that the dual variables on the nodes external to the window can predict how demands will shift in response to changes within the window. This implies that geographic decomposition does provide an approach for efficiently performing sub-area focusing.

1.7 REFERENCES

1. M. Frank and P. Wolfe, "An Algorithm of Quadratic Programming," Naval Research Logistics Quarterly, Vol. 3, p. 95-110, 1956.

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

3. G. B. Dantzig and R. M. Van Slyke, "Generalized Upper Bounding Techniques," Journal of Computer System Science, Vol 1, p. 213-226, 1967.

4. M. Florian and S. Nguyen, "A Method for Computing Network Equilibrium with Elastic Demands," Transportation Science, Vol 8, No. 4, p. 321-332, 1974.

5. A. M. Geoffrion, "Generalized Benders Decomposition," J. Optimization Theory and Applications, Vol 10, No. 4, p. 237-260, 1972.

6. P. A. Steenbrink, "Transport Network Optimization in the Dutch Integral Transportation Study," Transportation Research, Vol 8, p. 11-27, 1974.

7. P. A. Steenbrink, Optimization of Transport Networks, Wiley, New York NY, 1974.

8. E. K. Morlok, J. L. Schofer, W. P. Pierskalla, R. E. Marsten, S. K. Agarwal, J. L. Edwards, L. J. LeBlanc, D. T. Spacek, and J. W. Stoner, 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, 1973, ch. 4.

9. S. C. Dafermos, Traffic Assignment and Resource Allocation in Transportation Networks, Ph.D. Thesis, The Johns Hopkins University, Baltimore MD, 1968, ch. 3.

10. G. Bergendahl, "A Combined Linear and Dynamic Programming Model for Interdependent Road Investment Planning," Transportation Research, Vol. 3, p. 211-228, 1969.

11. Schimpeler-Corradino Associates, Optimum Staging of Projects in a Highway Plan, prepared for Federal Highway Administration, Office of Highway Planning, Washington DC, 1974.

12. P. O. Roberts, Transport Planning: Models for Developing Countries, Ph.D. Thesis, Department of Civil Engineering, Northwestern University, Evanston IL, 1966, ch. 3.3.

13. R. B. Dial, "The Development of UMTA's Transportation Planning System," Presented at the 1974 Annual Meeting of the Highway Research Board, Washington DC, January 1974.

2. TRAFFIC ASSIGNMENT PROBLEM -- FIXED DEMANDS

2.1 INTRODUCTION

The traffice assignment problems distributes interzonal trip demands to alternate routes between these zones. In this section we consider only the fixed demand case in which the demands do not respond to changes in travel costs. The traffic assignment problem with fixed demands is not only an interesting problem in itself, but as we will show in Sections 3-6, this problem also arises in the analysis of other problems, such as traffic assignment with elastic demands, network design, investment staging, and sub-area focusing. In this section we will examine decomposition methods for the solution of the traffic assignment problem with fixed demands, and we will be concerned only with methods that have proven or in our opinion have the potential to be effective in the solution of realistically sized problems. Accordingly, we will only examine methods that can solve a traffic assignment problem which has at least 1,000 nodes and 3,000 links.

A great number of papers have appeared that propose both elegant and sophisticated solutions of the traffic assignment problem. Reviews of the various proposals can be found in Leblanc [1], Nguyen [2], Ruiter [3], Potts and Oliver [4], Petersen [5], and Murchland [6]. These proposals can be conveniently classified as either capacity constraint methods or equilibrium methods (See Nguyen [2]). Capacity constraint methods are characterized by heuristic procedures for determining the traffic flow patterns. Unfortunately, these methods can neither guarantee convergence to the optimal solution, nor in most cases, can they determine a bound on how far the obtained solution is from the optimal solution.

On the other hand, equilibrium methods guarantee convergence to the optimal solution, are iterative in nature, and usually generate a bound on the difference between their current solution value and the optimal solution. Here, however, we must be careful to distinguish between methods that have the potential to solve problems of realistic size and those that do not. For example, one of the latest proposals is the primal-dual procedure of Petersen [5], which was tested on a 72 node problem with 147 links and required 3 minutes of CDC 6600 time. Another test, by Leventhal, Nemhauser, and Trotter [7], using a column generation scheme on a problem with 64 nodes, but only 5 origin-destination pairs, required between 1 and 4 minutes on an IBM 360/65. Such tests give little information on how the algorithm will perform on problems of a 1,000 nodes or more. Therefore, the focus of this analysis will be on traffic assignment codes that use equilibrium methods, but yet still have the potential to solve realistically sized problems.

2.2 PROBLEM FORMULATION

The traffic assignment problem with fixed demands, can be formulated as:

$$\text{MINIMIZE} \quad Z = \sum_{j \epsilon A} T_j(f_j) \quad , \tag{1}$$

with respect to: f_j^r and f_j, $j\epsilon A$, $r=1, \ldots, R$,

subject to

$$\sum_{j \epsilon W_i} f_j^r - \sum_{j \epsilon V_i} f_j^r = h_i^r \quad , \qquad (i\epsilon N;\ r=1, \ldots, R) \quad , \tag{2}$$

$$f_j = \sum_{r=1}^{r=R} f_j^r \quad , \qquad (j\epsilon A) \quad , \tag{3}$$

$$f_j^r \geq 0 \ , \qquad (j \varepsilon A;\ r=1,\ \ldots,R) \ , \qquad (4)$$

where

A = the set of links in the network

f_j^r = flow on link j from origin r

$$h_i^r = \begin{cases} -O_{ij} & \text{if } i \text{ is a destination node} \\ \sum_{j \varepsilon A} O_{rj} & \text{if } i = r \\ 0 & \text{otherwise} \end{cases}$$

N = the set of nodes in the network

O_{ij} = the number of trips originating at origin i and terminating at destination j

R = the number of origin nodes in the network

$T_j(f_j)$ = the total travel cost function for link j, which depends only on the flow f_j on link j

V_i = the set of links terminating at node i

W_i = the set of links originating at node i.

The cost functions $T_j(f_j)$ can be given two different interpretations dependi upon whether the planner is interested in a system optimal or a user equilibrium solution. To differentiate between these two alternatives, we must define the marginal link cost function:

$C_j(f_j)$ = the link marginal (per unit) cost function, where f_j is the total flow on link j.

The function $C_j(f_j)$ is assumed to be a nonnegative, differentiable, and non-decreasing function of f_j. With this interpretation for $C_j(f_j)$, we have the following alternative specifications for T_j:

a) System Optimal

$$T_j(f_j) \equiv C_j(f_j) \cdot f_j ,$$

where the flow pattern found by solving (1) - (4) will minimize the total system travel cost.

b) User Equilibrium

$$T_j(f_j) \equiv \int_0^{f_j} C_j(x)\, dx ,$$

where the flow pattern found by solving (1) to (4) will have the property that no traveler can decrease his cost by changing paths, given that all others remain on their present paths. This flow pattern may be referred to as a Nash equilibrium point (See Dafermos and Sparrow [8]).

The equations (2) to (4) of the constraints for this problem are often referred to as the node-arc formulation of the convex cost multicommodity flow problem. An alternative formulation of the problem may be derived along the lines of Leventhal, Nemhauser, and Trotter [7] and is usually referred to as the arc-chain formulation of the problem. Since the algorithms we will consider in the latter part of this section can be adequately described in terms of the node-arc formulation, we will not burden the reader with this additional mathematical notation.

It is important at this point to draw attention to a special case of the formulation (1) to (4), where instead of using the differentiable and convex cost functions $T_j(f_j)$ we choose to use piecewise linear approximations to these functions. If we assume there are M_j breakpoints associated with the piecewise linear approximation, then equation (3) may be replaced by:

$$\sum_{m=1}^{m=M_j} x_j^m = \sum_{r=1}^{r=R} f_j^r \ , \qquad (j\varepsilon A) \ , \qquad (3'$$

and the upper bound constraints,

$$0 \leq x_j^m \leq K_j^m \ , \qquad (j\varepsilon A;\ m=1, \ldots, M_j) \ , \qquad (3''$$

where K_j^m are the breakpoints in the piecewise linear approximation. If we now define C_j^m as the slope of the appropriate linear segments in the approximation, we obtain the linear form of the objective function as:

$$\text{MINIMIZE} \quad Z = \sum_{j\varepsilon A} \sum_{m=1}^{m=M_j} C_j^m \ x_j^m \ . \qquad (1'$$

This formulation of our problem, which we will refer to as the linear cost multicommodity flow problem, has a long history in the literature of network flows. In fact, it was a proposed solution method by Ford and Fulkerson [9] for this problem that eventually led to the Dantzig and Wolfe [10] decomposition principle. Moreover, this particular linear form of our basic model is very important in the remainder of this report, because our models of the network design (Section 4) and investment staging (Section 5) problems utilize this formulation. Finally, the proposed Convex-GUB procedure (Section 2.3.3) for solvin the original convex formulation (1) to (4) of the traffic assignment problem is ba principles that were developed to solve the linear cost multicommodity flow problem. See, for example, the work of Maier [11] and Hartman and Lasdon [12] on this subject. With the mathematical formulations specified, we are now in a position to discuss solution methods that meet our twin criteria of being based on sound algorithmic principles and yet still be capable of solving realistically sized problems.

2.3 SOLUTION METHODS

It is not our intent in this report to review, critique, and computationally compare all the proposed solution methods for the traffic assignment problem that have appeared in the literature, although we believe this to be an important area for future work. We will, however, discuss two basic algorithmic approaches which have been previously developed and implemented by other investigators, as well as present a new proposal.

Our selection of these three methods is based upon the following three criteria. First, we examined published meaningful computational results. Meaningful to us meant that the problems solved were based on actual transportation data and of realistic size. Only one computational study qualified under this criteria, and it is reported by Florian and Nguyen in [13], [14], and [2].

Second, we were only interested in algorithms that would have the flexibility to solve our formulation of the network design problem (described in Section 4), the investment staging problem (described in Section 5), and the geographic decomposition method for sub-area focusing (described in Section 6). These three problems place special requirements on a traffic assignment algorithm. The network design and investment staging problems generate objective functions that, although convex, are generally not differentiable. The traffic assignment algorithm will, therefore, require a derivative approximation technique. Since convergence can now become uncertain, an objective function bound on the algorithm will be most critical. The geographic decomposition approach, on the other hand, generates artificial origins and destinations. These artificial origins and destinations show up as modifications in the right-hand-side coefficients h_i^r.

This change in h_i^r can not be accomodated by many traffic assignment solution methods.

The final criterion in the selection of our algorithmic approaches can best be described as judgmental, and is based upon the experience of the research team with other large-scale mathematical programming problems. In this regard, the availability of the knowledge and resources collected by the Systems Optimization Laboratory at Stanford University is most important. The traffic assignment problem possesses much structure in its mathematical formulation. We have already pointed out that this problem can be described as a convex cost multicommodity flow problem. Furthermore, the constraints, as we have formulated them, possess the block-angular structure on which a great deal of research has been done. Therefore drawing upon the experience of the research team, taking account of available computational studies, and mindful of the special needs of our other problems, we have selected the following approaches as most meeting the needs of this study.

2.3.1 FRANK-WOLFE APPROACH

One of the earliest proposed approaches to the solution of a mathematical programming problem in which the constraints were linear, but the objective function was convex and differentiable, is the method of Frank and Wolfe [15]. A simple description of the method is as follows: the procedure starts with a feasible solution to the problem. Then a linear approximation to the objective function at the current solution value is generated. Using this linear approximation to the objective function, a linear program is formulated and solved. The solution to the linear program yields another feasible solution to the original problem, which is usually called the trial solution. The final step in the process consists of moving from the starting feasible solution in the direction of the trial solution and stopping when the maximum decrease in the objective function value is achieved.

This basic approach can be specialized to the traffic assignment problem by analyzing the generation of the trial solution. Nguyen [2] has shown that for this particular problem the generation of the trial solution does not require a linear program; rather, the trial solution is easily formed by assigning all traffic to the least cost path between each origin-destination pair. The computation of the least cost path can be found very efficiently by a shortest path algorithm, such as that proposed by Dial [16]. Once the trial solution is known, a simple one-dimensional search procedure, such as the Golden-Section method [17], may be used to generate the new feasible solution.

This procedure has a number of important attributes. First, it is extremely easy to program and requires a minimal amount of computer core storage. It is not surprising, therefore, to find that it is the basic algorithmic procedure used by the UTPS program UROAD. Second, the procedure produces an easily computed bound on the optimal objective function value. Third, some rather simple modifications to the procedure can be incorporated to allow for nondifferentiable cost functions, as described by Ruiter [3]. Finally, we have some published computational experience with the procedure. In analyzing the transportation system for the City of Winnepeg with 1,035 nodes, 2,789 links, and 140 origins, Florian and Nguyen [13] found that the Frank-Wolfe algorithm required 15 to 18 trial solutions, which in turn required 700 CPU seconds on a CDC Cyber 74 at a cost of $315. Although no published results were made available to us by UMTA or FHA, discussions with their key personnel indicated similar experience. UMTA also indicated that they had attempted problems of much larger dimensions; however, they did terminate the procedure after fewer iterations, and it was not clear whether a satisfactory solution had been obtained.

At this point we believe it important to voice one major concern about this procedure. The Frank-Wolfe algorithm is no longer widely used as a method to solve mathematical programs with convex costs and linear constraints. The computational experience gathered over time has shown that this procedure has one very undesirable characteristic: it converges very slowly. Unfortunately, Nguyen [2] has observed that this phenomenon appears to be present in the solution of traffic assignment problems as well. In his study of the City of Winnipeg he observed a significant "tailing off phenomenon in the vicinity of the equilibrium [solution]". We believe that in light of this history of convergence problems a careful examination of the computational characteristics of the Frank-Wolfe algorithm should be undertaken.

2.3.2 CONVEX SIMPLEX METHOD

Zangwill [18] has proposed a modification of the linear programming simplex method in the case where the objective function is convex and differentiable instead of linear. This procedure was specialized by Nguyen [14] to the traffic assignment problem and proposed as an alternative to the Frank-Wolfe method. In contrast to the simplicity of the Frank-Wolfe approach, the convex simplex method (while conceptually similar) is a somewhat more complex procedure. It is initiated by first applying to our original formulation (1) to (4 a natural decomposition proposed by Murchland [6] of flows by point of origin (or possibly by final destination). The result of this decomposition is a subproblem P(r) defined as:

$$\text{MINIMIZE} \quad Z_r = \sum_{j\varepsilon A} T_j(f_j) \quad , \qquad (5$$

with respect to: f_j^r and f_j, $j\varepsilon A$.

subject to:

$$\sum_{j \varepsilon W_i} f_j^r - \sum_{j \varepsilon V_i} f_j^r = h_i^r \qquad (i \varepsilon N) \ , \tag{6}$$

$$f_j = f_j^r + H_j^r \ , \quad (j \varepsilon A) \ , \tag{7}$$

$$f_j^r \geq 0 \ , \qquad (j \varepsilon A) \ , \tag{8}$$

where H_j^r is a constant equal to $\sum_{i \neq r} f_j^i$.

The subproblem P(r) is embedded in the following iterative decomposition scheme:

Step 1

Find an initial feasible flow pattern

f_j^r , $j \varepsilon A$, r=1, ...,R.

Set r=1.

Step 2

Solve the subproblem P(r).

Step 3

Revise the current flow pattern f_j^r based on the last solution of P(r).

If the current flow pattern is simultaneously optimal for all subproblems P(r), r=1, ...,R, terminate.

If r=R, set r=1; otherwise set r=r+1.

Return to Step 2.

The great appeal of this decomposition scheme is that it has dramatically reduced the size of the original multicommodity network flow problem (1) to (4) which one must work with at any point in the procedure. For the traffic assignment problems for the City of Winnipeg, which had 1,035 nodes, 2,789 links, and 140 origins, the reduction is from 145,000 equations and 393,000 variables in the original formulation (1) to (4) to only 1,000 constraints and 2,800 variables in subproblem P(r). Moreover, not only is the problem P(r) much smaller in size, but it is also a type of mathematical program that is easy to solve. The careful reader will no doubt recognize that this problem is the classical tra shipment problem with convex costs. At this point, a number of avenues would appear to be open with respect to the solution of the problem P(r).

Nguyen [4] in his analysis of the problem used a special adaptation of the convex simplex method. In his procedure he divised some rather elaborate techniques for maintaining a directed spanning tree upon which to make his flow adjustments. [Note: A directed spanning tree can be associated with an extreme point solution, or in linear programming parlance, a basis.] The algorithm itsel has many similarities to the work of Glover, Karney, and Klingman [19] on the transshipment problem with linear costs. There has also been some more recent work on the transshipment problem by Gordon Bradley, Gerald Brown, and Glenn Grav [20]. In personal discussions with the authors, this latter group has indicated they have devised a code that is a significant improvement over the work of Glove Karney, and Klingman. One thing is certain: great care must be taken in constructing a code to solve the problem P(r) if one is to capitalize on all of the previous work by other researchers. In our research on Convex Generalized Upper Bounding (to be discussed in Section 2.3.3), we have developed our own variant o Nguyen's approach which we believe has much promise.

