REPORT NO. DOT-TSC-RSPA-79-7

HYBRID OPTIMIZATION IN URBAN TRAFFIC NETWORKS

Han-Ngee Tan
Stanley B. Gershwin
Michael Athans

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Laboratory for Information and Decision Systems
Cambridge MA 02139

APRIL 1979

FINAL REPORT

DOCUMENT IS AVAILABLE TO THE PUBLIC
THROUGH THE NATIONAL TECHNICAL
INFORMATION SERVICE, SPRINGFIELD,
VIRGINIA 22161

Prepared for

U.S. DEPARTMENT OF TRANSPORTATION
RESEARCH AND SPECIAL PROGRAMS ADMINISTRATION
Office of Transportation Programs Bureau
Office of Systems Engineering
Washington DC 20590

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT-TSC-RSPA-79-7		

4. Title and Subtitle	5. Report Date
HYBRID OPTIMIZATION IN URBAN TRAFFIC NETWORKS	April 1979
	6. Performing Organization Code
7. Author(s) Han Ngee Tan, Stanley B. Gershwin, Michael Athans	8. Performing Organization Report No. DOT-TSC-RSPA-79-7
9. Performing Organization Name and Address Massachusetts Institute of Technology* Laboratory for Information and Decision Systems Cambridge MA 02139	10. Work Unit No. (TRAIS) RS905/R9516 11. Contract or Grant No. DOT-TSC-1456
12. Sponsoring Agency Name and Address U.S. Department of Transportation Research and Special Programs Administration Office of Transportation Programs Bureau Office of Systems Engineering Washington DC 20590	13. Type of Report and Period Covered FINAL REPORT 3/13/78 - 3/14/79 14. Sponsoring Agency Code

15. Supplementary Notes

*Under Contract to: U.S. Department of Transportation, Research and Special Programs Administration, Transportation Systems Center, Cambridge MA 02142

16. Abstract

The Hybrid Optimization Problem is formulated to provide a general theoretical framework for the analysis of a class of traffic control problems which takes into account the role of individual drivers as independent decisionmakers.

Different behavioral models for flow distribution are examined. Necessary conditions for this problem are derived, and a physical interpretation of these conditions is provided. Possible directions for the development of algorithms applicable for solving large-scale Hybrid Optimization Problems are proposed. A procedure for computing the upper and lower bounds of the optimal cost of the Hybrid Optimization Problem is outlines.

17. Key Words		18. Distribution Statement
Urban Traffic Traffic Newtworks User Optimization System Optimization Hybrid Optimization	Energy Optimization Math. Programming Driver Behavioral Models Traffic Control Strategies	DOCUMENT IS AVAILABLE TO THE PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
UNCLASSIFIED	UNCLASSIFIED	124	

Form DOT F 1700.7 (8-72)

Reproduction of completed page authorized

PREFACE

This study has been performed under the Transportation Advanced Research Program (TARP) of the Research and Special Programs Administration (RSPA), U.S. Department of Transportation (DOT), which is aimed at investigating the applicability of advanced techniques to large-scale transportation systems.

This report is based in part on the S.M. thesis of Han-Ngee Tan, "Hybrid Optimization in Traffic Networks," submitted to the M.I.T. Department of Electrical Engineering and Computer Sciences in January 1979.

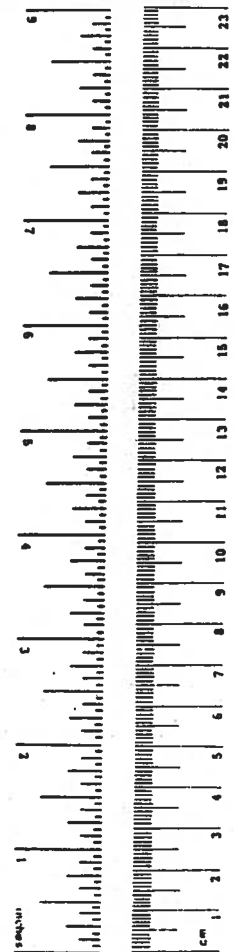
The research was conducted at the M.I.T. Laboratory for Information and Decision Systems (formerly, the Electronic Systems Laboratory), and was supported by DOT under contract DOT-TSC 1456. It was monitored by the Transportation Systems Center (TSC). We thank Diarmuid O'Mathuna of TSC and Robert Crosby of DOT/RSPA for their helpful criticism, considered remarks, and welcome encouragement.

We also express our gratitude to Leni Gross, Margaret Flaherty and Barbara Peacock for the outstanding typing, and to Arthur Giordani for the excellent drafting.

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

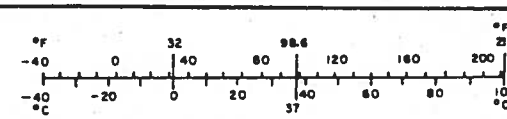


TABLE OF CONTENTS

TABLE OF CONTENTS (Contd.)

TABLE OF CONTENTS (Concl'd.)

ILLUSTRATIONS

TABLES

1. INTRODUCTION

1.1 Objective

Recently, there has been a considerable growth of interest in the solution of automobile transportation problems using modern control theory and system methodologies [1]. This reflects an increasing concern with urban traffic congestion which has become one of the most difficult and urgent problems facing our society. Environmental concerns and intense competition for limited resources make it imperative that more efficient use be made of the existing transportation infrastructures, such as the highway systems and the urban street networks, and of vehicles. Conventionally, the objective in most transportation control systems has been to ensure smooth vehicular traffic flow. There is a growing consensus that consideration must also be given to attaining broader national goals such as energy conservation.

The purpose of this report is to demonstrate the relationship between two important problems in automobile transportation: signal optimization, the choice of traffic signal control settings; and assignment, the distribution of vehicles in a network. It is clear that these problems affect one another and that they should be considered together. This report also demonstrates how properly treating them together has important payoffs in reducing congestion and energy consumption.

1.2 Background

The problem of real-time computer control of a transportation system can be viewed as a large-scale stochastic dynamic optimal control problem [2], [3], [4]. Such problems cannot be solved exactly, and some approximate solutions must be sought. For example, in a study of dynamic stochastic control of freeway corridor systems [2], the problem is decomposed into estimation, incident detection, and control activities. Control actions are then further subdivided into static and dynamic strategies. Static network optimization is shown to play an important role in this type of real-time control strategy.

Static approximation of network parameters and traffic flow variables also

play an important role in area traffic control systems in urban street networks. Most area traffic control systems use fixed-time signal plans because these strategies are efficient and easy to implement [5-9]. Fixed-time control strategies are based on observed daily patterns of traffic flows. Given any traffic situation characterized by flow volumes, link capacities, and vehicular arrival rates which are assumed approximately constant in time, a set of traffic control variables such as cycle times, green splits, and offsets which are optimal in some sense, can be calculated by a signal-optimizing program: e.g., TRANSYT [10], MITROP [11], SIGOP [7], [12]. Fixed-time signal plans are calculated in advance for a number of characteristic situations, such as morning and evening rush hours, and are implemented when such a situation arises.

In all the control strategies considered, the role of drivers as independent decision-makers in choosing among different available routes has not been incorporated in the process of analysis and synthesis. A remark due to Beckmann provides motivation for the research presented here [13]:

> ...a closer view of affairs shows that even in the most perfectly planned system the public retains certain freedoms, and that it may choose to ignore the intentions of the planners and play games of its own. Transportation analysis cannot afford to lose sight of how people use transportation.

Thus, there is a need to incorporate a behavioral model of drivers' route choice in the design, operation, and assessment phase of a transportation system. In the above calculations of optimal fixed-time signal plans, it is implicitly assumed that the route choice of the drivers is fixed and independent of the control parameters. In other words, it is assumed that the traffic volume on any particular link in the network is constant, regardless of the level of service offered by that link. This assumption is false since an individual driver cannot be prevented from taking an alternate route which may have been made more desirable (i.e., faster) by the implementation of a new control policy. In fact, the redistributional effects of traffic resulting from the implementation of an area traffic control policy have been studied and confirmed in a series of field experiments conducted in the city of Glasgow. See [14] and references cited therein.

It is observed [14], [15] that the new traffic pattern induced by some

"optimal" traffic control policy destroys the original optimality. It thus seems desirable to reoptimize periodically the controls based on new survey information on the traffic distribution [9], [14], [15]. However, this process of updating controls has seldom been carried out more than once or twice in practice due to the amount of effort and resources involved in the surveys [9].

In this report we consider the role of the driver's route choice behavior in the context of the steady-state network traffic control problem, the importance of which cannot be overemphasized. We attempt to provide some answers to the following questions:

Given the fact that the system has little or no control over the route selection decisions of individual drivers, how can it achieve a flow distribution which is optimal from the system's point of view, using the available controls? To what extent can control variables be used to influence drivers' route choice by exploiting the interdependence between the signal timing plan and the flow pattern? Given all the resources and effort, does the iterative reoptimization procedure mentioned in [14] lead to an optimal solution? Or more generally, given a certain predictive model of drivers' route selection behavior, how should one go about choosing a set of controls which, together with the eventual induced traffic pattern, is optimal with respect to a certain system cost criterion? We believe that this is a class of problem of fundamental importance in the various phases of decision making in transportation systems.