Florian and Nguyen [13] have performed some computational experiments with their version of the convex simplex method. They again attempted to solve the City of Winnipeg problem. One difficulty arose because of the following compact representation of an intersection:

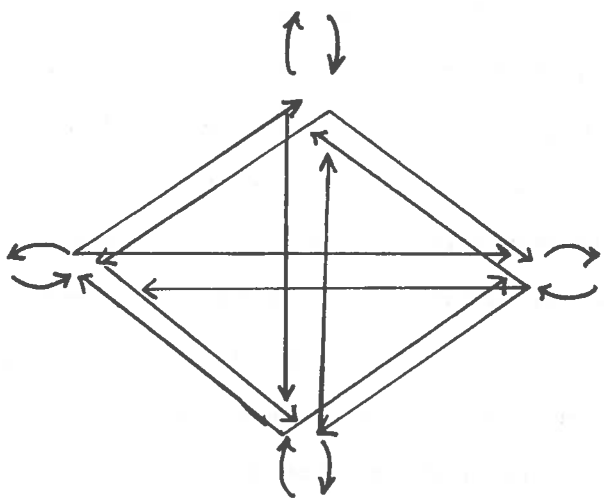


FIGURE 2.1--City of Winnipeg -- Original Problem.

It was found that for the convex simplex method solution, it was necessary to restructure such intersections as follows:

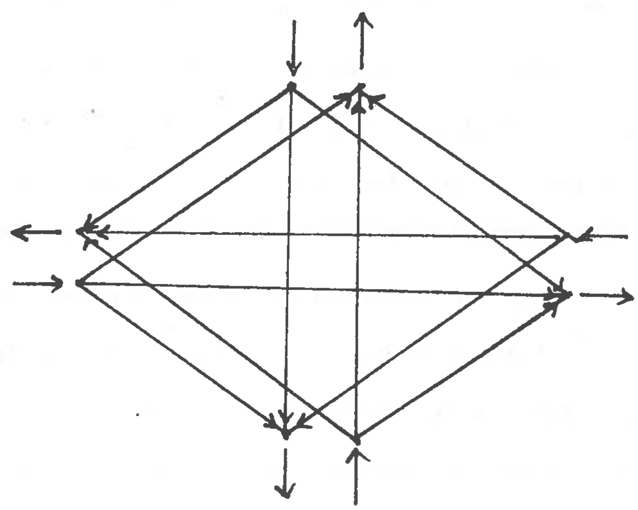


FIGURE 2.2--City of Winnipeg -- Restructured Solution.

This in turn increased the number of nodes in the system by 217. However, even with this larger network, the convex simplex method found a solution in about 500 CPU seconds on the CDC Cyber 74. This is somewhat less time than was consumed by the Frank-Wolfe type algorithm on the same problem. Besides actual running times, Nguyen [2] presents other data that show that the convex simplex method does not appear to suffer from the convergence problems of the Frank-Wolfe method. [It is possible to construct theoretical examples where the convergence of both methods are painfully slow.] Nguyen also points out that because the convex simplex method relies more heavily or auxiliary computer storage, the actual cost for this approach, $297, is very nearly the same as the Frank-Wolfe method cost of $315. We can summarize the computational results obtained by Florian and Nguyen as supporting the hypothesis that these two methods are at least comparabl

One final comment on computational results. The heart of the Frank Wolfe approach is an efficient shortest path algorithm. Our group has obtained from Nguyen a copy of his Frank-Wolfe code and verified that he used a very efficient version of this shortest path algorithm. On the otherhand, in the convex simplex method the key to efficiency rests in the storage and updatin of the spanning trees. Unfortunately, we were unable to obtain from Nguyen a copy of his convex simplex code. However, based on our own analysis of his published work, we believe that he may not have used the most efficient means of storing and updating the spanning trees. We believe, therefore, that it is very likely that a more efficient version of the convex simplex approach could be coded using the principles developed by Glover, Karney, and Klingman or Bradle Brown, and Graves mentioned previously or one based on the features contained in our Convex GUB algorithm, to be discussed in Section 2.3.3.

In order to complete our analysis of the convex simplex method, we will now comment on how it conforms to the requirements of our network design, investment staging, and geographic decomposition models. Perhaps of greatest importan is that the convex simplex method would permit the necessary modifications in the traffic assignment formulation to allow solution of our geographic decomposition problem. Nguyen in [14] indicates that a bound on the value of the objective function can be easily calculated for the convex simplex method. With such a bound, it would then be possible to easily modify the procedure to yield approximations to the derivative for the cost functions, similar to what has been done in the case of the UROAD program. A derivative approximation routine would then permit us to solve both the network design and staging problems by this method.

2.3.3 CONVEX GENERALIZED UPPER BOUNDING

The discussions of the foregoing two approaches for solving the traffic assignment problem should have brought out the relationship between this problem and the linear cost multicommodity flow problem. In fact, the lack of a differentiable cost function for the network design and investment staging problems so modifies the resulting models that it is perhaps better to think of them as being multicommodity flow problems with piecewise linear costs. It is in the examination of the literature concerned with the multicommodity flow problems that an interesting new idea was conceived. The idea is based on the Dantzig and Van Slyke algorithm called Generalized Upper Bounding [21]. Because we are applying this technique to a problem with a convex objective function, we have called the new procedure Convex Generalized Upper Bounding, or simply Convex-GUB.

Generalized Upper Bounding has been applied to the linear cost multicommodity flow problem by Maier in [11] and experimental results have been reported by Grigoriadis and White [22] and Hartman and Lasdon [12]. Furthermore, Carlos Winkler [23], in a study performed at the Stanford University System Optimization Laboratory, has shown that this procedure appears to run faster than the classical Dantzig - Wolfe decomposition principle [10] for general linear programs that possess the block angular structure of the multicommodity network flow problem.

The Convex-GUB algorithm can best be thought of as a modification of the convex simplex method discussed in the last section. In fact, any implementation of this procedure would first be preceded by developing an efficient convex simplex code. The principle behind the Convex-GUB method is to increase the interaction between the subproblems P(r) defined by applying the Murchland natural decomposition by sources. Currently, the only interaction that takes place in the Murchland procedure is through changes in the constants H_j^r, which are defined by the flow levels on all links except those in subproblem P(r). In the Murchland decomposition scheme, all flow changes must take place in only one subprogram P(r) at a time. In the Convex-GUB procedure, a special "working basis" is constructed that permits flow inter-changes to take place between several subproblems P(r) simultaneously. With the working basis, it is then possible to examine extremely congested links in a current solution and relieve this congestion by a reduction in traffic flow in a number of the subproblems P(r) simultaneously.

It is possible to construct hypothetical networks where the Frank-Wolfe or convex simplex approaches would converge slowly (or even terminate before reaching an optimum due to cut-offs built into the algorithm). We believe it is worthwhile testing: (1) Whether these approaches are in fact achieving a near optimum solution; and (2) whether convergence rates can be speeded up by use of the Convex GUB procedure. If the convergence rates are good, probably a formula that bounds the true minimum from below will suffice to test (1). An experimental version of the Convex-GUB procedure has been written in the MPL language to test (2). [Note: MPL is a mathematical programming language developed under a grant from the National Science Foundation to Stanford University.]

2.4 REFERENCES

1. L. J. LeBlanc, Mathematical Programming Algorithms for Large Scale Network Equilibrium and Network Design Problems, Ph.D. Dissertation, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston IL, 1973, ch. 1.

2. S. Nguyen, "A Unified Approach to Equilibrium Methods for Traffic Assignment," presented at the International Symposium on Traffic Equilibrium Methods, Montreal, CANADA, November 1974.

3. E. R. Ruiter, "Implementation of Operational Network Equilibrium Proceedures," Technical Memorancum CS.UE.05, Cambridge Systematics, Inc., Cambridge MA, January 1974.

4. R. B. Potts, and R. M. Oliver, Flows in Transportation Networks, Academic Press, New York NY, 1972.

5. E. R. Petersen, "A Primal-Dual Traffic Assignment Algorithm," Management Science, Vol. 22, No. 1, p. 85-87, 1975.

6. J. D. Murchland, "Road Network Traffic Distribution in Equilibrium," presented at the conference "Mathematical Methods in the Economic Sciences," Mathematisches Forschungsingstitut, Oberwolfach, 1969.

7. T. L. Leventhal, G. L. Nemhauser, and L. E. Trotter, "A Column Generation Algorithm for Optimal Traffic Assignment," Transportation Science, Vol 7, p. 168-176, 1973.

8. S. C. Dafermos and F. T. Sparrow, "The Traffic Assignment Problem for a General Network," Journal of Research, National Bureau of Standards, Vol 73B p. 91-118, 1969.

9. L. R. Ford and D. Fulerson, "A Suggested Computation for Maximal Multicommodity Network Flows," Management Science, Vol 5, No. 1, p. 97-101, 1958.

10. G. B. Dantzig and P. Wolfe, "Decomposition Principle for Linear Programs," Operations Research, Vol 8, p. 101-111, 1960.

11. S. F. Maier, "A Compact Inverse Scheme Applied to the Multicommodity Network with Resource Constraints," Optimization Methods for Resource Allocation, Crane, Russak and Co., New York NY, 1974, p. 179-204.

12. J. Hartman and L. Lasdon, "A Generalized Upper Bounding Algorithm for Multicommodity Network Flow Problems," Networks, Vol 1, p. 333-354, 1972.

13. M. Florian and S. Nguyen, "On Application and Validation of Equilibrium Trip Assignment Methods," Publication #28, Centre de recherche sur les transports, University of Montreal, CANADA, August 1975.

14. S. Nguyen, "An Algorithm for the Traffic Assignment Problem," Transportation Science, Vol 8, p. 203-216, 1974.

15. M. Frank and P. Wolfe, "An Algorithm of Quadratic Programming," Naval Research Logistics Quarterly, Vol 3, p. 95-110, 1956.

16. R. B. Dial, "Shortest-Path Forest with Topological Ordering," Communications of the ACM, Vol. 12, No. 11, p. 632-633, November 1969.

17. J. Kowalik and M. R. Osborne, Methods for Unconstrained Optimization Problems, Elsevier, New York NY, 1968.

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

19. F. Glover, D. Karney, and D. Klingman, "Implementation and Computational Study on Start Procedures and Basic Change Criteria for a Primal Network Code," Networks, Vol 4, No. 3, p. 191-212, 1974.

20. G. H. Bradley, G. G. Brown, and G. W. Graves, "A Comparison of Some Structure for Primal Network Codes," Naval Postgraduate School Technical Report (In Preparation), Monterey CA.

21. G. B. Dantzig and R. M. Van Slyke, "Generalized Upper Bounding Techniques," Journal of Computer System Science, Vol 1, p. 213-226, 1967.

22. M. D. Grigoriadis and W. W. White, "Computational Experience with a Multi-Commodity Network Flow Algorithm," Optimization Methods for Resources Allocation, Crane, Russak and Co., New York NY, 1974, p. 205-226.

23. C. Winkler, "Basic Factorization for Block-Angular Linear Programs: Unified Theory of Partitioning and Decomposition Using the Simplex Method," Ph.D. Dissertation, Operations Research Department, Stanford University, Stanford CA, December 1974.

3. TRAFFIC ASSIGNMENT PROBLEM -- ELASTIC DEMANDS

3.1 INTRODUCTION

The traffic assignment model formulated in Section 2 assumed a known and constant origin-destination trip matrix. In this section we extend this model to include the possibility of elastic demands for service; namely, we consider the case where the number of trips between origin i and destination j, denoted O_{ij}, is a function of the travel time between these points. Thus our problem is to determine the user equilibrium traffic assignment subject to these elastic demand functions. The systems optimal objective function does not apply to the elastic demand case.

We will present two formulations of this model. The first, which we call the circulation formulation, appears to have been first proposed by Murchland [1]. Based on this formulation, Florian and Nguyen [2] have proposed a solution methodology which uses Geoffrion's generalized Benders decomposition [3]. Florian and Nguyen's approach has considerable appeal, since these authors reported computational times that indicate the elastic demand problem can be solved with only 20 to 25 percent more effort than the fixed demand problem requires for the same network. There are, however, two drawbacks to the Florian and Nguyen approach. First, the approach requires the solution of a special subproblem corresponding to the generalized Benders decomposition master problem. Although Florian and Nguyen show this subproblem to be relatively easy to solve, it does require a considerably more complex computer program than would be required to solve the fixed demand traffic assignment problem. Second, after the decomposition is applied, the problem reduces to the solution of a fixed demand traffic assignment problem; but, unfortunately, this fixed demand problem requires a special algorithm for solution, because this decomposition technique only allows the flows between one origin-destination pair to be modified on any iteration. Because of this special requirement placed on the fixed demand traffic assignment algorithm, it does not appear as though either the present UROAD program or the three algorithms presented in Section 2 could be used to solve the elastic demand traffic assignment problem using the Florian and Nguyen approach.

In the second formulation that we will discuss, some additional nodes and links are added to obtain an expanded network model, which is nevertheless equivalent to the circulation formulation presented first. This second formulation, which we call the network formulation, has one major advantage over the circulation formulation: namely, it permits us to use any standard fixed demand traffic assignment algorithm to solve it. This means that either the current UROAD program or programs based on the approaches presented in Section 2 could be used to solve the elastic demand traffic assignment problem, and furthermore this solution could be obtained without making any modification to the fixed demand algorithm.

There is one disadvantage to this second formulation: we must solve a slightly larger fixed demand traffic assignment problem than that solved in the Florian and Nguyen approach. The network formulation requires 1 additional node and 1 additional link for each commodity. Since the basic node-arc approach uses R commodities (corresponding to the number of origin nodes in the network), this means a total of R additional nodes and R links. The fact that the network formulation has a larger number of nodes and links can be somewhat misleading. As we shall discuss below, both the Frank-Wolfe and convex simplex methods (outlined in Section 2) can be modified slightly so that the computational burden of the additional nodes and links is insignificant.

3.2 PROBLEM FORMULATIONS

The traffic assignment problem with elastic demands can be formulated as:

$$\text{MINIMIZE} \quad \sum_{j\varepsilon A} \int_0^{f_j} C_j(x)dx - \sum_{i=1}^{i=R} \sum_{j=1}^{j=R} \int_0^{O_{ij}} D_{ij}(x)dx \ , \tag{1}$$

with respect to: f_j^r, f_j, and O_{ir}, $j\varepsilon A$, $r=1, \ldots, R$, $i=1, \ldots, R$,

subject to:

$$\sum_{j\varepsilon W_i} f_j^r - \sum_{j\varepsilon V_i} f_j^r + \delta_i^r \ O_{ri} = 0 \ , \tag{2}$$

$$(r=1,\ldots,R;\ i\varepsilon N,\ i\neq r)$$

$$f_j = \sum_{r=1}^{r=R} f_j^r \ , \qquad (j\varepsilon A) \ , \tag{3}$$

$$f_j^r \geq 0 \ , \qquad (j\varepsilon A;\ r=1,\ldots,R) \ , \tag{4}$$

$$O_{ij} \geq 0 \ , \qquad (i=1,\ldots,R;\ j=1,\ldots,R) \ , \tag{5}$$

where

- A = the set of links in the network,
- $C_j(f_j)$ = the link marginal (per unit) cost function, where f_j is the total flow on link j,
- $D_{ij}(O_{ij})$ = the inverse of the demand function for service from origin i to destination j [Note: Following Wigan [4], if we let $g_{ij}(u_{ij})$ be the number of trips generated between origin i and destination j, which is a function of the cost of travel u_{ij} between these points, then $D_{ij}(O_{ij})$ is the inverse of this demand function.],

f_j^r = the traffic flow on link j from origin r,

N = the set of nodes in the network,

O_{ij} = the traffic flow generated by the demand functions from origin i to destination j,

R = the number of origin nodes in the network,

W_i = the set of links originating at node i,

V_i = the set of links terminating at node i,

$$\delta_i^r = \begin{cases} 1 \text{ if node } i \text{ is a destination node associated with node } r \text{ in the origin - destination matrix,} \\ 0 \quad \text{otherwise.} \end{cases}$$

An equivalent formulation of this problem may be stated as:

$$\text{MINIMIZE} \quad \sum_{j \varepsilon A} \int_0^{f_j} C_j(x)dx \quad - \sum_{i=1}^{i=R} \sum_{j=1}^{j=R} \int_0^{O_{ij}} D_{ij}(x)dx \ , \tag{6}$$

with respect to: f_j^r, f_j, s_r, O_{ri}, $j\varepsilon A$, $r=1, \ldots, R$, $i=1, \ldots, R$,

subject to:

$$\sum_{j\varepsilon W_i} f_j^r - \sum_{j\varepsilon V_i} f_j^r + \delta_i^r \quad O_{ri} = 0 \ , \tag{7}$$

$$(r=1,\ldots,R;\ i\varepsilon N,\ i\neq r) \ ,$$

$$\sum_{\substack{i\varepsilon N \\ i\neq r}} \delta_i^r \quad O_{ri} - s_r = -H_r \ , \qquad (r=1,\ldots,R) \ , \tag{8}$$

$$\sum_{j\varepsilon W_r} f_j^r - \sum_{j\varepsilon V_r} f_j^r + s_r = H_r \ , \quad (r=1,\ldots,R) \ , \tag{9}$$

$$f_j = \sum_{r=1}^{r=R} f_j^r \qquad (j\varepsilon A) \ , \tag{10}$$

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

$$O_{ij} \geq 0 \ , \qquad (i=1,\ldots,R;\ j=1,\ldots,R) \ , \qquad (12)$$

$$s_r \geq 0 \ , \qquad (r=1,\ldots,R) \ , \qquad (13)$$

where

s_r = a slack flow variable associated with commodity r

H_r = a large constant [Note: The value of H_r should be at least as large as the maximum flow anticipated between origin r and all associated destinations.] .

Equations (1) to (5) will be referred to as the circulation formulation and equations (6) to (13) as the network formulation of the elastic demand traffic assignment problem. Equation (9) in the second formulation can be omitted if desired, since it is a redundant equation.

3.3 SOLUTION METHODS

As mentioned in the introduction, the circulation formulation of the elastic demand model can be solved efficiently by the method proposed by Florian and Nguyen [2]. The disadvantages of the Florian and Nguyen approach are that it requires a special algorithm to solve the Benders master problem and another special algorithm to solve the resulting fixed demand traffic assignment problem.

The network formulation (6) to (13) will now be examined. In deriving this formulation we used a standard device from network flow theory, although to the best of our knowledge, this formulation of the problem has not previously been investigated in the traffic assignment literature. It is, however, a very useful formulation of the problem. If the reader groups all the flow variables $(f_j^r,\ O_{ri},\ s_r)$ together and treats them as though they were each associated with a corresponding link in the network, and if he appends R additional nodes to the network corresponding to equation (8), then the network formulation of the elastic demand traffic assignment problem can be seen to be exactly equivalent to a fixed demand traffic assignment problem. The resulting fixed demand traffic assignment problem has N+R nodes and $R \cdot |A| + R^2 + R$ links, where $|A|$ is the number of links in the set A.

It is interesting to compare this fixed demand traffic assignment problem to the fixed demand traffic assignment problem that is a result of the Florian and Nguyen procedure. The Florian and Nguyen procedure produces a fixed demand traffic assignment problem with only N nodes and $R \cdot |A|$ links. This substantial reduction in the number of nodes and links is of course gained at the expense of having to solve a great number of Benders master problems. We will now discuss how the Frank-Wolfe and convex simplex procedures (both presented in Section 2) can be easily modified so that this apparent advantage of the Florian and Nguyen approach almost disappears.

3.3.1 FRANK-WOLFE APPROACH

In what follows, we assume the reader has read Section 2.3.1 which describes the steps of the Frank-Wolfe approach. The critical step in the Frank-Wolfe algorithm is the calculation of the new trial solution. In the traffic assignment problem this is accomplished by the computation of the shortest path between each origin-destination pair. Using either the Dijsktra algorithm [5] or Dial's version of the Moore algorithm [6], we select a particular source node r and then find the shortest path from this node to all associated destination nodes. This same procedure is then repeated for all other origin nodes. For the elastic demand traffic assignment problem, the computation of the shortest path from node r to all associated destination nodes can be modified for computational efficiency.

We begin by assuming that the maximum flow from origin r to destination j is given as H_r^j, where we have defined

$$H^r \equiv \sum_{j=1}^{j=R} H_r^j ,$$

in the network formulation (6) to (13). Then let $COST_r^j$ be the minimum cost path from origin r to destination j found by applying the shortest path algorithm to the network composed of only the nodes in the set N and the links in the set A. We then compute the trial solution associated with origin r as follows:

Set $s_r = 0$.

For $j = 1,\ldots,R$ do:

If $COST_r^j < D_{rj}\ (O_{rj})$, then place H_r^j units of flow on the links associated with the minimum cost path from r to j. Set $O_{rj} = H_r^j$.

If $COST_r^j \geq D_{rj}(O_{rj})$, set $O_{rj} = 0$, and $s_r = s_r + H_r^j$ (No flow is placed on any links in the minimum cost path from r to j.)

By repeating this process for each origin node r in turn, we will generate a complete trial solution. Notice, however, that the shortest path algorithm is only applied to the network (N,A), which is exactly the same size network that Florian and Nguyen use in their procedure. We believe that this modified version of the Frank-Wolfe algorithm should be as efficient as the Florian and Nguyen procedure, although this is an open computational question.

3.3.2 CONVEX SIMPLEX METHOD

The important observation for the convex simplex method is that the subproblem P(r), defined in Section 2.3.2 by equations (5) - (8), does not involve all of N+R nodes and $R \cdot |A| + R^2 + R$ links of the network formulation (6) to (13). (We assume again that the reader is familiar with the convex simplex method described in Section 2.3.2.) In fact, the subproblem P(r) has only N+1 nodes and $|A| + R + 1$ links. It can be formulated as follows:

$$\text{MINIMIZE} \quad Z_r = \sum_{j \varepsilon A} \int_o^{f_j} C_j(x)dx - \sum_{j=1}^{j=R} \int_o^{O_{rj}} D_{rj}(x)dx ,$$

with respect to: f_j^r, f_j, s_r, and O_{ri}, $j\varepsilon A$, $i=1, \ldots, R$,

subject to:

$$\sum_{j \varepsilon W_i} f_j^r - \sum_{j \varepsilon V_i} f_j^r + \delta_i^r \; O_{ri} = 0 ,$$

$$(i\varepsilon N, \; i \neq r)$$

$$- \sum_{\substack{i\varepsilon N \\ i \neq r}} \delta_i^r \; O_{ri} - s_r = - H_r ,$$

$$\sum_{j \varepsilon W_r} f_j^r - \sum_{j \varepsilon V_r} f_j^r + s_r = H_r ,$$

$$f_j = f_j^r + K_j^r , \quad (j\varepsilon A) ,$$

$$f_j^r \geq 0 , \quad (j\varepsilon A) ,$$

$$O_{rj} \geq 0 , \quad (j=1, \ldots, R) ,$$

$$s_r \geq 0 ,$$

where K_j^r is a constant equal to $\sum_{i \neq r} f_j^i$.

Although the subproblem P(r) above is slightly larger than the one associated with the fixed demand traffic assignment case (having 1 additional node and R + 1 additional links), the authors believe this problem can be solved almost as efficiently as the fixed demand case. The reason for our optimism is that the simplex method and its variants appear to be most sensitive to the number of equations (which has only grown by 1) and much less sensitive to the number of variables (which has grown by R + 1). In any case, this is another computational question that needs further examination.

3.4 CONCLUSION

In this section, we have described three different computational approaches to the elastic demand traffic assignment problem. The first, by Florian and Nguyen [2], uses Generalized Benders decomposition to reduce the elastic demand problem to a fixed demand traffic assignment problem. Unfortunately, it also requires the solution of a master problem and restricts the solution of the fixed demand traffic assignment problem to a special class of algorithms. However, computationally the procedure has some appeal, since it obtains solutions in only 20 to 25 percent more time than the fixed demand case.

The second approach examined is the Frank-Wolfe algorithm originally described in Section 2.3.1. This algorithm can be applied without modification to an expanded network; but with only minor modification to the procedure for computation of trial solutions, the shortest path portion of the Frank-Wolfe algorithm can be applied to the same size network as used by Florian and Nguyen.

The final approach examined was the convex simplex algorithm described in Section 2.3.2. This algorithm also needs no modification, but does involve solving a somewhat larger equivalent fixed demand traffic assignment problem than solved by either of the two previous approaches.

The computational question as to which of these three procedures is most efficient still remains open, but the effort required to implement either of the last two methods is far less than that required to implement the first.

3.5 REFERENCES

1. J. D. Murchland, "Road Network Traffic Distribution in Equilibrium," presented at the conference "Mathematical Methods in the Economic Sciences," Mathematisches Forschungsinstitut, Oberwolfach, 1969.

2. M. Florian and S. Nguyen, "A Method for Computing Network Equilibrium with Elastic Demands," Transportation Science, Vol, 8, p. 321-332, 1974.

3. A. M. Geoffrion, "Generalized Benders Decomposition," Journal of Optimization Theory and Application, Vol 10, No. 4, p. 237-260, 1972.

4. M. R. Wigan, "Benefit Assessment for Network Traffic Models and Application to Road Pricing," Road Research Laboratories Report LT 417, Crowthorne, Berkshire ENGLAND, 1971.

5. E. W. Dijsktra, "Note on Two Problems in Connection with Graphs," Numerische Mathematik, Vol 1, p. 269-271, 1959.

6. R. D. Dial, "Shortest-Path Forest with Topological Ordering," Communications of the ACM, Vol. 12, No. 11, p. 632-633, November 1969.

4. NETWORK DESIGN MODEL

4.1 INTRODUCTION

The network design problem is concerned with the addition or modification of links within a transportation infrastructure so that the social costs of transportation are minimized. The network could refer to either rail, highway, or mass transit applications. Network design models can be classified in the following ways: whether the investment decision variables are discrete or continuous; whether the traffic assignment is user equilibrium (first principle of Wardrop) or systems optimal (second principle of Wardrop); and whether or not the possibility of congestion is included into the formulation. Table 4.1 displays some of the previous network design models and classifies these models according to the foregoing factors [1 through 11].

It is probably desirable to use user equilibrium assignment when designing highway networks and systems optimal assignment when designing rail or mass transit networks. Refer to Section 2.2 for the definitions of these two types of traffic assignment. In the case of user equilibrium assignment, the network design problem becomes: determine the optimal network design to minimize total users' travel time (plus possibly the investment costs) subject to user equilibrium traffic assignment and a possible budget constraint on total investment. This formulation is very difficult to solve for the following reason: the objective function appropriate for the network design problem (minimizing total travel time) is different from the objective function appropriate for user equilibrium traffic assignment (minimizing the integral of the travel time function). Because of this mathematical difficulty, user equilibrium network design formulations, such as LeBlanc [4], can only handle small network problems. In the case of systems optimal assignment, the network design problem becomes: determine the optimal network design to minimize total users' travel time

TABLE 4.1

CLASSIFICATION OF TRANSPORT NETWORK DESIGN MODELS

Author	Year	Investment Variable	Traffic Assignment	Total Link Travel Time as Function of Flow	Solution Algorithm	Objective
Agarwal [1]	1973	Discrete	Systems Optimal	Piecewise Linear	Mixed integer linear programming	Minimize total travel costs subject to a budget constraint
Carter and Stowers [2]	1963	Continuous	Systems Optimal	Piecewise Linear	Dantzig-Wolfe Decomposition	Minimize total travel costs plus investment costs subject to budget constraint
Dafermos [3]	1968	Continuous	Systems Optimal	Quadratic	Gradient	Minimize total travel costs subject to a budget constraint
LeBlanc [4]	1973	Discrete	User Equilibrium	Piecewise Linear	Branch and Bound	Minimize total travel costs subject to a budget constraint
Northwestern [5] Agarwal [6] LeBlanc [7]	1973	Continuous	Systems Optimal	Piecewise Linear	Several methods including simplex, Dantzig - Wolfe Decomposition and Boxstep	Minimize total travel costs subject to a budget constraint
Ridley [8]	1968	Discrete	No Difference	Linear	Branch and Bound	Minimize total travel costs subject to a budget constraint
Roberts [9]	1966	Discrete	Systems Optimal	Linear, with an upper bound on maximum flow	Mixed integer linear programming	Minimize total travel costs plus investment costs subject to a budget constraint
Steenbrink [10, 11]	1974	Continuous	Systems Optimal	Non-Linear differentiable	Constructs local optimal solution using link decomposition followed by solution of a traffic assignment problem	Minimize total travel costs plus investment costs
Proposed Model, No. 1	1976	Continuous	Systems Optimal	Convex	Constructs global optimal solution using link decomposition followed by solution of a traffic assignment problem.	Minimize total travel costs plus investment costs
Proposed Model, No. 2	1976	Continuous	Systems Optimal	Convex	This is same as above, except uses Lagrange multiplier to handle budget constraint	Minimize total travel costs subject to a budget constraint

(plus possibly the investment costs) subject to systems optimal traffic assignment and a possible budget constraint on total investment. Most of the network design models listed in Table 4.1 use systems optimal assignment, because the objective function is the same for both the network design and traffic assignment problems, and thus it is a much more tractable formulation.

For those instances in which user equilibrium assignment would be preferrable in the design application (such as for highway networks), 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 systems 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. For those instances in which these bounds are close (perhaps within 3%), these bounds would justify the use of systems optimal traffic assignment as being a valid approximation to user equilibrium assignmen

In general, it is necessary to use integer programming methods (such as branch and bound) to solve models having discrete investment decision variables (such as Agarwal [1], LeBlanc [4], Ridley [8] , and Roberts [9]). As a result, these models will not be able to handle networks with sizeable dimensions, and, as we have already indicated, neither will models which attempt user equilibrium traffic assignment (such as LeBlanc [4]). According to our discussions with representatives of DOT, in order for a network design model to be useful for planning purposes it is necessary for that model to be able to handle networks having at least a thousand nodes and several thousand links. Only a network design model that uses continuous investment variables and performs systems optimal assignment could handle these large dimensions. The model developed in this section is in that category and is similar to those proposed by Dafermos [3], the Northwestern group [5], and Steenbrink [10, 11].

In her dissertation, Dafermos [3] formulated a network design model which assumed that the total travel time on a link was a quadratic function of the flow on that link and that the only effect of investment was to change the link capacity, rather than the free-flow travel time. Using systems optimal traffic assignment with the differentiable objective function that she proposed, Dafermos showed that a gradient technique could be employed to determine the global optimal solution to the design problem; however, no computational results were given.

The Northwestern group [5] formulated the network design problem as a linear programming model by assuming that the total travel time on a link could be expressed as a piecewise linear function of the flow on that link and that the only effect of investment was to shift the locations of the breakpoints of this piecewise linear curve. The Northwestern group illustrated their approach by solving a highway network consisting of 24 nodes and 76 links, using the standard simplex method, in about 15 minutes on a CDC 6400 computer. Thus a significant difficulty with Northwestern's approach is the small size of problems which can be solved. The Northwestern group also experimented with Dantzig-Wolfe decomposition [12, 13] and the BOXSTEP method [14], but with little success.

Steenbrink [10, 11] used his model to plan the road investments in the Netherlands for a network consisting of 351 origins and destinations, about 2,000 nodes, and about 6,000 links. A non-linear differentiable function, similar to the highway travel time curve utilized by the U.S. Federal Highway Administration (FHWA), was used to represent the total travel time on a link as a function of the traffic flow, capacity, and free-flow travel time on that link, and the effect of the investment decision was to change the link capacity. Steenbrink used the following decomposition approach: a separate optimization problem was solved for each link, which created an objective function for a master problem. If the resulting objective function were convex, then the master problem would be equivalent to a traffic assignment problem and thus could be solved with a traffic assignment algorithm. If the objective function were not convex, then Steenbrink suggested that it could be approximated with a convex "envelope" curve. Because Steenbrink used a non-convex investment cost function in his study, the resulting objective function for his master problem was not convex. To solve the master problem, Steenbrink used an heuristic traffic assignment technique called SALMOF (Stepwise Assignment according to the Least Marginal Objective Function) which could only guarantee a local optimal solution, even if a convex objective function were used. Thus, the network design algorithm developed by Steenbrink could only guarantee local optimal solutions, although it could be used to solve the very large network problems with which he was dealing.

In this report we will present a general convex network design model that incorporates the following features: continuous investment decision variables; systems optimal traffic assignment; the total travel time on a link is expressed as a convex function of link flow and investment; and a convex investment cost function. We will show that the basic decomposition procedure devised by Steenbrink [10, 11] can be applied to obtain the global optimal solution to this general model, which includes the following as special cases:

a) FHWA Curve with Changes in Link Capacity. The non-linear differentiable FHWA curve is used to represent the total travel time on a link as a function of the capacity and free-flow travel time on that link. The only effect of investment is to change the link capacity, not the free-flow travel time. This basically was the case analyzed by Steenbrink.

b) Piecewise Linear Travel Time Function with Changes in Either or Both Link Capacity and Free-Flow Travel Time. We will show that Steenbrink's decomposition technique can be applied to give a very efficient solution to the network design model formulated by the Northwestern group [5], which includes a piecewise linear travel time function. We will also show that Northwestern's formulation can be extended to allow for changes to be made in either or both link capacity and free-flow travel time.

c) Quadratic Travel Time Curve with Change in Link Capacity. We will show that Steenbrink's decomposition technique can be applied to give a very efficient solution to the network design model formulated by Dafermos [3], which assumes a quadratic travel time curve with investment only affecting link capacity.