This is referred to as the Hybrid Optimization Problem (HOP) in this report. It has the following essential features.

a) The objectives of the traffic authority and the drivers are different.

b) On the system level, the problem for the traffic authority is to minimize some overall cost in the network, e.g., total travel time or total fuel consumption.

c) On the other hand, the individual driver wishes to minimize his trip cost in traveling through the network.

d) Individual drivers are independent decision-makers. This means that it is beyond the power of the traffic authority to specify traffic flows on all

links. The capability of the traffic authority is limited to the command of traffic control devices only. In most cases, the capability of the traffic authority is further restrained because practical limitations dictate that the traffic authority can exercise control only over a subset of the network.

1.3 Literature Survey

In the literature of transportation system planning, the problems of flow assignment and static control optimization are treated separately and are fairly well understood in terms of analysis and computation. Traffic flow assignment is the problem of computing flow configurations for a fixed set of network parameters, including control values such as signal settings and ramp metering rates, according to the two principles enunciated by Wardrop [16] on the distribution of vehicular traffic in a network. The static control optimization problem chooses appropriate values for control variables, assuming link flows are fixed and independent of the control variables.

1.3.1 Assignment

Traffic assignment is an indispensable part in the various complex phases of the transportation planning process [17], and has been an area of extensive research [18], [19], [20]. In system optimization, the traffic volume on any particular link in the network is assigned by a central traffic authority so that a certain system-wide cost is minimized. The resultant flow pattern has been referred to by different names in the literature: system-optimizing flow [18], efficiency flow [19], normative flow [20]. System-optimizing flow represents an optimal allocation of traffic and as such it should serve as a goal for any traffic control scheme. Studies on the properties of system-optimizing flow (e.g., existence and uniqueness) and computation methods can be found in many other disciplines also. This type of problem has been broadly classified as multicommodity minimum cost network flow problem (MFP) in the literature of management science and operations research [21]. A special class of problems in MFP with nonlinear link cost function has the same structure as the system optimization problem in traffic engineering. Research on computer network communication is also of the same nature [22], [23], [24]. A general discussion and analysis on system optimization can be found in a number of sources [18],

[19], [20]. Assad [21] presents an excellent survey of the literature to date on computation methods for MFPs. The propagation of perturbations of the system optimizing flow pattern due to disturbances has been investigated for a special class of networks [4].

The user optimization formulation of traffic assignment attempts to predict the flow distribution when a set of network parameters, including the controllable ones such as signal settings, and origin-destination (O-D) input flows are specified. This is a frequently encountered, critical problem in the planning and assessment phases in traffic engineering. The basic assumption is that each driver chooses his path in such a way as to minimize his own travel cost. Travel time is commonly used to represent travel cost [17], [25], [26]. Wardrop's first principle, or the principle of equal travel time, characterizes the equilibrium flow distribution which is the aggregate result of individual decisions. At an equilibrium, no single driver can reduce his own cost by unilaterally choosing an alternative route in the network. In the literature, terms such as equilibrium flow [19], descriptive flow [20], and user-optimized flow [18] are used interchangeably. In most of the large-scale transportation studies, a variety of heuristic methods have been traditionally used to determine equilibrium flow patterns. They include the "all or nothing" method [27], the "incremental loading method" [28]. The "all or nothing" method, applicable only to the case of constant-link cost (i.e., link cost per unit flow is independent of link flow) without link capacity constraints, assigns all traffic demand along the shortest (i.e., least costly) path for each origin-destination pair [27]. Modifications of the "all or nothing" method for the more realistic case of nonlinear link cost involve incrementally loading the network on the current shortest route and iteratively adjusting the link cost [28]. Heuristic rules are used to reallocate flows to "balance the system" [28]. Although this class of solution procedures has by far dominated the field in actual applications due to its conceptual simplicity, it is ad-hoc in nature and it has convergence and stability problems [29].

Attempts to formulate the equilibrium problem (the calculation of a user-optimized assignment) as an optimization problem are motivated by two different observations. Equilibria in many other fields, e.g., electrical circuit

theory [30], have been successfully formulated as solutions to some extremal problems; the necessary conditions for the multicommodity minimum cost flow problem bear a remarkable similarity to the equilibrium conditions stated in Wardrop's principle of equal travel time. Thus, it has been shown [18], [20], [31] that under certain mild restrictions on the link cost function, the determination of the equilibrium flow pattern is equivalent to the solution of a convex minimization problem. This equivalent minimization problem is in fact a multicommodity minimum cost flow problem, which has been discussed above.

Applying recent results in mathematical programming theory, coupled with successful exploitation of the special structures of the problem, several efficient algorithms, with the capability of solving medium-to-large scale network equilibrium problems, have been developed in recent years. See, for example, Nguyen [32], Leventhal [33], and LeBlanc [34]. More recently, application of nonlinear complementarity theory to traffic equilibrium problems represents a new and promising area of research [35]. Hall [36] has investigated a crucial, if not somewhat neglected, question in the study of traffic equilibrium concerning the sensitivity of equilibrium flow to variation of input flows. However, the area of sensitivity due to variation of control variables such as signal settings is still an unexplored area of research.

1.3.2 Signal Setting

Urban traffic congestion has motivated a great deal of research devoted to the synthesis of effective control of traffic signals. Fixed-time signal control policies have been widely used due to their simplicity in implementation and satisfactory performance [5], [6], [7], [8], [9]. A large number of mathematical techniques with traffic models of different degrees of sophistication have been developed [7]. May [7] provides an excellent survey on the fixed-time signal optimization methods that were developed and implemented in various parts of the world. Among the better known methods (e.g., TRANSYT [10], SIGOP [1], MITROP [11]) the optimization of control variables (which include green split, cycle time, and offset for coordination of neighboring signals) can be carried out for networks having on the order of 100 nodes.

1.3.3 Redistributional Effects

Until very recently, there has been very little attention paid to the significance of the relationship between traffic assignment and signal optimization, which have been studied as two mutually independent activities in traffic engineering. Recent investigations have led to the recognition of the interdependence of signal timing plans and network flow patterns [14], [15], [37], [38], [39]. For example, it has been established that "redistributional effects of an area traffic control policy are a possibility (that) can hardly be denined." [14]. It is also observed [14], [15] that in some cases the new traffic pattern indirectly induced by some "optimal" traffic control policy destroys the original optimality. There are several important implications as a consequence of these new findings.

It has been concluded [14] that the relative merits of alternative signal timing plans should be evaluated together with their redistributional effects. In addition, it seems desirable to reoptimize periodically the controls based on new information on the traffic distribution [15], [14]. Allsop [37], recognizing the interdependence between signal timing plans and flow patterns, suggested the idea of using control schemes to influence drivers' route choices. However, there exists no theoretical formulation for this problem. Such a formulation is the main contribution of the research reported here.

1.4 Contritubions and Main Results of This Report

The main contribution of this report is the introduction of the concept of explicitly incorporating the role of individual drivers as independent decision-makers in optimization of traffic control by the traffic authority. The Hybrid Optimization Problem (HOP) formulation provides a general theoretical framework for the study of this problem. In traffic engineering, the problems of system optimization, control optimization, and assignment have been studied independently of one another. The formulation of the Hybrid Optimization Problem can be viewed as a unified approach which combines these problems in transportation systems.

A Generalized System Optimization Problem is introduced, which extends the notion of system optimization to the case where the traffic authority

treats both traffic control parameters and flows simultaneously as decision variables in order to optimize some system-wide cost.

We have found a way to state the user optimization principle as a set of equations and inequalities. This is essential for the HOP formulation.

Our studies also show that the system cost, such as total travel time and total fuel consumption, may not be differentiable at some points in the space of traffic control variables when traffic is distributed according to user optimization. This is significant because it can make the construction of algorithms difficult. For example, numerical differentiation of the cost with respect to traffic control variables must be done with care in the neighborhood of a point where the derivative fails to exist.

Necessary conditions for the optimal solution of a specific HOP which assumes user optimization as the flow distribution principle, are derived. A physical interpretation of these conditions is summarized in the Extended Equilibrium Principle, which bears a remarkable similarity to the Equal Travel Time and Equal Marginal Cost Principle [19] of user and system optimizations.

An important contribution is a study of the iterative reoptimization procedure, which consists of successive alternations between control optimization and equilibrium assignment programs. This procedure has been proposed [14], [15], [27] as a solution algorithm for the problem addressed here, and which has also been described as a control strategy [9], [14]. We show that it leads to incorrect solutions.