Section 4.2 discusses the three basic travel time curves used in our analyses: the FHWA function, that is similar to the curve used by Steenbrink [10, 11]; the piecewise linear travel time curve that was used by Northwestern [5]; and the quadratic curve proposed by Dafermos [3]. Section 4.3 considers the problem of constructing or expanding links in order to minimize the sum of users' and investment costs; and Section 4.4 considers the problem of constructing or expanding links to minimize the users' cost subject to a budget constraint. Because of the nature of the decomposition procedure being used, we will show that the computational requirements for the network design formulation given in Section 4.3 are approximately the same as for assigning traffic on a network with the same dimensions. However, the computational requirements for the formulation given in Section 4.4 will be greater, as a Lagrange multiplier technique is used to handle the budget constraint.

4.2 TRAVEL TIME AS FUNCTION OF FLOW AND INVESTMENT

An important step in obtaining mathematical formulation of the network design problem is properly treating 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 $D_j(f_j, z_j)$.

4.2.1 FHWA CURVE

A useful curve for modeling the total travel time on a link as a function of flow on that link is in the following form:

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

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 FHWA uses $r = 0.15$ and $k = 4$ for modelling highway congestion (see COMSIS [15; page 35]). Steenbrink [10, 11] used a similar function in his study of road investments in the Netherlands, except that he used $k = 7$ for his initial test computations, and then later changed this parameter to $k = 5$.

We have defined $D_j(f_j, z_j)$ to be the total travel time on link j as a function of the flow f_j and investment decision z_j. If we assume that the only effect of the investment decision variable is to increase link capacity, then we may write

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

where the investment variable z_j is measured in units of capacity. This is the type of function which Steenbrink [10, 11] 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 affect the free-flow travel time on a link, rather than the link capacity. And in other applications, it may be desirable to have the investment modify both 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 affects the free-flow travel time. Unfortunately, it does not appear to be possible*, when using a differentiable travel time curve in the form (1), to allow the investment decision to affect only the free-flow travel time or both the free-flow travel time and capacity simultaneously; however this is possible when using a piecewise linear approximation for the travel time curve, as discussed next.

---

*Suppose, for example, that (1) were modified by replacing t_j with $t_j - z_j$, in order to have the investment decrease the free-flow travel time. Then $D_j(f_j, z_j)$ would not have the convexity that is required by the decomposition approach discussed in Section 4.3.1.

4.2.2 PIECEWISE LINEAR TRAVEL TIME FUNCTION

The Northwestern group [5] assumed that $T_j(f_j)$, the travel time as a function of flow, could be approximated with a piecewise linear function with M_j breakpoints. Define the components x_j^m such that

$$T_j(f_j) = \underset{x_j^m}{\text{MIN}} \sum_{m=1}^{m=M_j} C_j^m \; x_j^m \; , \tag{3.1}$$

subject to

$$f_j = \sum_{m=1}^{m=M_j} x_j^m \; , \tag{3.2}$$

and

$$0 \le x_j^m \le K_j^m \; , \qquad (m = 1, \ldots, M_j), \tag{3.3}$$

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. If the original travel time curve is convex, which is the case for a curve in the form (1) with nonnegative values of k and r, then $C_j^1 \le C_j^2 \le \ldots \le C_j^{M_j}$. The piecewise linear approximation used in the Northwestern study had two breakpoints ($M_j = 2$).

To incorporate investment into the analysis, the Northwestern group next 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 m^{th} 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

$$D_j(f_j, z_j) = \underset{x_j^m}{\mathrm{MIN}} \sum_{m=1}^{m=M_j} C_j^m \; x_j^m , \qquad (4.1)$$

subject to

$$f_j = \sum_{m=1}^{m=M_j} x_j^m , \qquad (4.2)$$

and

$$0 \leq x_j^m \leq K_j^m + F_j^m \; z_j \qquad (m = 1, \ldots, M_j) . \qquad (4.3)$$

By using an argument similar to that given in Section 4.6, 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 $\alpha_1 < \alpha_2 < \ldots < \alpha_M$. Thus

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

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:

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

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

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

because 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 foregoing Northwestern 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 the following could be done. Suppose that at some maximum investment it is desired to have both a new higher capacity c_j' and a new lower free-flow 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 (4.3) 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'. Thus, with this approach, the piecewise linear representation of the travel time curve can handle modifications in either or both the free-flow travel time and link capacity.

LeBlanc and Morlok [16] tested the accuracy of the piecewise linear approximation in a traffic assignment problem. They studied a 24 node, 76 link network that was used to model the city of Sioux Falls, South Dakota: when using systems optimal traffic assignment with the non-linear FHWA travel time curve (see Section 4.2.1), there were a total of 48,100 vehicle hours and 668,400 vehicle miles/day; and when using systems optimal assignment with the piecewise linear approximation ($M_j = 2$), there were a total of 48,700 vehicle hours and 641,900 vehicle miles/day.

4.2.3 QUADRATIC TRAVEL TIME FUNCTION

This subsection will review the results that Dafermos [3] obtained for a particular class of cost functions. She assumed that the total travel time on link j as a function of the flow f_j and investment z_j could be represented as

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

where $A_j(\cdot)$ and $B_j(\cdot)$ are functions which incorporate the effect of investment into the travel time relationship. Thus, the total travel time is a quadratic function of link flow for a particular value of z_j.

Investments which improve the interaction between vehicles on a road are reflected in the function $A_j(\cdot)$ and correspond to increasing the capacity parameter in the formulas given in Sections 4.2.1 and 4.2.2; examples of such improvements are widening a road, adding lanes, improving visibility, etc. Investments which reduce the free-flow travel time are reflected in the function $B_j(\cdot)$; examples of these improvements are shortening the distance, eliminating ramps, straightening curves, etc.

Assuming that $A_j(z_j) \neq 0$ for $z_j \geq 0$, Dafermos showed the the following are necessary and sufficient conditions for $D_j(f_j, z_j)$ to be convex over the region $f_j \geq 0$ and $z_j \geq 0$:

a) $B_j(z_j)$ is independent of z_j;

b) $A_j(z_j) > 0$ for $z_j \geq 0$;

c) $\frac{1}{A_j(z_j)}$ is concave for $z_j > 0$.

Note that condition a) implies that B_j is constant, so that it is not possible to include improvements that reduce the free-flow travel time; this was also the case for the FHWA curve in Section 4.2.1. Note also that conditions b) and c) imply that $A_j(z_j)$ is convex for $z_j \geq 0$.

Additional restrictions on $A_j(\cdot)$ can also be stated. Because investment will improve link conditions, $D_j(f_j, z_j)$ should be a decreasing function of z_j for fixed f_j; thus

d) $\frac{d}{dz_j} A_j(z_j) < 0$ for $z_j > 0$.

Note that conditions b) and d) imply that $\lim_{z_j \to \infty} A_j(z_j)$ exists and is nonnegative. It is reasonable to assume that when an "infinite" amount of investment is applied to a link (so that the width of the link becomes "infinite"), the interaction between vehicles disappears. This corresponds to having

e) $\lim_{z_j \to \infty} A_j(z_j) = 0$.

Dafermos showed that if γ_j, ℓ_j and a_j were positive constants, then

$$A_j(z_j) = a_j \left[\frac{\ell_j}{\ell_j + z_j}\right]^{\gamma_j}$$

would satisfy conditions b), d) and e). In order for condition c) to be satisfied also, it must be true that $0 \leq \gamma_j \leq 1$. In conclusion, a convex function $D_j(\cdot, \cdot)$ satisfying the foregoing conditions is given by

$$D_j(f_j, z_j) = a_j \left[\frac{\ell_j}{\ell_j + z_j}\right]^{\gamma_j} f_j^2 + B_j f_j, \tag{5}$$

where $a_j > 0$, $\ell_j > 0$, and $0 \leq \gamma_j \leq 1$. Because B_j is independent of z_j, we may interpret (5) as representing the case in which the effect of investment is to change only the link capacity, not the free-flow travel time.

4.3 INVESTING TO MINIMIZE USERS' AND INVESTMENT COSTS

This section will analyze the following problem: determine the optimal network design in order to minimize the sum of users' and investment costs, subject to systems optimal traffic assignment. A budget constraint will not be incorporated into the formulation until Section 4.4.

4.3.1 GENERAL CASE

In Section 4.2, we defined $D_j(f_j, z_j)$ as being 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. 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 is equal to the number of origin nodes or to the number of destination nodes, depending upon how it is formulated. Suppose that commodities are distinguished by origin node. Let f_j^r be the total flow on link j that originates from node r.

We now state the network design problem: determine the investment decisions z_j and flows f_j^r in order to minimize the total social transportation costs

$$\text{MINIMIZE} \sum_{j \varepsilon A} D_j(f_j, z_j) + \lambda G_j(z_j), \tag{6}$$

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

$$\sum_{j \varepsilon w_i} f_j^r - \sum_{j \varepsilon V_i} f_j^r = h_i^r \quad (i \varepsilon N;\ r = 1, \ldots, R), \tag{7}$$

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

$$f_j = \sum_{r=1}^{r=R} f_j^r \qquad (j \varepsilon A), \tag{8}$$

upper and lower bounds on the maximum improvement possible

$$L_j \leq z_j \leq P_j \qquad (j \varepsilon A), \tag{9}$$

and non-negativity restrictions

$$f_j^r \geq 0 \quad , \qquad (j \varepsilon A;\ r = 1, \ldots, R), \qquad (10)$$

where

A = set of links in the network

$D_j(f_j, z_j)$ = total travel time on link j as a function of the flow f_j and investment decision z_j

f_j^r = flow on link j from origin r

f_j = total flow on link j

$G_j(z_j)$ = cost for making decision z_j for link j

$$h_i^r = \begin{cases} -O_{ri} & \text{if } i \text{ is a destination node} \\ \sum_j O_{rj} & \text{if } i = r \\ 0 & \text{Otherwise} \end{cases}$$

L_j = minimum value for the investment decision for link j

N = set of nodes in the network

O_{ij} = number of trips from node i to node j

P_j = maximum value for the investment decision for link j

R = number of origin nodes

V_i = set of links terminating at node i

W_i = set of links originating at node i

z_j = investment decision for link j

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

Any continuous convex functions $D_j(\cdot,\cdot)$ and $G_j(\cdot)$ can be used. We assume that λ is nonnegative. For a given value of z_j, the objective function (6) minimizes the total travel time subject to the conservation of flow conditions; thus, the traffic is assigned according to the systems optimal criterion (see Section 2.2). The upper bound P_j on investment could be set by either a physical, technical, environmental, or financial constraint. As we will show in Section 5, the lower bound L_j is needed for multi-stage applications. If $L_j = P_j = 0$, then no improvement is possible on link j.

Next, we will present the decomposition procedure devised by Steenbrink [10, 11]. 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:

$$H_j(f_j) = \underset{z_j}{\text{MINIMUM}} \left[D_j(f_j, z_j) + \lambda G_j(z_j) \right] , \qquad (11)$$

subject to

$$L_j \leq z_j \leq P_j. \qquad (12)$$

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 (11) and (1]) for each value of f_j in order to construct the functions $H_j(\cdot)$ and $I_j(\cdot)$; however, as we will show in Sections 4.3.2 and 4.3.3, 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 (11) and (12) form the objective function for the master problem, and the purpose of the master problem is to compute the total flow on each link. Thus this problem becomes: determine f_j and f_j^r to

$$\text{MINIMIZE} \sum_{j \varepsilon A} H_j(f_j) \tag{13}$$

subject to

$$\sum_{j \varepsilon W_i} f_j^r - \sum_{j \varepsilon V_i} f_j^r = h_i^r \qquad (i \varepsilon N;\ r = 1, \ldots, R), \tag{14}$$

$$\sum_{r=1}^{r=R} f_j^r = f_j \qquad (j \varepsilon A), \tag{15}$$

$$f_j^r \geq 0 \qquad (j \varepsilon A;\ r = 1, \ldots, R). \tag{16}$$

We will show in Section 4.6 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 (13) - (16) is equivalent to the traffic assignment problem, and thus it can be solved by one of the approaches discussed in Section 2. It can be shown that the combination of solving the subproblems (11) and (12) and the master problems (13) to (16) is equivalent to solving the original problem, (6) to (10). Once the optimal value f_j is determined for the j^{th} link, then the optimal investment for that link is given by $I_j(f_j)$.

The success of this particular approach of solving the network design problem is dependent upon how easily the functions $H_j(\cdot)$ and $I_j(\cdot)$ can be obtained for each link, and upon whether $H_j(\cdot)$ is a convex function. Note that the traffic assignment algorithms discussed in Section 2 can be modified to handle non-differentiable objective functions, although they are best suited to work with differentiable functions. However, all of these traffic assignment algorithms require $H_j(\cdot)$ to be convex in order to ensure that a global optimal solution will be obtained. We have examined the following cases:

a) FHWA 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 on 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 Section 4.3.2, we will 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. The piecewise linear function (3) is used to represent the total travel time on a link. 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 (4). If $G_j(\cdot)$ is convex, then $H_j(\cdot)$ will be convex. In Section 4.3.3, we will 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 in the form (1), then we showed in Section 4.2.2 how the piecewise linear representation can be used to model changes made simultaneously in link capacity and 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 (5) 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 Section 4.3.2 and therefore is omitted.

4.3.2 DIFFERENTIABLE TRAVEL TIME FUNCTION WITH CHANGE IN LINK CAPACITY

Consider the case in which the travel time curve is in the form (1) and that the only effect of the investment decision is to change the capacity parameter c_j. Section 4.2.1 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

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

where the variable z_j is measured in units of capacity.

For this case, subproblem (11) - (12) becomes:

$$H_j(f_j) = \underset{z_j}{\text{MINIMUM}} \left\{ t_j\, f_j \left[1 + r\left(\frac{f_j}{c_j + z_j}\right)^k\right] + \lambda\, G_j(z_j) \right\}, \qquad (17)$$

subject to

$$L_j \leq z_j \leq P_j\ . \qquad (18)$$

The function $I_j(f_j)$ is defined as the value of z_j at which $H_j(f_j)$ attains its minimum. In Section 4.6 we will show that if the investment cost $G_j(\cdot)$ is a convex function, then $H_j(\cdot)$ will be convex also.

Next, we will derive explicit formulas for $H_j(\cdot)$ and $I_j(\cdot)$ for the case of a linear investment cost function: $G_j(z) = g_j\, z_j$. Define

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

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

$$\frac{\partial h_j}{\partial z_j} = - \frac{k\ r\ t_j\ (f_j)^{k+1}}{(c_j + z_j)^{k+1}} + \lambda\, g_j = 0.$$

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

$$\phi_j(\lambda) = \left[\frac{\lambda\ g_j}{k\ \ r\ \ t_j}\right]^{\frac{1}{k+1}}$$

It follows that the solutions to the subproblem (17) - (18) are the following:

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

$$H_j(f_j) = \begin{cases} t_j\, f_j \left[1 + r\left(\dfrac{f_j}{c_j + L_j}\right)^k\right] + \lambda g_j\, L_j & 0 \le f_j \le (c_j + L_j)\ \ \phi_j(\lambda) \\ t_j\, f_j \left[1 + (k+1)\, r\ \ \phi_j(\lambda)^k\right] - \lambda g_j\, c_j & (c_j + L_j)\ \ \phi_j(\lambda) \le f_j \\ & f_j \le (c_j + P_j)\ \ \phi_j(\lambda) \\ t_j\, f_j \left[1 + r\left(\dfrac{f_j}{c_j + P_j}\right)^k\right] + \lambda g_j\, P_j & f_j \ge (c_j + P_j)\ \ \phi_j(\lambda)\ . \end{cases}$$

By computing the derivative, the reader can verify that dH_j/df_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 (13) - (16) using the foregoing expression for $H_j(\cdot)$. Because $H_j(\cdot)$ is convex, this master problem can be solved by one of the traffic assignment algorithms discussed in Section 2. Once the optimal flow f_j is determined for the j^{th} link, the optimal investment for that link is $I_j(f_j)$.

4.3.3 PIECEWISE LINEAR TRAVEL TIME FUNCTION WITH CHANGE IN EITHER OR BOTH LINK CAPACITY AND FREE-FLOW TRAVEL TIME

Next consider the case in which the piecewise linear approximation (3) is used for the travel time curve and the only effect of investment is to shift the locations of the breakpoints. Section 4.2.2 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

$$D_j(f_j, z_j) = \underset{x_j^m}{\text{MIN}} \sum_{m=1}^{m=M_j} C_j^m \; x_j^m$$

subject to

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

and

$$0 \leq x_j^m \leq K_j^m + F_j^m \; z_j \qquad (m = 1, \ldots, M_j),$$

when M_j is the number of breakpoints for link j, C_j^m are the slopes of the linear curves, K_j^m are the initial segment lengths, and F_j^m is the proportion of the added capacity assigned to the m^{th} increment.

For this case, subproblem (11) and (12) become:

$$H_j(f_j) = \underset{x_j^m,\; z_j}{\text{MINIMUM}} \left\{ \sum_{m=1}^{M_j} C_j^m \; x_j^m + \lambda G_j(z_j) \right\} , \tag{19}$$

subject to

$$\sum_{m=1}^{M_j} x_j^m = f_j, \tag{20}$$

$$x_j^m \leq K_j^m + F_j^m \; z_j , \qquad (m = 1, \ldots, M_j), \tag{21}$$

$$L_j \leq z_j \leq P_j, \tag{22}$$

$$z_j \geq 0, \quad x_j^m \geq 0 , \qquad (m = 1, \ldots, M_j). \tag{23}$$

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 Section 4.6 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, $C_j^1 \leq C_j^2$. 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 subproblems (19) to (23):

$$H_j(f_j) = \begin{cases} C_j^1 f_j + \lambda g_j L_j & \text{for } f_j \leq K_j^1 + F_j^1 L_j \\[2ex] C_j^1 f_j + \dfrac{(f_j - K_j^1)\lambda g_j}{F_j^1} & K_j^1 + F_j^1 L_j < f_j \leq K_j^1 + F_j^1 P_j \\[2ex] C_j^1 (K_j^1 + F_j^1 P_j) & K_j^1 + F_j^1 P_j < f_j \leq K_j^1 + K_j^2 + (F_j^1 + F_j^2) P_j \\ + C_j^2 (f_j - K_j^1 - F_j^1 P_j) & \\ + P_j \lambda g_j & \end{cases}$$

and

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

And for the case

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

it can be shown that

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

and

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

The next step is to solve the master problem (13) - (16), using the above expression for $H_j(\cdot)$, in order to determine the optimal flow values f_j. Because $H_j(\cdot)$ is convex, problem (13) - (16) is equivalent to the traffic assignment problem with a non-differentiable objective function, and thus it can be solved by one of the approaches discussed in Section 2. The optimal investment on link j is then $I_j(f_j)$. If the original travel time curve is in the form (1), then we showed in Section 4.2.2 how the piecewise linear representation can be used to model changes made in either or both link capacity and free-flow travel time on existing links. It is also possible to handle the introduction of entirely new links by specifying their initial capacity as being zero.

4.4 INVESTING TO MINIMIZE USERS' COST SUBJECT TO BUDGET CONSTRAINT

The network design formulation in Section 4.3 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

$$
\text{MINIMIZE} \sum_{j \varepsilon A} D_j(f_j,\ z_j), \tag{24}
$$

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

$$
\sum_{j \varepsilon W_i} f_j^r - \sum_{j \varepsilon V_i} f_j^r = h_i^r \qquad (i \varepsilon N;\ r = 1, \ldots, R), \tag{25}
$$

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

$$f_j = \sum_{r=1}^{r=R} f_j^r \quad , \qquad (j \varepsilon A), \tag{26}$$

upper and lower bounds on the maximum improvement possible

$$L_j \leq z_j \leq P_j \quad , \qquad (j \varepsilon A), \tag{27}$$

nonnegatively restrictions

$$f_j^r \geq 0 \quad , \qquad (j \varepsilon A;\ r = 1, \ldots, R), \tag{28}$$

and total investment costs limited by the available budget B

$$\sum_{j \varepsilon A} G_j(z_j) \leq B, \tag{29}$$

where the input parameters are defined in Section 4.3.1.

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 (4) and $G_j(\cdot)$ is linear, then the model (24) to (29) is basically the one formulated by the Northwestern group [5]. If $D_j(\cdot,\cdot)$ is given by (5) and $G_j(\cdot)$ is linear, then (24) - (29) is basically the model proposed by Dafermos [3].

Our approach for solving the design problem (24) - (29) is illustrated in Figure 4.1 and involves using a Lagrange multiplier technique to handle the budget constraint (29). As Everett [17] 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 [18] show how to do this systematically with linear programming. As seen in Zangwill [19], this can be thought of as the

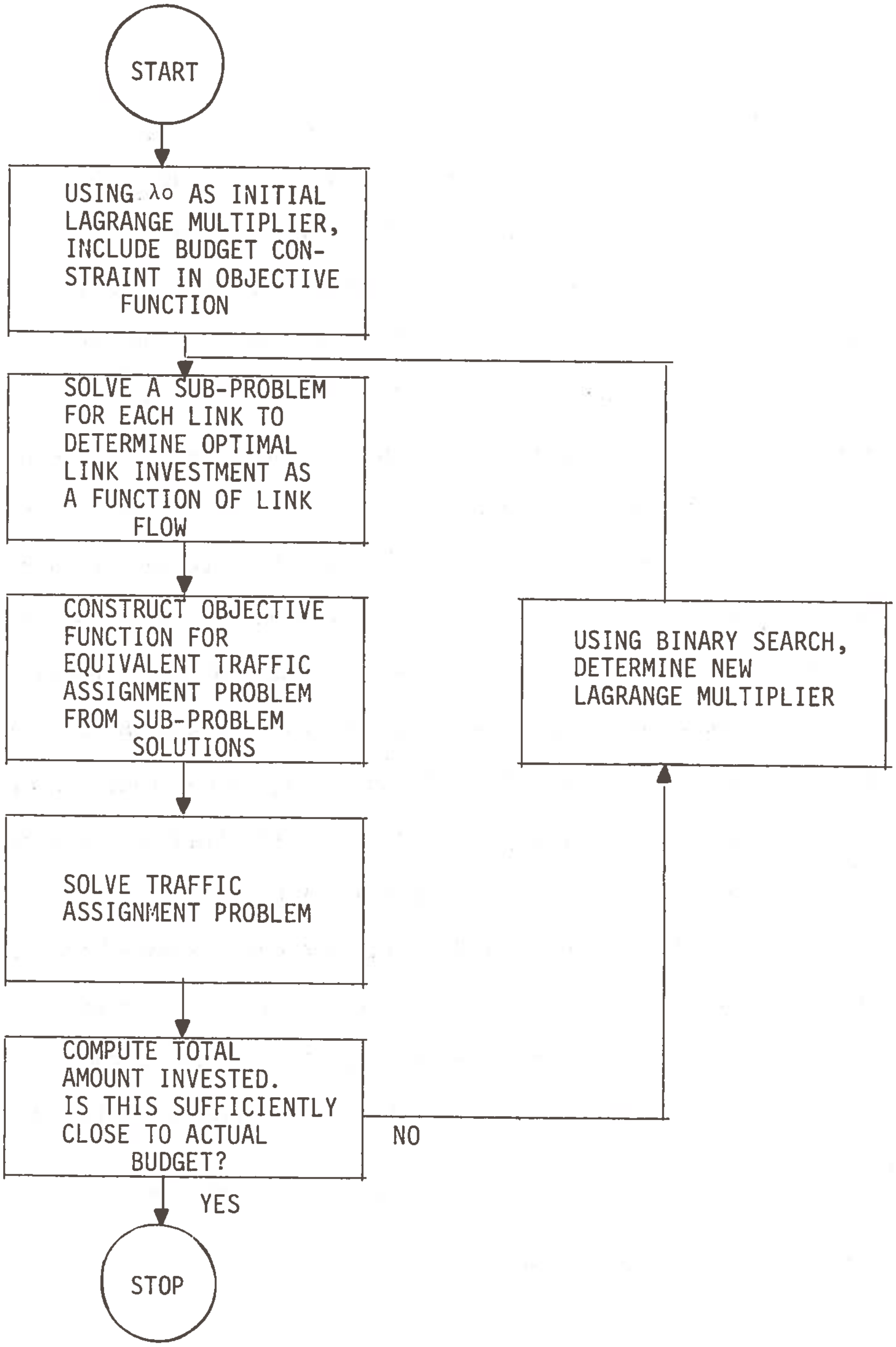


FIGURE 4.1--Solution of Network Design Model with Budget Constraint Using Decomposition

Dantzig-Wolfe decomposition method [12], [13]. However, we suggest using the method of Fox and Landi [20], which is to employ a binary sequential search procedure, sometimes called Bolzano's method (see Wilde and Beightler [21]). The problem of finding the appropriate Lagrange multiplier is equivalent to finding the zero-crossing of a monotone 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 [20] 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 (25) to (28), $Q(w)$ represent the objective function (24), and $R(w) \leq B$ represent the budget constraint (29). Thus (24) to (29) can be rewritten as:

$$\underset{w \in S}{\text{MINIMIZE}}\ Q(w), \text{ subject to } R(w) \leq B. \tag{30}$$

Using Everett's method [17], problem (30) can be (approximately) solved by finding a multiplier λ and the corresponding solution $w(\lambda)$ to

$$\underset{w \in S}{\text{MINIMIZE}}\ [Q(w) + \lambda R(w)]. \tag{31}$$

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 (31) 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 (31) is identical to the model (6) - (10) formulated in Section 4.3.1. Thus, the decomposition approach discussed in Section 4.3.1 shows that problem (31) 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.

4.5 CONCLUSION

The network design approach developed in this section has the following characteristics:

a) Continuous Investment Decision Variables. The algorithm determines the optimal solution based upon continous decision variables. If, however, a discrete solution is needed for a particular application, then one approach is to simply use a discrete solution that is close to the optimal continuous solution.

---

*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.

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. Section 4.1 gave bounds on the user equilibrium design problem using only the solution to the systems optimal design problem; if these bounds were close, then this would justify using the network design based on systems optimal assignment in an application in which user equilibrium assignment was 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 FHWA travel time function, used by Steenbrink [10, 11]; the piecewise linear curve used by the Northwestern group [5]; and the quadratic curve used by Dafermos [3].

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 on existing links, then Section 4.2.1 showed how this could be implemented with a differentiable travel time curve; and if the effect of investment is to change either or both the free-flow travel time and capacity on existing links, then we showed in Section 4.2.2 how this could be done with a piecewise linear travel time curve. The piecewise linear approach can also handle the introduction of entirely new links by specifying their initial capacity as being zero.

f) Solution Algorithm. For the case in which there is no budget constraint but the investment cost is included in the objective function, Section 4.3 showed how the solution to the network design problem could be obtained by solving a traffic assignment problem. For the case in which a budget constraint is used, then Section 4.4 gave a Lagrange multiplier technique that obtains a solution to the design problem by solving a series of traffic assignment problems, one for each value of the multiplier.

4.6 PROOF OF CONVEXITY OF $H_j(\cdot)$

In the decomposition technique discussed in Section 4.3.1, it was necessary for the objective function of the master problem, denoted as $H_j(\cdot)$, to be convex in order for the master problem to be solvable 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), (4), and (5) given for $D_j(\cdot,\cdot)$ in Section 4.2 are convex.

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

$$H_j(f_j) = \underset{z_j}{\text{MINIMUM}} \left[D_j(f_j, z_j) + \lambda\, G_j(z_j)\right] , \qquad (32)$$

subject to

$$L_j \le z_j \le P_j \quad . \qquad (33)$$

<u>THEOREM</u>: Assume that $D_j(f_j, z_j)$ and $G_j(z_j)$ are continuous convex functions defined for $f_j \ge 0$ and z_j satisfying (33), that λ is nonnegative, and that L_j and P_j are finite; then $H_j(\cdot)$ is convex.

<u>PROOF</u>: By the assumptions,

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

is a continuous convex function and is defined for $f_j \geq 0$ and z_j satisfying (33). 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 (33) so that $H_j(f_j^i) = h_j(f_j^i, z_j^i)$. Thus

$$\alpha H_j(f_j^1) + (1 - \alpha) H_j(f_j^2)$$

$$= \alpha h_j(f_j^1, z_j^1) + (1 - \alpha) h_j(f_j^2, z_j^2) \tag{34}$$

$$\geq h_j\left[\alpha (f_j^1, z_j^1) + (1 - \alpha) (f_j^2, z_j^2)\right] \tag{35}$$

$$\geq H_j \left[\alpha f_j^1 + (1 - \alpha) f_j^2\right]. \tag{36}$$

Equation (34) follows from the definition of z_j^i, (35) follows from the convexity of $h_j(\cdot, \cdot)$, and (36) from the definition of $H_j(\cdot)$ given in (32) and (33). Thus $H_j(\cdot)$ is convex, which completes the proof.

4.7 REFERENCES

1. S. K. Agarwal, Optimizing Techniques for the Interactive Design of Transportation Networks Under Multiple Objectives, Ph.D. Thesis, Department of Civil Engineering, Northwestern University, Evanston IL, ch. 2.1.

2. E. C. Carter and J. R. Stowers, "Model for Funds Allocation for Urban Highway Systems Capacity Improvements," Highway Research Record, No. 20, p. 84-102, 1963.

3. S. C. Dafermos, Traffic Assignment and Resource Allocation Transportation Networks, Ph.D. Thesis, The Johns Hopkins University, Baltimore MD, 1968, ch. 3.

4. L. J. LeBlanc, Mathematical Programming Algorithms for Large Scale Network Equilibrium and Network Design Problems, Ph.D. Thesis, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston IL, 1973, ch. 3.

5. E. K. Morlok, J. L. Schofer, W. P. Pierskalla, R. E. Marsten, S. K. Agarwal, J. L. Edwards, L. J. LeBlanc, D. T Spacek, and J. W. Stoner, 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, 1973.

6. Agarwal, op. cit., ch. 2.2

7. LeBlanc, op. cit., ch. 4.

8. T. M. Ridley, "An Investment Policy to Reduce the Travel Time in a Transportation Network," Transportation Research, Vol. 2, p. 409-424, 1968.

9. P. O. Roberts, Transport Planning: Models for Developing Countries, Ph.D. Thesis, Department of Civil Engineering, Northwestern University, Evanston IL, 1966, ch. 3.2.

10. P. A. Steenbrink, "Transport Network Optimization in the Dutch Integral Transportation Study," Transportation Research, Vol 8, p. 11-27, 1974.

11. P. A. Steenbrink, Optimization of Transport Networks, Wiley, New York NY, 1974.

12. G. B. Dantzig and P. Wolfe, "Decomposition Principle for Linear Programs," Operations Research, Vol. 8, p. 101-111, 1960.

13. G. B. Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton NJ, 1963, ch. 23.

14. R. E. Marsten, W. W. Hogen, and J. W. Blankenship, "The BOXSTEP Method for Large-Scale Optimization," Operations Research, Vol 23, p. 389-405, 1975.

15. Comsis Corporation, Traffic Assignment, Prepared for Urban Planning Division, Office of Highway Planning, Federal Highway Administration, Washington DC, August 1973, p. 35.

16. L. J. LeBlanc, and E. K. Morlok, "An Analysis and Comparison of Behavioral Assumptions in Traffic Assignment," Presented at the International Symposium on Traffic Equilibrium Methods, Montreal, CANADA, November 1974.

17. H. Everett, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," Operations Research, Vol 11, p. 399-417, 1963.

18. R. B. S. Brooks and A. M. Geoffrion, "Finding Everett's Lagrange Multipliers by Linear Programming," Operations Research, Vol. 14, p. 1149-1153, 1966.

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

20. B. L. Fox and D. M. Landi, "Searching for the Multiplier in One-Constraint Optimization Problems," Operations Research, Vol. 18, p. 253-262, 1970.

21. D. J. Wilde and C. S. Beightler, Foundations of Optimization, Prentice-Hall, Englewood Cliffs NJ, 1967, p. 230-232.

5. OPTIMAL STAGING OF INVESTMENTS OVER TIME

5.1 INTRODUCTION

The network design model discussed in Section 4 only treated a static problem: a network was to be examined for possible link additions or modifications, without regard to the sequence or timing of the implementation of these changes. However, it may be desirable to perform long-range planning for a highway, mass transit, or rail network in which investments are to be planned during each stage (or year) in a multi-stage horizon, subject to a budget constraint on the total investment in each stage. The staging problem is based on the observation that all network improvements would not contribute equally towards the efficient functioning of the system: some should be added early and some can be delayed. The analysis is complicated by the fact that the user benefits derived from improvements on different links are interrelated in a complex way through the network structure. There are two cases we will consider:

a) The final configuration is not specified. In this case the staging model would determine both the final configuration and the order in which the investments would occur.

b) The final configuration is specified. In this case the staging model would determine only the order in which the recommended network investments should be constructed over the horizon.

Table 5.1 classifies some of the previous approaches for solving the staging problem with respect to the following factors: whether the investment variables are discrete or continuous, type of traffic assignment, and type of objective function. The staging model formulated in this section can be viewed as being a multi-stage version of the network design model formulated in Section 4; thus this staging model will include many of the characteristics of that network design model, such as: continuous investment decision variables, systems

TABLE 5.1

CLASSIFICATION OF STAGING MODELS

Author	Year	Investment Variable	Traffic Assignment	Total Link Travel Time as Function of Flow	Solution Algorithm	Objective
Bergendahl [1]	1969	Discrete	Systems Optimal	Piecewise Linear	First solves a traffic assignment problem using linear programming for each potential configuration at each stage, and then uses dynamic programming to get optimal solution over time	Minimum of discounted sum of travel costs plus investments over horizon plus salvage costs at end of horizon
Roberts [2]	1966	Discrete	Systems Optimal	Linear, with an upper bound on maximum flow	Decomposes the problem into solving a single network design model for each stage, and each of these is solved using integer programming	Lexicographic minimum of the vector of travel costs for each stage, subject to a budget constraint for each stage
Schimpeler-Corradino Associates [3]	1974	Discrete	Systems Optimal	Piecewise Linear	Uses a heuristic procedure which ignores the interdependency of the links in the network	Minimum of sum of total travel time in all stages, subject to a budget constraint for each stage
Proposed Model	1976	Continuous	Systems Optimal	Convex	Decomposes the problem into solving a single network design model for each stage, and each of these is solved using a traffic assignment algorithm	Lexicographic minimum of the vector of travel costs for each stage, subject to a budget constraint for each stage

optimal traffic assignment, travel time being a convex function of link flow and investment, convex investment cost function, etc.