Practical difficulties in solving large-scale Hybrid Optimization Problems are discussed. These are the necessity of a priori enumeration of all paths and the high dimensionality of the resulting optimization problem.

1.5 Outline

Most of the notation and terms used in this report are defined in Section 2. We first introduce A Generalized System Optimization Problem (GSOP) which extends the conventional system optimization problem to the case where the central traffic authority chooses controllable network parameters, such as signal settings, in addition to assigning link flows. We discuss various phys-

ical considerations and limitations imposed on the traffic authority in the real-life situations. The Hybrid Optimization Problem (HOP) is then formulated and stated in general mathematical terms.

Section 3 presents a discussion on the different behavioral models on the route selection process and the resulting overall network flow distribution principles. We discuss the flow distribution according to Wardrop's First Principle in considerable detail since this is used primarily as a flow distribution principle in this study. However, it should be emphasized that the HOP formulation presented in this report is not restricted by any particular flow distribution principle. A new mathematical characterization of the user-optimized flow pattern is presented. This characterization, which is a set of equalities and inequalities,can be incorporated as constraints in an optimization problem. (An alternative flow distribution model is also discussed.)

In Section 4, we devote our attention to specific HOP which assumes user optimization as the flow distribution principle. Necessary conditions for the optimal solution of the HOP are derived. A physical interpretation of these conditions is summarized in the Extended Equilibrium Principle, which bears a remarkable similarity to the Equal Travel Time and Equal Marginal Cost Principles of user and system optimization in traffic engineering.

To demonstrate and verify the concepts and formulation of the HOP and the Extended Equilibrium Principle, numerical examples using small networks are discussed in Section 5. Two different approaches are used in solution strategies. The solutions obtained are described and compared, and the two solution approaches are evaluated.

In Section 6, we address the issue of solving HOPs involving large-scale networks. Some practical difficulties are pointed out. They include the generation of considerable path information and the large number of variables. Two algorithms are proposed in an attempt to alleviate these problems. In these proposed algorithms, we show how some of the special structures of the HOP may be exploited; however, there are still some problems associated with these algorithms. In one algorithm, we are able to avoid the generation of all the paths a priori. Instead, paths are generated when required. Although we are able to reduce the size of the problem, it is still too cumbersome for

large networks to be solved on a computer. In another algorithm, we encounter the nondifferentiability of the cost function with respect to the independent variables. We also provide a procedure for computing the upper and lower bounds of the optimal cost of the HOP. Based on previous results in the literature of network flow patterns, we present a conjecture on these bounds, and an approximate solution to the HOP is proposed.

There are several outstanding problems requiring further research which are discussed in Section 7. Several problems in traffic engineering to which the HOP formulation appears to be applicable are identified.

2. GENERALIZED SYSTEM OPTIMIZATION AND HYBRID OPTIMIZATION PROBLEMS

The main purpose of this section is to provide a precise mathematical formulation of the Hybrid Optimization Problem (HOP). To do this, we define the various terminologies and notations to be used throughout this report in section 2.1. A list of symbols used in this report is provided in Appendix A. In section 2.2, we discuss a related problem, the Generalized System Optimization Problem (GSOP). GSOP extends the conventional System Optimization Problem (SOP) to the case where the central traffic authority chooses controllable network parameters, such as signal settings, in addition to assigning link flows. It chooses these parameters and flows to optimize a certain system-wide cost. GSOP is, of course, a highly idealized abstraction of the real-life traffic control problem. In section 2.3, we discuss various practical limitations imposed on the traffic authority. Taking into consideration these additional constraints, we formulate the Hybrid Optimization Problem in general and precise mathematical terms amenable for analysis in section 2.4.

2.1 Terminology and Notation

We represent a transportation network by a directed graph G. G consists of a pair (N,L), where N is a collection of elements which will be called "nodes", and L is a set of elements called "links" or "arcs." A node in N represents an intersection of streets in the network. Every road or street segment is represented by a link in L, which can also be viewed as an ordered pair of nodes in N.

For example, link i, which goes from node $\alpha(i)$ to node $\beta(i)$ can be represented by the ordered pair $[\alpha(i), \beta(i)]$. Nodes $\alpha(i)$ and $\beta(i)$ are called the initial and terminal nodes of link i, respectively. In this report, we use the notation $|S|$ for the total number of elements in the set S. Let NN = $|N|$; i.e., the total number of nodes and NL = $|L|$; i.e., the total number of links. Associated with a network, there is a node-arc incidence matrix, $\underline{A}$,* of dimension

* Underlined upper-case and lower-case letters represent matrices and column vectors, respectively.

NN by NL. The (i,j) element of matrix $\underline{A}$, a_{ij}, is defined as follows:

$$a_{ij} = \begin{cases} +1 \text{ if node i is the initial node of link j;} \\ -1 \text{ if node i is the terminal node of link j;} \\ 0 \text{ otherwise.} \end{cases} \qquad (2.1)$$

The node-arc incidence matrix contains all the topological information of the network under consideration.

For every link i in L, there are several quantities of interest: f_i, e_i, t_i. The amount of traffic flow (in veh/hr) on link i is denoted by f_i. The quantity t_i represents the total amount of time each vehicle spends in traveling through link i; t_i includes the transversal time and possibly the waiting time due to queuing at a signalized junction. Similarly, e_i is the total amount of fuel in gal/veh each vehicle consumes in traveling through link i. We let $\underline{f} \varepsilon R^{NL}$, $\underline{e} \varepsilon R^{NL}$, $\underline{t} \varepsilon R^{NL}$* be the link flow, link fuel consumption, and link time vectors, respectively.

Suppose there are K O-D (origin-destination) pairs in the network under consideration. For the k^{th} O-D pair, let H^k be the total traffic requirement, or the input flow, to go from the origin node, O^k, to the destination node, D^k. We denote the set of all loop-free paths connecting the k^{th} O-D pair by P_k. A loop-free path in P_k is defined as a sequence of nodes in N leading from O^k to D^k with the restriction that every node in the sequence is distinct. For each path i in P_k, we define R_i^k as a set of links on the i^{th} path of the K^{th} O-D pair. Let $\underline{T}^k$, $\underline{h}^k$ (each of dimension $|P_k|$) be the path time and path flow vector for the k^{th} O-D pair. T_i^k and h_i^k are, respectively, the path time and path flow on the i^{th} path of the k^{th} O-D pair. From the path flow vectors of all O-D pairs, $\underline{h}^k$, k=1, ..., K, we form a vector $\underline{h}$, which is of dimension $\sum_{k=1}^{K} |P_k|$. Associated with the k^{th} O-D pair, we designate $\underline{f}^k \varepsilon R^{NL}$ as the link flow vector of the k^{th} O-D pair. The i^{th} element of $\underline{f}^k$, denoted by f_i^k, is the amount of the traffic flow of the k^{th} O-D pair on link i. The link flow vector, $\underline{f}^k$, and path flow vector, $\underline{h}^k$, are related by the following equation:

---

* R^a represents the a-dimensional Euclidean space.

$$\underline{f}^k = \underline{B}^k \underline{h}^k , \qquad (2.2)$$

where $\underline{B}^k$, of dimension NL by $|P_k|$, is the <u>link-path incidence matrix</u> for the k^{th} O-D pair. The (i,j) element of $\underline{B}^k$, b_{ij}^k, is defined as follows:

$$b_{ij}^k = \begin{cases} 1 \text{ if link } i \text{ lies on the } j^{th} \text{ path of the } k^{th} \text{ O-D pair;} \\ 0 \text{ otherwise.} \end{cases} \qquad (2.3)$$

With this definition, path and link time vector can also be related by

$$\tau_j^k = \sum_{i=1}^{NL} b_{ij}^k t_i ,$$

or (2.4)

$$\underline{\tau}^k = \underline{B}^{k'} \underline{t} ,$$

where $\underline{B}^{k'}$ denotes the transpose of matrix $\underline{B}^k$.