There are two basic types of objective functions that can be used in a staging model. The objective of Bergendahl [1] and Schimpeler-Corradino Associates [3] was to minimize the weighted sum of costs for each individual stage. As we show in Section 5.2, the difficulty with this approach is that it becomes intractable for large networks. An alternative objective function can be specified by the application of the lexicographic ordering rule of vector analysis, and it is the approach used by the staging model formulated in this section. This can be viewed as being a generalization of a method proposed by Roberts [2]: our algorithm allows the order of importance of the stages to be arbitrary, whereas in Roberts' algorithm the most weight was placed upon the costs in the final stage, followed by the next to final stage, then second to final stage, etc.

Section 5.2 discusses the differences between the weighted sum and lexicographic objective functions, Section 5.3 gives a mathematical formulation of the lexicographic staging model, and Section 5.4 presents the solution algorithm. We will show that a T-stage staging problem can be decomposed into T single-stage network design models if the final configuration is not specified, and that it can be decomposed into T-1 single-stage network design models if the final configuration is specified. As demonstrated in Section 4, the network design model with budget constraint, continuous investment decision variables, and systems optimal assignment can be decomposed into a series of traffic assignment problems using the Lagrange multiplier technique. We are, therefore, able to decompose the staging model into a series of traffic assignment problems.

5.2 OBJECTIVE FUNCTIONS

This section compares the use of the weighted sum and lexicographic objective functions when staging investments over a planning horizon. Suppose that there are T stages in the horizon. Let

U_t = total user travel cost (or time) during the t^{th} stage ($t = 1,\ldots,T$),

where U_t could refer to the costs summed over all days in the stage, or to only the costs on the final day of the stage, or to only the costs associated with the peak hours on the final day of the stage. For definiteness, we will assume in this section that the staging model is to determine both the final configuration at the end of the horizon and the order of construction during the horizon.

5.2.1 WEIGHTED SUM OBJECTIVE

Perhaps the most straightforward objective would be to stage the investment decisions so as to

$$\text{MINIMIZE} \sum_{t=1}^{T} w_t U_t,$$

where w_t are appropriate weights. For example, suppose that each stage corresponded to one year and δ is the annual discount rate. If we assume that the demands and costs are stationary after the end of the horizon, then

$$w_t = \frac{1}{(1+\delta)^t} \qquad \text{for } t = 1, \ldots, T-1,$$

and

$$w_T = \sum_{t=T}^{\infty} \frac{1}{(1+\delta)^t} = \frac{1}{\delta(1+\delta)^{T-1}} \tag{1}$$

are the weights that would allow the present value of future travel costs to be computed. Note that for realistic values of δ and T (such as T=20 and δ=.08), $w_T > w_t$ for $t < T$. This corresponds to the observation that the greatest benefits from investment will occur after construction is completed, not while construction is still in progress.

One approach that could be used for solving the staging model with a weighted sum objective function is to formulate this problem as a linear program. This can be done if the following assumptions are made*: a piecewise linear approximation is used (see Section 4.2.2) to express the total link travel time as a function of link flow, and the only effect of investment is to shift the locations of the breakpoints in a linear manner; traffic is assigned according to the systems optimal criterion; the investment decision variables are continuous; and the investment cost function is linear. The number of constraints in this formulation is equal to

$$N\ R\ T + (M + 2)\ A\ T + T,$$

where

N = number of nodes
R = number of origins
T = number of stages
A = number of links
M = number of segments in piecewise linear approximation for travel time function.

---

*These assumptions were made in the network design model that was formulated and solved as a linear program by the Northwestern group [4]. The staging model (with a weighted sum objective) formulated by Schimpeler-Corradino Associates [3] makes these same assumptions, except that they used discrete investment decision variables and thus obtained a mixed integer linear program. In Section 5.3 the staging model is formulated with more general assumptions and a lexicographic objective, but it could provide a guide to the reader as to how the linear programming version would be written.

For even moderate values of N, R, T, A, and M, the above expression quickly exceeds the capacity of commercially available linear programming codes. For example, if N = 1000, R = 250, T = 4, A = 3000, and M = 2, then the linear programming model would have over one million constraints. However, these constraints do possess the staircase structure that is characteristic of dynamic capacity expansion models. Thus, the nested decomposition approach of Glassey [5] and Ho and Manne [6] could be employed. However, the value of the nested decomposition procedure appears to be in reducing storage requirements, rather than in reducing computer time (see specifically the test problem results given in [5] and [6]). Thus, we conclude that in practice it would not be feasible to use the weighted sum objective function when staging investments over large networks. It is for this reason that Schimpeler-Corradino Associates [3] suggested using heuristics when solving large problems with the weighted sum objective function.

5.2.2 LEXICOGRAPHIC OBJECTIVE

An alternative approach to the weighted sum objective function is to employ the concept of lexicographic ordering of vectors (refer to Dantzig [7, page 294]). A vector R is defined to be greater than zero in the lexicographic sense (or lexico-positive and denoted $R \succ 0$), if it has at least one non-zero component, the first of which is positive. If vectors R and S have the same number of components, then R is defined to be greater than S in the lexicographic sense (denoted $R \succ S$), if R - S is greater than zero in the lexicographic sense (or $R - S \succ 0$). The term "lexicographic" is used because this ordering is similar to the way one orders a set of words in a dictionary.

Let

$$Z = [U_{\alpha(1)}, U_{\alpha(2)}, \ldots, U_{\alpha(T)}]$$

be the vector of travel costs, where the function $\alpha(t)$, $t=1, \ldots, T$, defines the order in which the components U_t are included in the vector Z. For example, if $\alpha(1) = T$, $\alpha(2) = T - 1$, ..., $\alpha(T) = 1$, then

$$Z = [U_T, U_{T-1}, \ldots, U_1]. \quad (2)$$

Or if $\alpha(1) = T$, $\alpha(2) = 1$, $\alpha(3) = 2$, ..., $\alpha(T) = T - 1$, then

$$Z = [U_T, U_1, U_2, \ldots, U_{T-1}]. \quad (3)$$

Thus, the objective for the staging model is to find a vector Z that is the lexico-minimum of all travel cost vectors (i.e., $R \succ Z$ for all other possible travel cost vectors R). We express this notationally as:

$$\text{LEXICO-MINIMIZE } [U_{\alpha(1)}, U_{\alpha(2)}, \ldots, U_{\alpha(T)}].$$

With this ordering, the optimal investment schedule during stages $t = 1, 2, \ldots, T$ will

(1) minimize travel costs $U_{\alpha(1)}$ among all possible investment schedules during horizon;

(2) minimize travel costs $U_{\alpha(2)}$ among all investment schedules satisfying (1);

(3) minimize travel costs $U_{\alpha(3)}$ among all investment schedules satisfying (1) and (2);

.

.

.

(T) minimize travel costs $U_{\alpha(T)}$ among all investment schedules satisfying conditions (1), (2), ..., (T - 1).

In comparing the weighted sum and lexicographic objective functions, it may be useful to think of the lexicographic objective as being a limiting case of the weighted sum objective, where a very large weight is placed on stage $\alpha(1)$; a much smaller in comparison, but still large weight is placed on stage $\alpha(2)$; etc. Notationally, this becomes:

$$w_{\alpha(1)} >> w_{\alpha(2)} >> \ldots >> w_{\alpha(T)} .$$

It is probably desirable to have $\alpha(1) = T$, so that the lexicographic objective would guarantee that the best possible network design is obtained at the end of the last stage in the planning horizon. This corresponds to the fact that $w_T > w_t$ for $t < T$, when realistic values are used for T and δ in the expressions for w_t in (1).

The staging algorithm proposed by Roberts [2] determined the lexicographic minimum of the vector of travel costs, but required this vector to be in the form (2); i.e., the most weight was placed upon the last stage, followed by next to last, second to last, etc. The staging algorithm that we propose in this report generalizes Roberts' algorithm to allow the order function $\alpha(t)$ to be specified in any arbitrary way; this means that the terms U_t can appear in any order in the travel cost vector. In particular, our algorithm will allow the travel cost vector to be in the form (3), which is perhaps the most useful objective function in staging applications; i.e., the most weight is placed upon the last stage, followed by the first stage, second stage, etc. Note that this is the order of importance of the weights given in (1). Thus, we may view the solution obtained from solving a staging problem with a lexicographic objective in the form (3) as being an approximation to the solution that would have been obtained, if the staging problem were solved using a weighted sum objective with the weights (1).

5.3 FORMULATION

A mathematical description of the staging model with lexicographic objective function is given in this section. Because this staging model can be viewed as being a multi-stage version of the network design model discussed in Section 4, it will include many of the characteristics of the network design model: continuous investment decision variables, systems optimal traffic assignment, travel time on a link is expressed as a convex function of link flow and investment, convex investment cost function, etc. Also, the same basic notation that was developed in Section 4 will be used here, with the addition of the subscript t to refer to the t^{th} stage. With this modification, the investment decision variables become:

z_{jt} = total capacity added to link j during the first t stages.

Thus, $z_{jt} - z_{j,t-1}$ is the amount of capacity added to the j^{th} link in the t^{th} stage.

Let $D_{jt}(f_{jt}, z_{jt})$ be the total travel time (or cost) in stage t on link j as a function of the link flow f_{jt} and investment decision z_{jt} for that link. Any continuous convex function can be used for $D_{jt}(\cdot,\cdot)$, including the following: the non-linear differentiable curve, similar to the FHWA (U.S. Federal Highway Administration) travel time function, that was used in the network design model formulated by Steenbrink [8, 9] (see Section 4.2.1); the piecewise linear curve used in the network design model formulated by the Northwestern group [4] (see Section 4.2.2), which includes as special cases the functions used in the staging models formulated by Roberts [2] and Schimpeler-Corradino Associates [3] ; and the quadratic curve used in the network design model formulated by Dafermos [10] (see Section 4.2.3).

Let B_t be the cumulative budget for the first t stages, and $G_j(z_{jt})$ be the investment cost for adding capacity z_{jt} to the j^{th} link. Any continuous convex function can be used for $G_j(\cdot)$. We require that the staging of investments over the planning horizon meet the following conditions:

a) No construction should be removed after it has been installed; this corresponds to requiring that

$$z_{jt} \leq z_{j,t+1} \quad ,$$

for all links j and stages t.

b) The total investment costs during the first t stages cannot exceed the cumulative budget, or

$$\sum_{j \varepsilon A} G_j(z_{jt}) \leq B_t,$$

where A is the set of links, and this inequality holds for each stage t.

We are now ready to formulate the staging model and will consider first the case in which both the final configuration and the staging of investments are to be chosen. Thus, the staging problem becomes: determine the decision variables f^r_{jt} and z_{jt} in order to obtain the lexicographic minimum of the vector of travel costs

$$\text{LEXICO-MINIMIZE} \left[U_{\alpha(1)}, U_{\alpha(2)}, \ldots, U_{\alpha(T)} \right], \tag{4}$$

where the total travel costs for the t^{th} stage is computed as

$$U_t = \sum_{j \varepsilon A} D_{jt}(f_{jt}, z_{jt}) \qquad (t=1, \ldots, T), \tag{5}$$

conservation of flow equations are defined for each node i, origin r, and stage t

$$\sum_{j \varepsilon W_i} f^r_{jt} - \sum_{j \varepsilon V_i} f^r_{jt} = h^r_{it} , \tag{6}$$

$$(i \varepsilon N;\ r=1, \ldots, R;\ t=1, \ldots, T),$$

total flow on link j is equal to the sum of flows from each source r

$$f_{jt} = \sum_{r=1}^{r=R} f^r_{jt} \qquad (j \varepsilon A;\ t=1, \ldots, T), \tag{7}$$

total capacity added to link j during first t stages cannot exceed the capacity added during the first $t+1$ stages or the upper bound P_j

$$z_{jt} \leq z_{j,t+1} \text{ and } z_{jT} \leq P_j \qquad (j \varepsilon A;\ t=1, \ldots, T-1), \tag{8}$$

nonnegativity restrictions

$$f^r_{jt} \geq 0, \quad z_{jt} \geq 0 \qquad (j \varepsilon A;\ r=1, \ldots, R;\ t=1, \ldots, T), \tag{9}$$

and total investment costs during the first t stages cannot exceed the available budget

$$\sum_{j \varepsilon A} G_j(z_{jt}) \leq B_t \qquad (t=1, \ldots, T), \tag{10}$$

where

A = set of links in the network

B_t = available budget during first t stages

$D_{jt}(f_{jt}, z_{jt})$ = total travel cost on link j in stage t as a function of the flow f_{jt} and investment decision z_{jt}

f_{jt} = total flow on link j in stage t

f^r_{jt} = flow on link j from origin r in stage t

$G_j(z_{jt})$ = investment cost for adding capacity z_{jt} to link j

$$h^r_{it} = \begin{cases} -O_{rit} & \text{if } i \text{ is a destination node} \\ \sum_j O_{rjt} & \text{if } i = r \\ 0 & \text{otherwise} \end{cases}$$

N = set of nodes in the network

O_{ijt} = number of trips from node i to node j in stage t

P_j = maximum capacity that can be added to link j

R = number of origin nodes

T = number of stages

U_t = total travel costs for the t^{th} stage

V_i = set of links terminating at node i

W_i = set of links originating at node i

solve problems of this size using UMTA's computer package called the Transportation Planning System (UTPS). Although this package uses a somewhat different approach than that suggested by the decomposition scheme above (See the Frank-Wolfe algorithm in Nguyen [9] and for a detailed description see Ruiter [10]), the performance of the algorithms is similar. A difficulty with this general methodology arises when very large problems are considered, such as the planning model for the city of Los Angeles ($|N| \cong 10{,}000$; $|A| \cong 30{,}000$; $R \cong 1{,}000$). Here the sheer size of the network permits only a very rough solution to be computed; whereas if the network could in some way be reduced in size, a more accurate solution could be obtained. The geographic decomposition approach presented in this report is aimed at reducing to a more manageable size these very large networks.

6.3 GEOGRAPHIC DECOMPOSITION

Geographic decomposition is based on the observation that very large networks are often only loosely connected; in other words, if a small set of links are deleted from such a network, it will decompose into a series of disjoint subnetworks. For example, if four links are deleted from the network in figure 6.1,

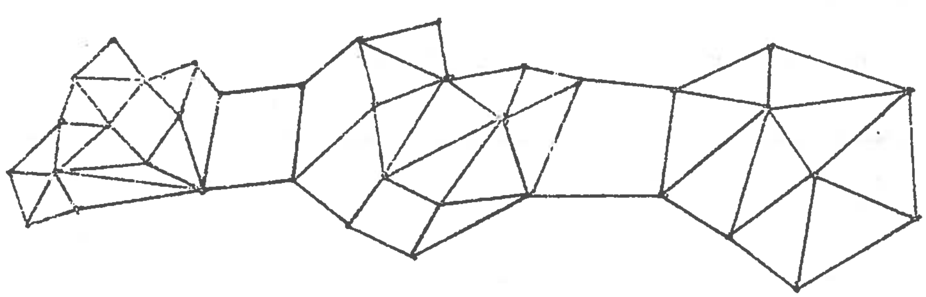


FIGURE 6.1--Disjoint Sub-Networks.

we could decompose the network into a series of unconnected sub-networks, as in Figure 6.2.

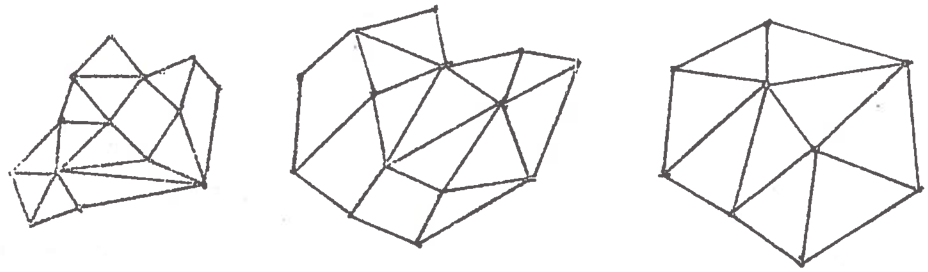


FIGURE 6.2--Unconnected Sub-Networks.