The total link flow vector, $\underline{f}$, is the sum of the link flow vectors of all O-D pairs.

$$\underline{f} = \sum_{k=1}^{K} \underline{f}^k = \sum_{k=1}^{K} \underline{B}^k \underline{h}^k , \qquad (2.5)$$

or, written component-wise:

$$f_i = \sum_{k=1}^{K} f_i^k .$$

The following example is used to clarify the difference between link flows and path flows. Consider the network shown in Figure 2.1. A list of paths for each O-D pair is shown in Table 2.1. We observe that there are two paths for the first O-D pair going from node 1 to node 5. The path flows h_1^1, h_2^1 are, respectively, the flows along the first path, (1,4,5), and the second path, (1,3,4,5), of the first O-D pair. From Figure 2.1, it can be seen that h_1^1 passes through links 3,5 and h_2^1 passes through links 1,4,5. The arc-chain incidence matrix for

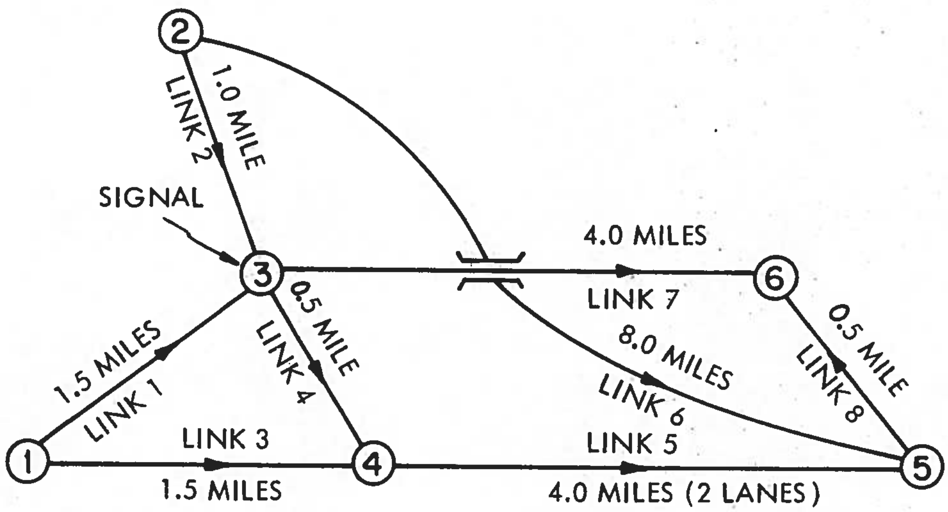


Figure 2.1 Network for Examples 2 and 3

the first O-D pair, denoted by $\underline{B}^1$, can then be written according to the definition (2.3)

$$\underline{B}^1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} .$$

Of the two paths of the first O-D pair, only the second passes through links 1 and 4. Therefore, the amount of flow between the first O-D pair on links 1 and 4, denoted by f_1^1 and f_4^1, are both equal to h_2^1. On the other hand, both h_1^1 and h_2^1 pass through link 5, and hence, f_5^1 is equal to $h_1^1 + h_2^1$. The link flow vectors from each O-D pair, $\underline{f}^k$, are listed in Table 2.2. They can also be computed using (2.2). The total amount of flow on any particular link i, f_i, is the total amount of path flow of all O-D pairs passing through link i. For example, it can be observed from Table 2.1 that h_1^1, h_2^1, h_1^2, h_3^2, h_1^3, h_1^4 all pass through link 5. Hence,

$$f_5 = h_1^1 + h_2^1 + h_1^2 + h_3^2 + h_1^3 + h_1^4 .$$

The total flow on all links are listed in the last column of Table 2.2.

2.1.1 Feasible Flows

A basic principle governing network flows is the Principle of Conservation of flow: in steady state, the total amount of flow entering a node equals the total amount of flow leaving the node. Mathematically, for each O-D pair,

$$\underline{Af}^k = \underline{w}^k, \; k=1,\ldots,K , \tag{2.6}$$

where $\underline{w}^k$ is the traffic requirement vector for the k^{th} O-D pair with the i^{th} component defined as follows:

TABLE 2.1 All Available Paths

O-D Pair k	Origin Node O^k	Dest. Node D^k	Path Flow h_i^k	Path	Link along Path
1	1	5	h_1^1 h_2^1	1 = (1, 4, 5) 2 = (1,3,4,5)	(3, 5) (1,4,5)
2	1	6	h_1^2 h_2^2 h_3^2	1 = (1,4,5,6) 2 = (1, 3, 6) 3 = (1,3,4,5,6)	(3,5,8) (1, 7) (1,4,5,8)
3	2	5	h_1^3 h_2^3	1 = (2,3,4,5) 2 = (2, 5)	(2,4,5) (6)
4	2	6	h_1^4 h_2^4 h_3^4	1 = (2,3,4,5,6) 2 = (2, 5, 6) 3 = (2, 3, 6)	(2,4,5,8) (6, 8) (2, 7)

TABLE 2.2 <u>Link Flows</u>

Link	Initial Node	Terminal Node	Partial Flow Due To Each O-D Pair				Total Flow
i	(i)	(i)	f_i^1	f_i^2	f_i^3	f_i^4	f_i
1	1	3	h_2^1	$h_2^2 + h_3^2$	0	0	$h_2^1 + h_2^2 + h_3^2$
2	2	3	0	0	h_1^3	$h_1^4 + h_3^4$	$h_1^3 + h_1^4 + h_3^4$
3	1	4	h_1^1	h_1^2	0	0	$h_1^1 + h_1^2$
4	3	4	h_2^1	h_3^2	h_1^3	h_1^4	$h_2^1 + h_3^2 + h_1^3 + h_1^4$
5	4	5	$h_1^1 + h_2^1$	$h_1^2 + h_3^2$	h_1^3	h_1^4	$h_1^1+h_2^1+h_1^2+h_3^2+h_1^3+h_1^4$
6	2	5	0	0	h_2^3	h_2^4	$h_2^3 + h_2^4$
7	3	6	0	h_2^2	0	h_3^4	$h_2^2 + h_3^4$
8	5	6	0	$h_1^2 + h_3^2$	0	$h_1^4 + h_2^4$	$h_1^2 + h_3^2 + h_1^4 + h_2^4$

$$w_i^k = \begin{cases} H^k \text{ if node } i \text{ is the origin node } O^k; \\ -H^k \text{ if node } i \text{ is the destination node } D^k; \\ 0 \quad \text{otherwise.} \end{cases} \tag{2.7}$$

Using Eq. (2.5), conservation of total link flow can be written as

$$\underline{Af} = \underline{w} \quad , \tag{2.8}$$

where

$$\underline{w} = \sum_{k=1}^{K} \underline{w}^k . \tag{2.9}$$

Conservation of flows may also be written in terms of path flows:

$$\sum_{i \varepsilon P_k} h_i^k = H^k . \tag{2.10}$$

In addition, link and path flows are required to be non-negative:

$$\underline{f} \geq 0 \;, \tag{2.11}$$

$$\underline{f}^k \geq 0 \;, \quad k = 1, \ldots, K \;, \tag{2.12}$$

$$\underline{h}^k \geq 0 \;, \quad k = 1, \ldots, K \;. \tag{2.13}$$

Finite link capacity also requires that the total amount of flow on every link be less than or equal to the saturation flow,

$$\underline{f} \leq \underline{c} \;, \tag{2.14}$$

where $\underline{c} \varepsilon R^{NL}$ is a vector of link saturation flows, or link capacities. Equations (2.5) and (2.14) imply

$$\sum_{k=1}^{K} \underline{f}^k \leq \underline{c} \;, \tag{2.15}$$

$$\sum_{k=1}^{K} \underline{B}^k \underline{h}^k \leq \underline{c} . \tag{2.16}$$

Definition: A set of path flow vectors $\{\underline{h}^k\}$ is a feasible path flow if and only if $\{\underline{h}^k\}$ satisfies Eqs. (2.10), (2.13), and (2.16).

Definition: $\{\underline{f}^k\}$ is said to be a set of feasible link flow vectors for all O-D pairs if and only if Eqs. (2.6), (2.12) and (2.15) are satisfied. The resultant total link flow, $\underline{f}$, is called a feasible total link flow vector.

2.1.2 Control Parameters

Let $\underline{g}\varepsilon R^U$ be a vector of control parameters which includes green splits, cycle time, offsets, and ramp-metering rates in the most general formulation. Ramp metering rates are excluded from consideration here since in this study we focus our attention on urban grid network. Furthermore, we assume a fixed-cycle time of one minute and ignore the effect of offsets. Thus, we consider only the green splits as control parameters. This simplifying approach is taken to avoid mathematical complications without losing the important conceptual qualitative features of the problem. G denotes the set of feasible controls which defines the physical constraints on $\underline{g}$. Specifically, g_i, which is the green split on the i^{th} signal expressed as a fraction of a cycle, must satisfy

$$0 \leq g_i \leq 0.9 \quad i = 1, \dots U, \tag{2.17}$$

The upper limit of 0.9, instead of 1.0, accounts for the lost time in every cycle due to reaction time at the beginning of the green phase [11]. Thus,

$$G = \{\underline{g}\varepsilon R^U \mid 0 \leq g_i \leq 0.9; \; i = 1, \dots, U\}. \tag{2.18}$$

Throughout this report, we assume that there are two competing streams of traffic entering each signalized junction i. The green split facing one of the streams is g_i while the green split facing the opposing stream is $(0.9-g_i)$. Which split is called g_i and which is called $0.9-g_i$ is arbitrary.

2.1.3 Cost Function

The appropriate cost function often chosen in transportation system decisionmaking is total travel time or total fuel consumption [11], [15]. In the case of travel-time minimization, the cost can be written as

$$J = \sum_{i=1}^{NL} f_i t_i \; . \tag{2.19}$$

If the concern is total fuel consumption, then the cost is

$$J = \sum_{i=1}^{L} f_i e_i \quad . \tag{2.20}$$

The average transit time on an unsignalized link is modeled by the conventional fourth-order polynomial in link flow

$$t_i^s = t_i^o \left| 1 + 0.15 \left(\frac{f_i}{c_i}\right)^4 \right| \quad , \tag{2.21}$$

where t_i^o is the average free flow speed on link i. This model has been used by the Federal Highway Administration [40] and widely adopted in various transportation studies [17], [32], [40], [41].

For links entering a signalized intersection, the additional component of average waiting time is modeled by Webster formula [15]

$$t_i^w(f_i, \underline{g}) = \begin{cases} 0.45 \left(\dfrac{CYC(1-g_j)^2 c_i}{c_i - f_i} + \dfrac{f_i}{g_j c_i (g_j c_i - f_i)} \right) ; & \text{for } g_j c_i > f_i , \\ \infty & \text{for } g_j c_i \leq f_i , \end{cases} \tag{2.22}$$

where g_j is the green split facing link i; CYC is the cycle time in hours; c_i is the link capacity.

Thus, the total time spent on link i is

$$t_i(f_i, \underline{g}) = \begin{cases} t_i^s(f_i) + t_i^w(f_i, \underline{g}) & \text{if link i is signalized ,} \\ t_i^s(f_i) & \text{otherwise .} \end{cases} \tag{2.23}$$

Recent studies have shown that e_i, the fuel consumption on link i, is a linear function of the link distance d_i, and link time, t_i.

$$e_i(f_i, \underline{g}) = k_1 d_i + k_2 t_i(f_i, \underline{g}) \ , \tag{2.24}$$

where k_1 and k_2 are known constants [42], [43].

Since the time and fuel consumption functions on each link i can be written as functions of link flows and control parameters as shown in (2.23) and (2.24), the total system cost can also be expressed as a function of link flows and

controls.

$$J = J(\underline{f}, \underline{g}) \ . \tag{2.25}$$

Using Eq. (2.5),

$$\underline{f} = \sum_{k=1}^{K} \underline{B}^k \underline{h}^k \ , \tag{2.26}$$

which relates link flow vector to path flow vectors, the total system cost can also be expressed in terms of path flows and controls.

$$\begin{aligned} J &= J(\underline{f},\underline{g}) , \\ &= J(\sum_{k=1}^{K} \underline{B}^k \underline{h}^k, \underline{g}) . \end{aligned} \tag{2.27}$$

It is also important to recognize that there is an implicit constraint in the cost function that prevents flow on signalized links from exceeding the effective capacity $g_j c_i$. Since $0.0 \leq g_j \leq 0.9$, flows on signalized links are also prevented from violating the capacity constraint (2.14). For unsignalized links, the fourth-power time function (2.21) serves as a penalty function to prevent excessive violation of the capacity constraints since the fourth-power time function rises steeply for flows exceeding the capacity.

2.2 Generalized System Optimization Problem (GSOP)

Consider the following Generalized System Optimization Problem (GSOP):

$$\underset{\underline{f},\underline{g}}{\text{Minimize}} \quad J(\underline{f},\underline{g}) ,$$

subject to

$$\underline{A}\underline{f}^k = \underline{w}^k \qquad k = 1,\ldots,K \ , \tag{2.28}$$

$$\underline{f}^k \geq 0 \qquad k = 1,\ldots,K \ , \tag{2.29}$$

$$\underline{f} = \sum_{k=1}^{K} \underline{f}^k \ , \tag{2.30}$$

$$f_i \leq \begin{cases} g_j c_i & \text{if link i is signalized;} \\ c_i & \text{otherwise ,} \end{cases} \tag{2.31}$$

$$\underline{g} \varepsilon G \, . \tag{2.32}$$

Constraints (2.28)-(2.32) constitute the requirement that $\underline{f}$ is a feasible flow. In (2.32), $\underline{g}$ is restricted to the set of feasible controls, G.

The GSOP is an optimization problem faced by the traffic authority: to choose a feasible traffic control vector and a feasible network flow such that a certain sytem-wide cost function is minimized.

Throughout this study, the following assumption is made on the input flow vectors $\{\underline{w}^k\}$. We restrict our study to those $\{\underline{w}^k\}$s with the property that there is at least one $\underline{g} \varepsilon G$ such that there exists a feasible flow. From now on, we drop the capacity constraints (2.31) since there is an implicit constraint in the cost function which prevents the flow from exceeding the effective capacity, $g_j c_i$, for signalized links. For unsignalized links, the fourth-power link time function (2.21), with a steep gradient in the region $f_i \geq c_i$, serves as a penalty function to prevent excessive violation of the capacity constraint. In addition, our assumption on the requirement flows which guarantees the existence of a feasible flow also indicates that if there are some violations of the capacity constraints, the violation will not be excessive. From a practical point of view, a slight violation of the capacity constraint is tolerable because the practical capacity of a link is only a very rough estimate [44].

The GSOP is a generalized problem of the conventional System Optimization Problem in the following way. The System Optimization Problem assigns flows in an optimal way, given a set of fixed control variables, whereas in GSOP, both flows and controls are considered as decision variables. In GSOP, the traffic authority chooses feasible controls in addition to assigning a feasible network flow to minimize a system-wide cost. Hence, GSOP extends the scope of decisionmaking of the conventional System Optimization Problem to include traffic controls as decision variables.

2.3 Limitations

The GSOP formulation assumes that the traffic authority possesses the power of arbitrarily assigning any feasible flow to the network in addition to the capability of regulating the traffic-control devices, i.e., traffic signals. There are important reasons why this assumption does not hold in practical situations. First, individual drivers are independent decisionmakers who autonomously choose among different available paths to reach their respective destinations. Second, traffic control devices do not allow the traffic authority to enforce link flow to any arbitrary volumne. They can, at most, limit the access of vehicles to certain links in the network.

To achieve a traffic distribution from the optimal solution, it is necessary to have all users of the system fully cooperative with the central traffic authority, so that each driver is required to surrender his freedom of route choice to the central control agency even though in so doing he may be subjected to higher personal cost.

Following the consideration of these practical limitations, it may be concluded that the GSOP formulation is a highly idealized characterization of the traffic control problem. However, the GSOP formulation has some value in the understanding and analysis of the traffic control problem in spite of the failure to incorporate the constraints of drivers' route choice behavior. The optimal solution of GSOP represents what can be achieved in the best of all possible worlds. Therefore, the optimal cost of GSOP is a lower bound for the cost achievable by any feasible controls. From this point of view, the solution of GSOP can be used to serve as a standard against which different control strategies may be compared and evaluated.

2.4 Hybrid Optimization Problem (HOP)

With the recognition of the limited capability of the traffic authority, it is necessary to reformulate the problem to take into consideration realistically the role of individual drivers as independent decisionmakers in choosing among the different routes available to them. Thus, an important characteristic in this formulation is that there is no one single decisionmaker in the system

anymore. The traffic authority and the drivers are two different sets of decisionmakers, each with a different collection of control or decision variables and different objectives; both are capable of affecting the overall system performance. The traffic authority directly controls the traffic control devices, while the drivers choose among different routes. An important and subtle point to be emphasized is that the route choice among different drivers depends on the traffic control imposed by the traffic authority. Therefore, the traffic authority does possess some control over the flow variables. However, this control is indirect and possibly limited.

In this report, the term "hybrid" is used to describe this special class of transportation network optimization problem. It can be viewed as a specialization of the Generalized System Optimization Problem. More specifically, it is a GSOP with additional constraints on the flow distribution in the network; these constraints reflect driver behavior and decisions. It is hybrid in the sense that the objective function to be minimized is a certain system-wide cost, but the flows are constrained to distribute according to some behavioral model, e.g., individual minimization of personal trip cost. The Hybrid Optimization Problem can now be stated as follows:

HOP

$$\min_{\underline{f},\underline{g}} J(\underline{f},\underline{g}),$$

subject to

$$\underline{Af}^k = \underline{w}^k , \quad k = 1,\dots,K , \tag{2.33}$$

$$\underline{f}^k \geq 0 , \quad k = 1,\dots,K , \tag{2.34}$$

$$\underline{f} = \sum_{k=1}^{K} \underline{f}^k , \tag{2.35}$$

$$\underline{g} \varepsilon G , \tag{2.36}$$

$\underline{f}$ satisfies traffic distribution laws under some behavioral assumptions on drivers' route choice . (2.37)

An explicit mathematical description of (2.37) is presented in Section 3.