Geographic decomposition of the network for the problem of determining only the shortest path between the origin-destination pairs has been previously proposed by Hu [3], [4]; Lansdowne [5]; and Ramsburg and Aronoff [6]. Our approach will be to apply this general concept directly to the traffic assignment problem using Goeffrion's generalized Benders Decomposition [1].

Let S be the set of deleted links. With the removal of the links in S, the network (N,A) can be decomposed into a series of disjoint subnetworks (N_ℓ, A_ℓ) $\ell = 1, \ldots, L$, where

$$N = N_1 \cup N_2 \cup \ldots \cup N_L$$

$$A = A_1 \cup A_2 \cup \ldots \cup A_L \cup S\ .$$

The first version of the algorithm can now be stated.

6.3.1 GEOGRAPHIC DECOMPOSITION -- VERSION I

Step 0. (Initialization)

Determine an initial set of flows $\bar{f}_j^r$ for $j \varepsilon S$; $r = 1, \ldots, R$.

let $\bar{f}_j = \sum_{r=1}^{r=R} \bar{f}_j^r$ for $j \varepsilon S$.

Set $UBD \leftarrow \infty$ "upper bound on solution value"

$\bar{Z}_M \leftarrow -\infty$ "lower bound on solution value"

$K \leftarrow 0$ "index for subproblem's dual feasible solutions"

$K' \leftarrow 0$ "index for subproblem's dual unbounded solutions"

Select convergence parameters ε_1 and ε_2.

Step 1A (Find feasible solutions for the subproblems)

Solve the subproblems for $\ell = 1,..,L;\ r = 1,...,R$:

$$\text{MINIMIZE} \quad W^{\ell}_{SUB(r)} = \sum_{i\varepsilon N_\ell} (v_i^+ + v_i^-) \ , \tag{9}$$

with respect to: f_j^r, v_i^+, and v_i^-, $j\varepsilon A_\ell$, $i\varepsilon N_\ell$,

subject to:

$$\sum_{j\varepsilon W_i - S} f_j^r - \sum_{j\varepsilon V_i - S} f_j^r + v_i^+ - v_i^- = h_i^r - \left(\sum_{j\varepsilon W_i \cap S} \bar{f}_j^r - \sum_{j\varepsilon V_i \cap S} \bar{f}_j^r\right), \tag{10}$$

$$(i\varepsilon N_\ell) \ ,$$

$$v_i^+ \geq 0, \ v_i^- \geq 0 \ , \qquad (i\varepsilon N_\ell), \tag{11}$$

$$f_j^r \geq 0 \ , \qquad (j\varepsilon A_\ell) \ . \tag{12}$$

IF $W^{\ell}_{SUB(r)} > 0$ for any ℓ or r, then:

Let $K' \leftarrow K' + 1$.

Let $\bar{\lambda}^{K'}_{ri}$ be the optimal dual variables associated with constraint (10).

Go to Step 2.

Else, if $W^{\ell}_{SUB(r)} = 0$ for all ℓ and r, then let $\bar{f}_j^r$ be the optimal solution obtained for the problem (9) to (12) and go to Step 1B.

Step 1B. (Find optimal solutions for the subproblems)

Set $K \leftarrow K + 1$.

Solve the traffic assignment problems for $\ell = 1,\ldots,L$:

$$\text{MINIMIZE} \quad Z_{SUB}^{K,\ell} = \sum_{j\varepsilon A_\ell} \int_0^{f_j} C_j(x)dx \quad , \tag{13}$$

with respect to: f_j^r and f_j, $j\varepsilon A_\ell$, $r=1, \ldots,R$,

subject to:

$$\sum_{j\varepsilon W_i - S} f_j^r - \sum_{j\varepsilon V_i - S} f_j^r = h_i^r - (\sum_{j\varepsilon W_i \cap S} \bar{f}_j^r - \sum_{j\varepsilon V_i \cap S} \bar{f}_j^r) \quad , \tag{14}$$

$$(i\varepsilon N_\ell;\ r = 1,\ldots,R) \quad ,$$

$$f_j = \sum_{r=1}^{r=R} f_j^r \quad , \qquad (j\varepsilon A_\ell) \quad , \tag{15}$$

$$f_j^r \geq 0 , \qquad (j\varepsilon A_\ell;\ r = 1,\ldots,R). \tag{16}$$

Let $\bar{u}_{ri}^K$ be the dual variables associated with constraints (14) on iteration K. The $\bar{u}_{ri}^K$ can be interpreted as the cost of sending an additional vehicle from origin r to node i, which may or may not be a destination.

Note: In order to determine the $\bar{u}_{ri}^K$ uniquely, we will select one node $\bar{i}$ in each set N_ℓ so that

$$\bar{u}_{r\bar{i}}^K = 0 \text{ for } \ell = 1,\ldots,L;\ r = 1,\ldots,R.$$

This corresponds to deleting a redundant constraint from (14) for $\ell = 1,\ldots L$ and $r = 1,\ldots,R$.

Compute:

$$Z_{SUB}^K = \sum_{\ell=1}^{\ell=L} Z_{SUB}^{K,\ell} \quad .$$

$$\text{Set UBD} \leftarrow \text{minimum } \{\text{UBD}, Z^K_{SUB} + \sum_{j\varepsilon S} \int_0^{\bar{f}_j} C_j(x)dx\}$$

$$\bar{M}^K = Z^K_{SUB} - \sum_{r=1}^{r=R} \sum_{i\varepsilon N} \bar{u}^K_{ir} \left(\sum_{j\varepsilon W_i \cap S} \bar{f}^r_j - \sum_{j\varepsilon V_i \cap S} \bar{f}^r_j\right).$$

If $\bar{Z}_M \geq \text{UBD} - \varepsilon_1$

or

$\left|\bar{u}^K_{ra} - \bar{u}^K_{rb} - C_j(\bar{f}_j)\right| \leq \varepsilon_2$ where the link $j = (a, b)$,

for all $r = 1,\ldots,R$; $j\varepsilon S$

terminate the algorithm.

Go to Step 2.

Step 2 (Optimize the master)

Solve the generalized Bender's master problem:

$$\text{MINIMIZE } Z_M = \sum_{j\varepsilon S} \int_0^{f_j} C_j(x)dx + x_0 , \tag{17}$$

with respect to: f^r_j, f_j and x_0, $j\varepsilon S$, $r=1, \ldots, R$.

subject to:

$$\sum_{i\varepsilon N_\ell} [h^r_i - (\sum_{j\varepsilon W_i \cap S} f^r_j - \sum_{j\varepsilon V_i \cap S} f^r_j)] = 0 \tag{18}$$

$(r = 1,\ldots,R;\ \ell = 1,\ldots,L-1)$,

$$x_0 \geq \bar{M}^K + \sum_{r=1}^{r=R} \sum_{i \varepsilon N} \bar{u}^k_{ri} \left(\sum_{j \varepsilon W_i \cap S} f^r_j - \sum_{j \varepsilon V_i \cap S} f^r_j \right) , \tag{19}$$

$$(k = 1, \ldots, K) ,$$

$$\sum_{r=1}^{r=R} \sum_{i \varepsilon N} \bar{\lambda}^k_{ri} \left[h^r_i - \left(\sum_{j \varepsilon W_i \cap S} f^r_j - \sum_{j \varepsilon V_i \cap S} f^r_j \right) \right] \leq 0 , \tag{20}$$

$$(k = 1, \ldots, K') ,$$

$$f_j = \sum_{r=1}^{r=R} f^r_j \qquad (j \varepsilon S) , \tag{21}$$

$$f^r_j \geq 0 \qquad (j \varepsilon S,\ r = 1, \ldots, R) , \tag{22}$$

and obtain $\bar{Z}_M$, $\bar{f}_j$, and $\bar{f}^r_j$ $(j \varepsilon S,\ r = 1, \ldots, R)$.

Go to Step 1A.

6.3.2 DISCUSSION

Version I of the geographic decomposition algotithm is based on applying generalized Benders Decomposition directly to the formulation (1) - (4). The major shortcoming of this approach is the very large master problem that must be solved in Step 2. If the set S of deleted links contains then for a problem with 1,000 origin nodes we could have a master problem with over 100,000 variables. If the set S divided the network into 5 subnetworks, then the number of constraints would be 4,100 + K + K', which of course grows as the iteration counters K and K' increase. However, the situation is far from hopeless, since the master problem possesses a great deal of structure.

First, one should recognize that by far the largest number of constraints (4,000) are generated by (18). However, these equations are just the network conservation of flow constraints, which are separable for each distinct origin node r, r = 1,...R.

Second, the equations (20) can be rewritten in a form that will also permit them to be separated by origin node, namely,

$$\sum_{i \in N} \bar{\lambda}^k_{ri} \, [h^r_i - (\sum_{j \in W_i \cap S} f^r_j - \sum_{j \in V_i \cap S} f^r_j)] \leq 0 \, , \qquad (20')$$

$$(r = 1,\ldots,R;\ k = 1,\ldots,K').$$

This alternative formulation is based on the observation that the subproblem of Step 1A has already been separated by origin node r, and each such independent subproblem can make its own feasibility test.

Finally, we observe that equations (19) and (21) can be entirely eliminated if the objective function is rewritten in the form:

$$\text{MINIMIZE } Z_M = \sum_{j \in S} \int_0^{f^1_j + \ldots + f^R_j} C_j(x)\,dx \; +$$

$$\underset{k}{\text{MAXIMUM}} \{ \bar{M}^k + \sum_{r=1}^{r=R} \sum_{i \in N} \bar{u}^k_{ri} \, (\sum_{j \in W_i \cap S} f^r_j - \sum_{j \in V_i \cap S} f^r_j)\}. \qquad (17')$$

This form of the objective function is jointly convex in f^r_j, since the functions over which the maximum is taken are convex. Since the only interaction that remains between origin nodes occurs in the convex objective function, we can now apply to the formulation (17'), (18), (20'), (22) the natural decomposition procedure suggested by Murchland [8] and Nguyen [7] for the traffic assignment problem to obtain:

Step 2. (Revised)

Substep 2A.

Let $\bar{f}_j^r$, $j \varepsilon S$, $r = 1,\ldots,R$ be the last obtained solution.

Let $r = 1$.

Substep 2B.

Solve the master problem M(r) defined as:

$$\text{MINIMIZE} \quad Z_{M(r)} = \sum_{j \varepsilon S} \int_0^{f_j} C_j(x)dx + x_0 \quad ,$$

with respect to: f_j^r, f_j, and x_0, $j \varepsilon S$,

subject to:

$$\sum_{i \varepsilon N_\ell} [h_i^r - (\sum_{j \varepsilon W_i \cap S} f_j^r - \sum_{j \varepsilon V_i \cap S} f_j^r)] = 0 \quad ,$$

$$(\ell = 1,\ldots,L-1) \quad ,$$

$$x_0 \geq \bar{M}_r^k + \sum_{i \varepsilon N} \bar{u}_{ri}^k (\sum_{j \varepsilon W_i \cap S} f_j^r - \sum_{j \varepsilon V_i \cap S} f_j^r) \quad ,$$

$$(k = 1,\ldots,K) \quad ,$$

$$\sum_{i \varepsilon N} \bar{\lambda}_{ri}^k [h_i^r - (\sum_{j \varepsilon W_i \cap S} f_j^r - \sum_{j \varepsilon V_i \cap S} f_j^r)] \leq 0 \quad ,$$

$$(k = 1,\ldots,K') \quad ,$$

$$f_j = f_j^r + K_j^r \qquad (j \varepsilon S) \quad ,$$

$$f_j^r \geq 0 \qquad (j \varepsilon S) \quad ,$$

where

$$K_j^r = \sum_{p \neq r} \bar{f}_j^p ,$$

and

$$\bar{M}_r^k = \bar{M}^k + \sum_{p \neq r} \sum_{i \in N} \bar{u}_{pi}^k \left(\sum_{j \in W_i \cap S} \bar{f}_j^p - \sum_{j \in V_i \cap S} \bar{f}_j^p \right) ,$$

where $\bar{f}_j^p$, $j \in S$, $p = 1,\ldots,R$ is the current flow pattern.

Substep 2C

Revise the flow pattern $\bar{f}_j^r$, $j \in S$, $r = 1,\ldots,R$ based on the last solution of M(r). If the flow pattern is simulatenously optimal for all problems M(r), r = 1,...,R, terminate. If r = R, set r = 1, else set r = r + 1. Return to substep 2B.

At the completion of Step 2 (Revised), set $\bar{Z}_M = Z_{M(r)}$ for the last r solved, and obtain $\bar{f}_j$ and $\bar{f}_j^r$ for $j \in S$, $r = 1,\ldots,R$.

Each master problem M(r) of Step 2B now has only 100 variables and 4 + K + K' constraints. This size problem can be solved directly as a non-linear programming problem; or if we choose to use a linear approximation for the convex separable objective function, we can solve the problem as a linear program. In either case, the size of the master problem no longer presents a computational difficulty.

6.3.3 REMARKS

The subprogram in Step 1B can be solved by any algorithm for the traffic assignment problem. If the natural decomposition of Nguyen and Murchland, which we have just applied to the master problem is used, then the problem will decompose by commodities (origin nodes) into a set of very simple convex cost transhipment problems.

The Bender's master problem of Step 2 is somewhat unusual because of the addition of constraints (18). These equations were included in the formulation to increase the probability that the subproblems will be feasible. In effect, they simply state that for each commodity r and subnetwork ℓ, the sum of the trips generated at all origins must equal the sum of the trips terminated at all destinations. This condition is necessary for the subproblems to be feasible. (Note: Because of the deletion of the arcs in the set S, there will be generated many pseudo origins and destinations. These must be included in the calculations.) If these constraints were omitted, then we would expect to find many cases in which the subproblems would prove to be infeasible. Each such occurrence would generate a constraint (20), and eventually we would force the master problem to a feasible solution. However, this would be a very inefficient procedure, and as we argue below, the inclusion of (18) will eliminate most of our infeasibility problems. Furthermore, we can show that (20) can generate (18) as follows. Notice that $\bar{\lambda}^{k}_{ri}$ will only take on the values $\pm$ 1 or 0. If $\bar{\lambda}^{k}_{ri} = 1$ for $i \epsilon N_\ell$ and 0 for all other i, and then at a later iteration $\bar{\lambda}^{k+1}_{ri} = -1$ for $i \epsilon N_\ell$ and 0 for all other i, we will generate precisely one of the constraints in (18).

If every network (N_ℓ, A_ℓ) is strongly connected, as defined below, then the subproblems will never be infeasible, and the test in Step 1A can be omitted. The computational advantage of this property rests in the fact that we will not generate any constraints (20) for the master. The mere fact that Step 1a can be omitted is relatively unimportant, since we must still find a feasible starting solution for the subproblem (13) - (16).

Let the set of nodes $N^r \subseteq N$ be defined as all nodes in the network (N_ℓ, A_ℓ) for commodity r that are either:

(1) an original origin or destination node, i.e., $h_i^r \neq 0$.

(2) a node at which an arc in the set S either originated or terminated.

If for each pair of nodes i and j in the set N_ℓ^r there exists a directed chain from i to j and a directed chain from j to i, and such chains exists for all N_ℓ^r, $r = 1,\ldots,R$, the network (N_ℓ, A_ℓ) will be called strongly connected. (A directed chain is a set of links that connect i and j which are all oriented in the same direction.)

In general, one would expect a transportation network to be strongly connected, since the set of nodes in N_ℓ^r are usually centers for demand zones, and a route always exist between such centers. A simple condition that will guarantee a strongly connected network is: if there exists an arc $j \varepsilon N_\ell$ from the node a to the node b, then there also exists an arc $j' \varepsilon N_\ell$ from the node b to the node a.

The second version of the geographic decomposition procedure is obtained by first applying the natural decomposition by commodity described earlier, and then applying generalized Bender's decomposition to the problem of Step 2. The conceptual advantage of this procedure is a greatly reduced size for the master problem, and the processing of each commodity r in sequence and independent of all other commodities.

6.3.4 GEOGRAPHIC DECOMPOSITION -- VERSION II

Step 1. (Initialization)

Determine an initial set of flows $\bar{f}_j^r$ for $j \varepsilon S$; $r = 1,\ldots,R$.

Also, select an initial set of flows f_j^r for $j \varepsilon N_\ell$; $\ell = 1,\ldots,L$; $r = 1,\ldots,R$.

Let $\bar{f}_j = \sum_{r=1}^{r=R} \bar{f}_j^r$ for $j \varepsilon S$.

Set UBD $\leftarrow \infty$ "upper bound on solution for current commodity"

$\bar{Z}_M \leftarrow -\infty$ "lower bound on solution for current commodity"

K $\leftarrow 0$ "index for subproblem's dual feasible solutions"

K' $\leftarrow 0$ "index for subproblem's dual unbounded solutions"

r $\leftarrow 1$ "index for the current commodity being optimized"

Select convergence parameters ε_1 and ε_2.