The difference between the GSOP and HOP formulations is the flow distribution constraint (2.37). Without the constraint (2.37), the HOP formulation is equivalent to the GSOP formulation. In the HOP formulation, the traffic authority is not allowed to assign any arbitrary feasible flow, as in the GSOP formulation, to minimize the system cost. Instead, the flow is required to satisfy some distribution principles under certain assumptions on the drivers' route selection behavioral model. In the HOP formulation presented here, the constraint (2.37) on flow distribution is intentionally left as general as possible because there exists in the literature several different behavioral models on the drivers' route selection process. This is the subject to be discussed in the next section.

Conceptually, the HOP formulation can be viewed in a more general setting. It is a system-wide network optimization in an environment where flows cannot be arbitrarily assigned, but are distributed in a prescribed way which is influenced by the decision variables chosen by the traffic authority. Hence for this class of problems, it is necessary that decision variables be chosen in anticipation of the reaction of the drivers.

We quote the Equilibrium Network Design Problem considered by Abdulaal et al. [41] as an example which illustrates the generality of the HOP formulation. In the Equilibrium Network Design Problem, one is faced with the problem of determining the practical capacities of a set of links in a network to be constructed or improved, so that a certain system-wide cost is minimized, e.g., total fuel consumption, with the assumption that flow is distributed in the improved network according to User Optimization. In this case, the control parameters $\{g_i\}$ are the link capacities under consideration. The maximum available budget for expenditures on construction and improvement imposes a constraint on the control parameters and must be properly incorporated in the set of feasible controls, G.

The dependence of the system cost on the decision variable appears in the link travel-time function (2.21),

$$t_i^s(\underline{f},\underline{g}) = t_i^o(1 + 0.15(\frac{f_i}{g_i})^4), \qquad (2.38)$$

where g_i, capacity on link i, is one of the decision variables.

The Network Equilibrium Design Problem can then be stated as follows

NEDP

$$\underset{\underline{f},\underline{g}}{\text{Minimize}}\ J(\underline{f},\underline{g}),$$

subject to

$$\underline{A}\underline{f}^k = \underline{w}^k, \quad k = 1,\dots,K, \tag{2.39}$$

$$\underline{f}^k \geq 0, \quad k = 1,\dots,K, \tag{2.40}$$

$$\underline{f} = \sum_{k=1}^{K} \underline{f}^k, \tag{2.41}$$

$$\underline{g} \in G, \tag{2.42}$$

$$\underline{f} \text{ is an equilibrium flow.} \tag{2.43}$$

Comparing the NEDP (2.39)-(2.43) and the HOP (2.33)-(2.37), it is clear that the NEDP can be formulated as another HOP.

2.5 Conclusion

In this section, we have defined the notation and terminology used in this report. A Generalized System Optimization Problem was introduced as an extension of the conventional system optimization problem in traffic engineering. We have pointed out several practical issues which are not incorporated in the GSOP formulation. A Hybrid Optimization Problem has been formulated as a modification of the GSOP to take into account realistically the role of drivers' route choice behavior in the traffic control problem.

3. USER OPTIMIZATION: FLOW DISTRIBUTION PRINCIPLE

An important and perhaps one of the most complicating issues in the study of the traffic control problem is the role of individual drivers as independent decisionmakers when they choose among the routes available in the network. This section is devoted to the discussion of user optimization, the flow distribution principle to be used in this study, and its underlying behavioral assumptions on the process of route selection of individual drivers.

The objective of this section is to derive a mathematical representation of the user-optimized flow pattern which in turn can be incorporated as constraints to the Hybrid Optimization Problem. User-optimized flow pattern is usually characterized mathematically in ways that are not convenient for our purposes: either as a solution to an optimization problem [20], or as a system of logical statements of the form "If a, then b." Section 3.2 provides a mathematical characterization of the equilibrium flow pattern expressed as a set of equalities and inequalities. This representation can then be conveniently posed as constraints to the Hybrid Optimization Problem. In section 3.3, we present the probabilistic assignment models which are commonly used in the literature as alternative models for flow distribution.

3.1 Principle of Equal Travel Time (Wardrop's First Principle) on Traffic Distribution

The way traffic distributes itself in a transportation network is a complicated process which depends primarily on the drivers' route selection behavior and the intersections among the drivers in the network. The choice of route varies among individual drivers and depends on a large number of factors, including the traveling time, the distance along the route, and the number of stops. For this reason, it is a very complex task to formulate a mathematical model which describes precisely the drivers' route choice behavior. Therefore, some simplifying approximations are necessary.

In this study, we adopt the Equal Travel-Time Principle, or Wardrop's First Principle, which is frequently used in traffic assignment to describe network flow distribution [14], [15], [17], [18], [19], [20]. The underlying behavioral assumptions on drivers' route selection processes are:

a) Every driver has a perfect knowledge of the prevailing traffic condition,

b) Each individual driver chooses his route so as to minimize his personal trip cost, which is assumed to be the travel time. That is, he chooses the shortest (in time) available path.

Neither of the hypotheses is grossly unrealistic in a long-term average situation. Through travel experience in the network and traffic broadcasts, a fairly good knowledge of the traffic condition can be acquired. The hypothesis of individual trip-time minimization can be regarded as a close approximation since it has been shown that trip time is the most important determining factor in the drivers' route selection decision [26].

The model of link travel time as an increasing function of flow, which models the congestion effects, together with the two hypotheses of drivers' route choice behavior, results in an equilibrium situation described by Wardrop's First Principle [16], which states that the "journey time on all routes actually used is equal and less than that which would be experienced by a single vehicle on an unused route".

3.2 Mathematical Characterization of Equilibrium Flow

In this section, we express the Equal Travel-Time Principle in mathematical terms. This results in a set of logical mathematical statements of the form, "If a, then b." Unfortunately, a mathematical statement of this form cannot be conveniently posed as constraints to the HOP because a standard mathematical programming problem formulation admits constraints in the form of equality and inequality only. To meet this requirement, we show that it is possible to characterize mathematically the equilibrium conditions in Wardrop's First Principle by a set of equalities and inequalities. Mathematical characterizations of the equilibrium flow in both link flow and path flow formulation are presented. These mathematical characterizations of equilibrium flow are

believed to be an original contribution to the study of traffic network flows.

3.2.1 Path Flow Formulation

It has been shown by a number of studies [45], [36], [31], [18-20], that Wardrop's First Principle can be expressed mathematically as follows:

a) Every path flow must be non-negative. Mathematically,

$$h_i^k \geq 0; \; i \varepsilon P_k: \quad k = 1,\ldots,K, \tag{3.1}$$

b) All path flows of the same O-D pair sum to the requirement flow.

$$\sum_{i \varepsilon P_k} h_k^k = H^k; \quad k = 1,\ldots,K, \tag{3.2}$$

c) Utilized paths have the same trip time, which is equal to the minimum trip time. Mathematically,

$$h_i^k > 0 \text{ implies that } \tau_i^k(\underline{h},\underline{g}) = \min_{j \varepsilon P_k} \tau_j^k \; (\underline{h},\underline{g}); i \varepsilon P_k; \quad k = 1,\ldots,K. \tag{3.3}$$

d) Paths having greater trip time carry no flow.

$$\tau_i^k(\underline{h},\underline{g}) > \min_{j \varepsilon P_k} \tau_j^k(\underline{h},\underline{g}) \text{ implies that } h_i^k = 0, \quad i \varepsilon P_k, \; k = 1,\ldots,K. \tag{3.4}*$$

<u>Definition:</u> A set of path flows $\{\underline{h}^k\}$ is an equilibrium flow if,and only if, $\{\underline{h}^k\}$ satisfies Eqs. (3.1)-(3.4).

The mathematical characterization of equilibrium flow in (3.1), (3.2), (3.3), and (3.4) cannot be conveniently posed as constraints to the Hybrid Optimization Problem because (3.3) and (3.4) are logical statements. We show in Theorem 3.1 a transformation that can be used to overcome this difficulty.

---

* Conditions (3.3) and (3.4) are of the form "If a, then b."

Theorem 3.1

$\{\underline{h}^k\}$ satisfies (3.1)-(3.4) if, and only if, $\{\underline{h}^k\}$ satisfies (3.1), (3.2), and

$$\tau_i^k(\underline{h},\underline{g}) \geq [\sum_{j\epsilon P_k} h_j^k \tau_j^k(\underline{h},\underline{g})]/H^k; \quad i\epsilon P_k; \quad k = 1,\ldots,K. \tag{3.5}$$

Proof:

For clarity and notational convenience, we suppress the arguments of $\tau_j^k(\underline{h},\underline{g})$ in this proof.