Step 1A. (Find feasible solutions for subproblem r)

Solve the subproblem for $\ell = 1,\ldots,L$:

$$\text{MINIMIZE} \quad W^{\ell}_{SUB(r)} = \sum_{i \varepsilon N_\ell} (v_i^+ + v_i^-) \;, \tag{23}$$

with respect to: f_j^r, v_i^+, and v_i^-, $j \varepsilon A_\ell$, $i \varepsilon N_\ell$,

subject to:

$$\sum_{j \varepsilon W_i - S} f_j^r - \sum_{j \varepsilon V_i - S} f_j^r + v_i^+ - v_i^- = h_i^r - \Big(\sum_{j \varepsilon W_i \cap S} \bar{f}_j^r - \sum_{j \varepsilon V_i \cap S} \bar{f}_j^r\Big), \tag{24}$$

$$(i \varepsilon N_\ell) \;,$$

$$v_i^+ \geq 0, \; v_i^- \geq 0 \qquad (i \varepsilon N_\ell) \;, \tag{25}$$

$$f_j^r \geq 0 \qquad (j \varepsilon A_\ell) \;. \tag{26}$$

If $W^{\ell}_{SUB(r)} > 0$ for any ℓ, then

Set $K' \leftarrow K' + 1$

Let $\lambda_{ri}^{K'}$ be the optimal dual variables associated with constraint (24).

Go to Step 2.

Else, if $W^{\ell}_{SUB(r)} = 0$ for all ℓ, let $\bar{f}^r_j$, $j \varepsilon A_\ell$ be the optimal solution obtained for the problem (23) - (26) and go to Step 1B.

Step 1B (Find an optimal solution for subproblem r)

Set $K \leftarrow K + 1$.

Solve the traffic assignment problems for $\ell = 1,\ldots,L$:

$$\text{MINIMIZE} \quad Z^{K,\ell}_{SUB(r)} = \sum_{j \varepsilon A_\ell} \int_0^{f_j} C_j(x)dx \quad , \tag{27}$$

with respect to: f^r_j and f_j, $j \varepsilon A_\ell$.

subject to:

$$\sum_{j \varepsilon W_i - S} f^r_j - \sum_{j \varepsilon V_i - S} f^r_j = h^r_i - \Big(\sum_{j \varepsilon W_i \cap S} \bar{f}^r_j - \sum_{j \varepsilon V_i \cap S} \bar{f}^r_j\Big) \ , \tag{28}$$

$$(i \varepsilon N_\ell) \ ,$$

$$f_j = f^r_j + K^r_j, \qquad (j \varepsilon A_\ell) \ , \tag{29}$$

$$f^r_j \geq 0 \ , \qquad (j \varepsilon A_\ell) \ , \tag{30}$$

where K^r_j is a constant equal to $\sum_{i \neq r} f^i_j$.

Let $\bar{u}^K_{ri}$ be the dual variables associated with constraints (28) for subproblems r on iteration K. In order to determine $\bar{u}^K_{ri}$ uniquely, it is sufficient that for each N_ℓ a node $\bar{i}$ be selected and

$$\bar{u}^K_{r\bar{i}} = 0 \text{ for } \ell = 1,\ldots,L.$$

Compute:

$$Z^K_{SUB(r)} = \sum_{\ell=1}^{\ell=L} Z^{K,\ell}_{SUB(r)} \quad .$$

Set UBD $\leftarrow$ minimum $\{UBD, Z^K_{SUB(r)} + \sum_{j\varepsilon S} \int_0^{\bar{f}_j} C_j(x)dx\}$

$$\bar{M}^K = Z^K_{SUB(r)} - \sum_{i\varepsilon N} \bar{u}^K_{ir} (\sum_{j\varepsilon W_i \cap S} \bar{f}^r_j - \sum_{j\varepsilon V_i \cap S} \bar{f}^r_j)$$

If $\bar{Z}_M \geq UBD - \varepsilon_1$

or

$\left|\bar{u}^K_{ra} - \bar{u}^K_{rb} - C_j(\bar{f}_j)\right| \leq \varepsilon_2$ where the arc $j = (a, b)$ for all $j\varepsilon S$, go to Step 3.

Go to Step 2.

Step 2. (Optimize the master)

Solve the generalized Bender's master problem:

$$\text{MINIMIZE} \quad Z_M = \sum_{j\varepsilon S} \int_0^{f_j} C_j(x)dx + x_0 \;, \tag{31}$$

with respect to: f^r_j, f_j and x_0, $j\varepsilon S$,

subject to:

$$\sum_{i\varepsilon N} [h^r_i - (\sum_{j\varepsilon W_i \cap S} f^r_j - \sum_{j\varepsilon V_i \cap S} f^r_j)] = 0 \;, \tag{32}$$

$(\ell = 1,\ldots,L-1)$,

$$x_0 \geq \bar{M}^k + \sum_{i\varepsilon N} \bar{u}^k_{ri} (\sum_{j\varepsilon W_i \cap S} f^r_j - \sum_{j\varepsilon V_i \cap S} f^r_j) \;, \tag{33}$$

$(k = 1,\ldots,K')$,

$$\sum_{i \epsilon N} \bar{\lambda}^{k}_{ri} [h^r_i - (\sum_{j \epsilon W_i \cap S} f^r_j - \sum_{j \epsilon V_i \cap S} f^r_j)] \leq 0 \quad , \tag{34}$$

$$(k = 1,\ldots,K') \quad ,$$

$$f_j = f^r_j + K^r_j \qquad (j \epsilon S) \quad , \tag{35}$$

$$f^r_j \geq 0 \qquad (j \epsilon S) \quad , \tag{36}$$

and obtain $\bar{Z}_M$, $\bar{f}_j$ and $\bar{f}^r_j$ for $j \epsilon S$.

Go to Step 1A.

Step 3. (An optimal solution for subproblem r has been found)

Revise the current flow pattern $F = (f^1,\ldots,f^R)$ based on the last solution for f^r_j, $j \epsilon A-S$ in Step 1B and $\bar{f}^r_j$, $j \epsilon S$ in Step 2.

If the flow pattern F is simultaneously optimal for all subproblems r in Step 1B, terminate.

If r = R, set r = 1; else set r = r + 1.

UBD $\leftarrow \infty$

$\bar{Z}_M \leftarrow -\infty$

$K \leftarrow 0$

$K' \leftarrow 0$.

Go to Step 1A.

6.3.5 DISCUSSION

There is a great similarity between versions I and II of the geographic decomposition procedure. In point of fact, the feasibility testing procedures of Step 1A are identical; while if one were to apply the natural decomposition by commodities to the subproblem of Step 1B in version I, one would obtain the subproblem of Step 1B in version II. As we have shown in the previous discussion, the application of this same natural decomposition provedure to the master problem of Step 2 in version I results in a master problem that is identical to the master problem of Step 2 in version II. There are, however, distinctions that do remain between the two versions of the algorithm.

First, version I can be viewed as being a more flexible procedure, since the traffic assignment problem of Step 1B can be solved by methods other than that recommended by Nguyen [7], such as a modified version of the UTPS algorithm of UMTA, or the column generation procedure of Leventhal, Nemhauser, and Trotter [11]. Since there is very little computational evidence as to which of the many proposed algorithms is most efficient, this could prove to be a very useful property of version I. (The only large scale computational study known to the author is that of Nguyen [9], which proved inconclusive).

Second, version I also permits additional flexibility in the procedure that can be used to solve the master problem of Step 2. Although we have already shown that this problem can be totally decomposed by origin nodes, this might not prove to be the most desirable procedure. For instance, in a problem of more modest size, it may be possible to solve the master problem of Step 2 directly, possibly taking into account only the network structure of constraints (18), or the classical block angular structure with coupling equations given by (19), (20) and (21).

As a counterbalance to the flexibility of version I, version II possesses the very desirable property of working with only one commodity, defined by an origin node, at a time. This should permit more efficient utilization of computer memory and a reduced amount of information transfer between high speed memory and auxiliary storage. Thus, a conclusion as to which of the two versions is more efficient can only be ascertained by extensive empirical testing.

6.4 SUB-AREA FOCUSING

Dial [2] has posed the following problem in connection with a network design model for transportation planning. A small area of a large transportation network is extracted for intensive study. This area, often called the window, will be examined for possible changes in the infrastructure of the available transportation system. A key to an accurate model of these changes is the reaction of vehicles outside of the window to changes within. Equally important is the ability to expand the detail of the infrastructure within the window, while still maintaining a reasonable size for the overall problem. Current methodology calls for the aggregation of the network outside of the window, both to permit more detail within the window and to increase computational efficiency. As an alternative to this procedure, we now propose to use geographic decomposition.

The key to the use of geographic decomposition is the selection of the set of areas for the cut set S, so that the window becomes one of the subnetworks (N_ℓ, A_ℓ). If the original optimal solution for the problem is available, changes to only the window subnetwork can now be examined. Such changes will take the form of the addition or deletion of links. A first order approximation to the effect of such changes can be computed by iterating between the window subproblem in Step 1B and the master problem of Step 2 in version I of the geographic decomposition algorithm. Effectively, this procedure uses the dual solutions $\bar{u}_{ri}^K$ of the surrounding subproblems to model the effect on the window of shifts in trip routes within the surrounding areas. The approximation can be made more precise by periodically resolving the subnetwork problems for the surrounding areas in Step 1B of the algorithm, and thus generating new dual solutions $\bar{u}_{ri}^K$ for these areas.

Since the use of geographic decomposition involves only the repeated solution of the window subproblem and the master problem, the procedure should prove to be computationally efficient, and therefore, a reasonable alternative to the use of aggregation. Moreover, geographic decomposition does have the added benefit of the ability to easily refine its approximation, something that is more difficult with an aggregation procedure.

6.5 CONCLUSION

In this section, we have presented a decomposition procedure that should prove useful in the solution of very large transportation planning problems, and especially useful for the problem of sub-area focusing. In fact, the proposed methodology should find many other applications in the area of transportation modeling. For example, it requires only minor changes in our original traffic assignment formulation to compute a system optimal solution that obeys Wardrop's second principle, or to add the flexibility of elastic trip demand functions.

6.6 REFERENCES

1. A. M. Geoffrion, "Generalized Benders Decomposition," Journal of Optimization Theory and Applications, Vol 10, No. 4, p. 237-260, 1972.

2. R. B. Dial, "The Development of UMTA's Transportation Planning System," Presented at the 1974 Annual Meeting of the Highway Research Board, Washington DC, January 1974.

3. T. C. Hu, "A Decomposition Algorithm for Shortest Paths in a Network," Operations Research, Vol 16, p. 21-102, 1968.

4. T. C. Hu and W. T. Torres, "Shortcut in the Decomposition Algorithm for Shortest Paths in a Network," IBM Journal of Research and Development, Vol. 13, No. 4, July 1969.

5. Z. F. Lansdowne, "Analysis of Intercity Transport Improvements: Forecasting Demand and Evaluating User Benefits," Report No. RM-6255-DOT, The Rand Corp., May 1970.

6. R. E. Ramsburg and M. H. Aronoff, "Calculation by Digital Computer of Minimum Station-to-Station Distances in the United States Railroad Network," Department of Transportation Report DOT-P-6300.1, June 1972.

7. S. Nguyen, "An Algorithm for the Traffic Assignment Problem," Transportation Science, Vol. 8, p. 203-216, 1974.

8. J.D. Murchland, "Road Network Traffic Distribution in Equilibrium," presented at the conference "Mathematical Methods in the Economic Sciences," Mathematisches Forschungsingstitut, Oberwolfach, 1969.

9. S. Nguyen, "A Unified Approach to Equilibrium Methods for Traffic Assignment," Presented at the International Symposium on Traffic Equilibrium Methods, Montreal, CANADA, Novermber 1974.

10. E. R. Ruiter, "Network Equilibrium Capabilities for the UMTA Transportation Planning System," Presented at the International Symposium on Traffic Equilibrium Methods, Montreal, CANADA, November 1974.

11. T. Leventhal, G. Nemhauser, and L. Trotter, Jr., "A Column Generation Algorithm for Optimal Traffic Assignment," Transportation Science, Vol. 7, p. 168-176, 1973.

APPENDIX A

SELECTIVE REVIEW OF DECOMPOSITION METHODOLOGY

In this appendix, we will briefly review the decomposition methods used in the body of this report. This is by no means meant to be an exhaustive review. The interested reader who wishes to see a complete review of decomposition methodology is referred to the excellent book by Lasdon [1]. In addition there are the reviews by Balinski [2], Gomory [3] and Geoffrion [4]. More recently, Lasdon [5] has prepared a technical report that updates the previously mentioned text. Since the traffic assignment problem is a special type of network flow problem, the reader may also be interested in examining the survey by Bradley [6], which describes the state of the art in the solution of large network flow problems.

Three basic decomposition approaches are used in the algorithms presented in this paper. The first we classify as a price directed decomposition. The classic paper in this area is by Dantzig and Wolfe [7], which can be viewed as an extension of the work of Ford and Fulkerson [8] on multicommodity network flow problems. In a price directed decomposition, the subproblems communicate with the master problem by the mechanism of prices and proposals. The master problem sets prices, which in turn modify the objective function of the subproblems. The subproblem solutions are passed to the master problem which will then recompute the prices and the process is repeated. The natural decomposition used in the convex simplex method of Sections 2 and 3 and in the geographic decomposition of Section 6 are variants of this approach.

In Section 4 we introduce a decomposition scheme due to Steenbrink [9], [10]. This is probably most appropriately considered a form of price decomposition, although the analogy is far from perfect. In this procedure, a separate optimization problem was solved for each link, which created an objective function for a traffic assignment problem. If we associate the entire group of link optimization problems with a master problem, then the Steenbrink decomposition falls into the mold of a price directed approach.

The second basic decomposition approach can be classified as resource directed. In this approach, the master problem sends resources to the subproblem and the subproblem in turn computes the marginal value of the resources it has been allocated and returns this information to the master problem. Geoffrion's generalized Bender's decomposition [11] is an example of such a procedure. We have used this methodology in deriving the geographic decomposition of Section 6.

The third basic decomposition approach that we used was the optimization principle of dynamic programming (Bellman [12], [13]). This was used in Section 5 to decouple the investment staging problem, so that the multi-stage problem could be solved by solving a series of single stage problems.

In summary, we have used three basic decomposition approaches, price directed, resource directed, and dynamic programming, in developing the analysis presented in this report.

REFERENCES

1. L. S. Lasdon, Optimization Theory for Large Systems, Macmillan, New York NY, 1970.

2. M. L. Balinski, "On Some Decomposition Approaches in Linear Programming," Presented at the Conference on Recent Mathematical Advances in Operations Research, Engineering Summer Conferences, University of Michigan, June 1967.

3. R. E. Gomory, "Large and Non-Convex Problems in Linear Programming," Experimental Arithmetic, High Speed Computing and Mathematics, Proceedings of Symposium in Applied Mathematics, Vol. XV, American Mathematical Society, 1963.

4. A. Geoffrion, "Elements of Large-Scale Mathematical Programming," Management Science, Vol. 16, No. 11, p. 652-691, 1970.

5. L. S. Lasdon, "A Survey of Large Scale Mathematical Programming," Technical Memo No. 349, Department of Operations Research, School of Management, Case Western Reserve University, Cleveland OH, December 1974.

6. G. H. Bradley, "Survey of Deterministic Networks," AIIE Transactions, Vol. 7, p. 222-234, 1975.

7. G. B. Dantzig and P. Wolfe, "Decomposition Principle for Linear Programs," Operations Research, Vol 8, p. 101-111, 1960.

8. L. R. Ford, Jr., and D. R. Fulkerson, "A Suggested Computation for Maximal Multi-Commodity Network Flows," Management Science, Vol 5, p. 97-101, 1958.

9. P. A. Steenbrink, Optimization of Transport Networks, Wiley, New York NY, 1974.

10. P. A. Steenbrink, "Transport Network Optimization in the Dutch Integral Transportation Study," Transportation Research, Vol. 8, p. 11-27, 1974.

11. A. M. Geoffrion, "Generalized Benders Decomposition," Journal of Optimization Theory and Applications, Vol. 10, No. 4, p. 237-260, 1972.

12. R. Bellman, Dynamic Programming, Princeton University Press, Princeton NJ, 1957.

13. R. Bellman and S. E. Dreyfus, Applied Dynamic Programming, Princeton University Press, Princeton NJ, 1967.

APPENDIX B

REPORT OF INVENTIONS

The following features are regarded as innovations and improvements.

1. Convex Generalized Upper Bounding to Solve the Traffic Assignment Problem (see pages 23-24).

2. The Use of the Frank-Wolfe and Convex Simplex Algorithms for the Elastic Demand Traffic Assignment Problem (see pages 31-35).

3. The Solution of a Broad Class of Network Design Problems by the use of a Traffic Assignment Algorithm and a Language Multiplier Technique (see pages 44-69).

4. The Lexico-Graphic Objective Function was Introduced as a Technique to Solve the Problem of the Optimal Staging of Investments over time. (see pages 76-89).

5. Geographic Decomposition Applied to the Traffic Assignment and Sub-Area Focusing Problems (see pages 92-112).

The innovations of "3" and "4" have been implemented in a computer code.