A. To show that (3.1), (3.2), (3.3), (3.4) imply (3.5)

$$\text{Let } \tau_*^k = \min_{j\epsilon P_k} \tau_j^k; \quad k = 1,\ldots,K. \tag{3.6}$$

(3.3) and (3.4) imply

$$h_i^k(\tau_i^k - \tau_*^k) = 0, \text{ for all } i\epsilon P_k; \quad k = 1,\ldots,K. \tag{3.7}$$

Summing (3.7) over all $i\epsilon P_k$, τ_*^k can be expressed in terms of $\underline{h}^k$ and $\underline{\tau}^k$,

$$\tau_*^k = \left(\sum_{j\epsilon P_k} h_j^k \tau_j^k\right) / H^k; \quad k = 1,\ldots,K. \tag{3.8}$$

From (3.6), it follows that

$$\tau_i^k \geq \min_{j\epsilon P_k} \tau_j^k = \tau_*^k = \left(\sum_{j\epsilon P_k} h_j^k \tau_j^k\right) / H^k \text{ for all } i\epsilon P_k; \; k - 1,\ldots,K,$$

which is the desired result, (3.5).

B. To show that (3.1), (3.2), (3.5) imply (3.3), (3.4)

$$\text{Let } \tau_*^k = \left(\sum_{j\epsilon P_k} h_j^k \tau_j^k\right) / H^k; \quad k = 1,\ldots,K.$$

(3.5) can be rewritten as

$$\tau_i^k - \tau_*^k \geq 0, \text{ for all } i\epsilon P_k, \quad k = 1,\ldots,K, \tag{3.9}$$

which also implies

$$\tau_*^k \leq \min_{j \varepsilon P_k} \tau_j^k; \quad k = 1,\ldots,K. \tag{3.10}$$

Since, by (3.1), $h_i^k \geq 0$, it follows that

$$h_i^k(\tau_i^k - \tau_*^k) \geq 0, \text{ for all } i\varepsilon P_k, \ k = 1,\ldots,K. \tag{3.11}$$

Let us examine the quantity

$$\begin{aligned} S &= \sum_{i\varepsilon P_k} h_i^k \, (\tau_i^k - \tau_*^k), \\ &= \sum_{i\varepsilon P_k} h_i^k \, \tau_i^k - (\sum_{i\varepsilon P_k} h_i^k) \, \tau_*^k, \\ &= \sum_{i\varepsilon P_k} h_i^k \tau_i^k - (H^k) \left((\sum_{j\varepsilon P_k} h_j^k \tau_j^k) \, / \, H^k \right), \\ &= 0. \end{aligned}$$

Since every individual term in S is, by (3.11), non-negative, and since S equals zero, it follows that

$$h_i^k \, (\tau_i^k - \tau_*^k) = 0 \text{ for all } i\varepsilon P_k, \quad k = 1,\ldots,K, \tag{3.12}$$

$$h_i^k > 0 \Longrightarrow \tau_i^k = \tau_*^k, \text{ for some } i\varepsilon P_k, \quad k = 1,\ldots,K, \tag{3.13}$$

and that

$$\tau_i^k > \tau_*^k \Longrightarrow h_i^k = 0, \text{ for some } i\varepsilon P_k, \ k = 1,\ldots,K. \tag{3.14}$$

It remains to be proved that

$$\tau_*^k = \min_{j\varepsilon P_k} \tau_j^k.$$

We show this as follows. The flow conservation constraints (3.2) states

$$\sum_{i\varepsilon P_k} h_i^k = H^k > 0,$$

which implies there exists some $i\varepsilon P_k$ for all $k = 1,\ldots,K$ such that (3.13) is true; i.e., there is at least one $i\varepsilon P_k$ for every $k = 1,\ldots,K$ such that $\tau_i^k = \tau_*^k$.

But, by Eq. (3.10),

$$\tau_*^k \leq \min_{j\varepsilon P_k} \tau_j^k .$$

Hence,

$$\tau_*^k = \min_{j\varepsilon P_k} \tau_j^k \quad k = 1,\ldots,K.$$

Q.E.D.

3.2.2 Link Flow Formulation

In the study of large-scale networks, it is often desirable to avoid the enumeration of all available paths [33]. This can often be done if we work in the space of link flows. In this section, a mathematical characterization of equilibrium flow in the link flow formulation similar to that obtained in section 3.2.1 is derived.

We introduce a vector, $\underline{v}^k$, of dimension NN, in association with each O-D pair k, $k = 1,\ldots,K$. The i^{th} element of $\underline{v}^k$ is defined as

$$v_i^k \overset{\Delta}{=} \text{the minimum time taken to go from } O^k\text{, the origin node of the } k^{th} \text{ O-D pair, to node } i. \tag{3.15}$$

We note that

$$v_i^k \geq 0 \quad \text{for all } i = 1,\ldots,NN;\ k = 1,\ldots,K,$$

and

$$v_i^k = 0 \quad \text{if and only if } i = O^k. \tag{3.16}$$

In addition, it can be deduced from the definition of v_i^k in (3.15) that all O-D pairs k with the same origin node O^k have the same value of v_i^k for all i. The quantity v_i^k can be viewed as a potential.

It has been shown [46] that Wardrop's First Principle can be expressed mathematically in link flow formulation as follows:

$$\underline{f}^k \geq 0, \tag{3.17}$$

$$\underline{Af}^k = \underline{w}^k, \tag{3.18}$$

$$t_i(\underline{f},\underline{g}) \geq v^k_{\beta(i)} - v^k_{\alpha(i)}. \tag{3.19}$$

$$\text{If } t_i(\underline{f},\underline{g}) > v^k_{\beta(i)} - v^k_{\alpha(i)},$$

$$\text{then } f^k_i = 0 \quad i = 1,\ldots,NL; \quad k = 1,\ldots,K. \tag{3.20}$$

$$\text{If } f^k_i > 0,$$

$$\text{then } t_i\ (\underline{f},\underline{g}) = v^k_{\beta(i)} - v^k_{\alpha(i)} \quad i = 1,\ldots,NL;\ k = 1,\ldots,K, \tag{3.21}$$

where $\alpha(i)$ and $\beta(i)$, as defined in section 2.1, are the initial and terminal nodes, respectively, of link i.

By making use of the node-arc incidence matrix, $\underline{A}$, (3.19) can be written more compactly in vector form:

$$\underline{t}(\underline{f},\underline{g}) + \underline{A}'\underline{v}^k \geq 0, \quad k = 1,\ldots,K. \tag{3.22}$$

The system (3.17)-(3.21), with further manipulations, can be shown to be equivalent to the following

$$\underline{f}^k \geq 0, \tag{3.23}$$

$$\underline{Af}^k = \underline{w}^k, \tag{3.24}$$

$$\underline{t}(\underline{f},\underline{g}) + \underline{A}'\underline{v}^k \geq 0, \tag{3.25}$$

$$(\underline{f}^k)'\underline{t}(\underline{f},\underline{g}) + (\underline{w}^k)'v^k = 0. \tag{3.26}$$

Definition: A set of link flow vectors $\{\underline{f}^k\}$ is an equilibrium flow if and only if there exists $\{\underline{v}^k\}$ such that (3.23)-(3.26) are satisfied.

In this section, we have presented two mathematical characterizations of the equilibrium flow. They consist of equalities and inequalities, and therefore, can be used as constraints in a standard mathematical programming problem formulation. The system (3.1), (3.5) is in path flow formulation, while (3.23)-(3.26) is in link flow formulation. The link flow formulation has an advantage that no path information (i.e., the arc-chain incident matrices $\{\underline{B}^k\}$) is required.

However, this is possible only with the introduction of new variables $\{\underline{v}^k\}$. Consequently, the HOP in link flow formulation is a much larger problem in terms of the number of variables. Thus, there is no clear-cut advantage of using the link flow formulation when computation time and on-line storage involved in solving the HOP are taken into consideration.

3.3 Alternative Flow Distribution Models

In this study, it is assumed that traffic distributes itself according to Wardrop's First Principle. It will be shown that this is a reasonable assumption in later sections. However, there are a number of other models for flow distribution, and the HOP formulated in section 2 is general enough to admit alternative models.

We present in this section a class of probabilistic models cosidered in [47], [48]. These models differ from that of User Optimization in that they attempt to account for (1) the random characteristics of the different perceptions of the traffic condition,and (2) the non-uniformity in preferences among individual drivers.

These models are developed on the following premises. Due to the fact that drivers do not have a perfect knowledge of the traffic condition, and different drivers may have different criteria in the route selection process, for a given O-D pair, paths which are less desirable, say,in terms of travel time, may still carry a positive amount of traffic. The flows on these paths are less than that on the most desirable path. This is the feature distinguishing these models from the User Optimization model, where such less desirable paths carry no flow at all. Empirical equations are hypothesized for flow distribution in a network to capture these qualitative features. In words, these models state that the fraction of drivers taking any particular path between a given O-D pair depends on how attractive that path is as compared with all other available paths. The degree of attractiveness of a particular path is assumed to be a function of the path characteristics, usually the total travel time or the total distance or a weighted linear combination of both. The longer the travel time or the distance of a path, the less attractive it is to the drivers. Parameters are introduced in these models to account for the random nature of the route selection process. These parameters are usually calibrated using observed traffic data.

In the model of diversion curve assignment, the following functional form is hypothesized [47].

$$\frac{h_i^k}{H^k} = \ell_k \sum_{j\varepsilon R_i^k} \{[d_j + \ell_t t_j(\underline{h},\underline{g})]\}^{-\ell_s} \; ; \; i\varepsilon P_k; \quad k = 1,\ldots,K, \tag{3.27}$$

where the summation is taken over all those link j's belong to R_i^k, the set of all links on the i^{th} path of the k^{th} O-D pair; d_j and $t_j(\underline{h},\underline{g})$, as we recall, are the distance and travel time on link j; ℓ_k, ℓ_t, ℓ_s are parameters to be explained below.

Briefly, the model of diversion curve assignment, expressed in Eq. (3.27), states that the fraction of traffic flow of the k^{th} O-D pair taking the i^{th} path, i.e., $\frac{h_i^k}{H^k}$, is inversely proportional to the route characteristic on that path raised to the power ℓ_s. The route characteristic considered in the drivers' route selection is assumed to be a weighted combination of the total travel distance and total travel time. The tradeoff between distance and travel time in the decisionmaking of the drivers is expressed by ℓ_t. The degree which drivers are sensitive to route characteristics when choice is made between alternative routes is parameterized in ℓ_s. This reflects in part the quality of information the drivers have on the traffic condition, and the strength of preference of the drivers. Given two models with different values of ℓ_s, the one with a larger value will have less traffic flow on the less desirable paths. The constant ℓ_k is a normalizing factor which may be obtained by summing (3.27) over all paths available between the k^{th} O-D pair, i.e., over all i such that $i\varepsilon P_k$. This results in

$$\ell_k = \sum_{i\varepsilon P_k} \{\sum_{j\varepsilon R_i^k} [d_j + \ell_t t_j(\underline{h},\underline{g})]\}^{\ell_s}; \; k = 1,\ldots,K. \tag{3.28}$$

In the multipath probabilistic assignment model considered by Dial [48], the following functional form is assumed for flow distribution.

$$\frac{h_i^k}{H^k} = \ell_k \exp\{-\ell_s[\tau_i^k(\underline{h},\underline{g}) - \tau_*^k(\underline{h},\underline{g})]\}, \quad \text{for all } i\varepsilon P_k; \quad k = 1,\ldots,K, \tag{3.29}$$

where

$\tau_i^k(\underline{h},\underline{g})$ is the total travel time along the i^{th} path of the k^{th} O-D pair;

$$\tau_*^k(\underline{h},\underline{g}) \triangleq \min_{j\varepsilon P_k} \tau_j^k(\underline{h},\underline{g}) \tag{3.30}$$

is the path travel time along the fastest route connecting the k^{th} O-D pair; ℓ_k and ℓ_s are parameters to be explained below.

The multipath probabilitic assignment model states that the fraction of traffic of the k^{th} O-D pair taking the i^{th} path is a function of the excess travel time of that path over that of the fastest path. The functional form exp $(\cdot)$ is chosen to capture the qualitative feature that paths having longer travel time carry less flow. The parameter ℓ_s measures how sensitive the drivers are to the total travel time in choosing among alternative paths. The normalizing factor, ℓ_k, can be obtained explicitly by summing (3.29) over all paths in P_k as in the case of the model of diversion curve assignment,

$$\ell_k = \frac{1}{\sum\limits_{i\varepsilon P_k} \exp\{-\ell_s[\tau_i^k(\underline{h},\underline{g}) - \tau_*^k(\underline{h},\underline{g})]\}} \; ; \; k = 1,\ldots,K \; . \tag{3.31}$$

Because of the particular functional form chosen in this model, the multi-path probabilitic flow assignment model as expressed in (3.29) and (3.31) can be further simplified to the following form:

$$\frac{h_i^k}{H^k} = \frac{\exp\{-\ell_s\tau_i^k(\underline{h},\underline{g})\}}{\sum\limits_{j\varepsilon P_k} \exp\{-\ell_s\tau_j^k(\underline{h},\underline{g})\}} \; ; \; i\varepsilon P_k \, , \quad k = 1,\ldots,K \; . \tag{3.32}$$

3.4 Discussion and Summary

In this section, we have presented different models for network flow distribution. Two mathematical characterizations of the equilibrium flow pattern are presented. These characterizations are in the form of inequalities, and can be used as constraints in the HOP formulation in Section 2. Alternative models,

broadly classified as the probilistic assignment models, which include the diversion curve assignment model and the multipath probabilistic assignment model are also discussed.

In the remainder of this study,we restrict our attention to a user optimization model for a number of reasons. In modeling, the fact is appreciated that any mathematical behavioral model, no matter how complicated or sophisticated, can only be an approximation to the real system. User optimization is among one of the most commonly used and better understood models in the area of traffic assignment. Another important justification for using user optimization as a flow distribution model is that equilibrium flow has been shown [49] to be a reasonably good approximation of the actual traffic distribution. Because of these considerations, much emphasis has been given to the formulation of equilibrium flow. In this section,the equilibrium flow pattern has been mathematically characterized in a useful way in the sense that it can be conveniently posed as constraints to the HOP.

4. NECESSARY CONDITIONS AND EXTENDED EQUILIBRIUM PRINCIPLE OF HYBRID OPTIMIZATION PROBLEM

The general formulation of the Hybrid Optimization problem in Section 2, together with the mathematical representation of flow distribution models in Section 3, comprises a precise mathematical statement of a specific HOP which assumes Wardrop's First Principle as a model for the flow distribution. In section 4.1, we present the derivation and analysis of the necessary conditions of this problem. From the analysis, we obtain an Extended Equilibrium Principle for the HOP. It is an interesting fact that the Extended Equilibrium Principle bears a remarkable similarity to the Equal Marginal Cost Principle in the system optimization problem (20). This seems quite intuitive since the HOP combines features of the user optimization problem and the system optimization problem. This is discussed in Section 4.2. We defer the presentation of the specific numerical examples to illustrate the Extended Equilibrium Principle to the next section.

4.1 Necessary Conditions for Hybrid Optimization Problem in Path Flow Formulation

4.1.1 Problem Statement

By using the general formulation of HOP in Section 2 and the results on equilibrium flow representation, a precise mathematical statement for the HOP may be written. However, before doing so, we rewrite the inequalities in (3.5) in vector form for the sake of compactness,

$$\underline{\tau}^k(\underline{f},\underline{g}) \geq (\underline{h}^{k'}\underline{\tau}^k/H^k)\underline{u}^k, \tag{4.1}$$

where $\underline{u}^k$ is a vector of dimension $|P_k|$ with all elements having a value of 1. By using the arc-chain incidence matrices, $\underline{B}^k$, (4.1) can be rewritten as

$$\underline{B}^{k'}\underline{t}(\underline{f},\underline{g}) \geq \left(\underline{f}^{k'}\underline{t}(\underline{f},\underline{g})/H^k\right)\underline{u}^k \quad . \tag{4.2}$$

Problem Statement of the HOP

HOP: $\underset{\underline{h},\underline{g}}{\text{minimize}} \; J(\underline{f},\underline{g})$

subject to

$$\sum_{i\in P_k} h_i^k = H^k , \quad k = 1,\ldots,K, \tag{4.3}$$

$$\underline{h}^k \geq 0 , \quad k = 1,\ldots,K, \tag{4.4}$$

$$\underline{B}^{k'} \underline{t}(\underline{f},\underline{g}) \geq \left(\underline{f}^{k'} \underline{t}(\underline{f},\underline{g})/H^k\right) \underline{u}^k, \; k = 1,\ldots,K , \tag{4.5}$$

$$0.9 \geq g_i \geq 0 , \quad i = 1,\ldots,U . \tag{4.6}$$

In this problem statement of HOP, the decision variables are $\underline{h}$ and $\underline{g}$. Recall that $\underline{h}$ is a vector of path flows and $\underline{g}$ is a vector of control parameters, the green splits. The link flow variables, $\{f_i\}$, are not independent decision variables. They are related to the path flow variables, $\underline{h}^k$, by

$$\underline{f} = \sum_{k=1}^{K} \underline{B}^k \underline{h}^k . \tag{4.7}$$

4.1.2 Necessary Conditions

Let L be the Lagrange function of the minimization problem of the HOP.

$$\begin{aligned} L = J(\underline{f},\underline{g}) &+ \sum_{k=1}^{K} \lambda_k \left(\sum_{i\in P_k} h_i^k - H^k\right) - \sum_{k=1}^{K} \underline{\mu}^{k'} \underline{h}^k \\ &- \sum_{k=1}^{K} \underline{\gamma}^{k'}\left[\underline{B}^{k'} \underline{t}(\underline{f},\underline{g}) - \left(\underline{f}^{k'} \underline{t}(\underline{f},\underline{g})/H^k\right) \underline{u}^k\right] \\ &- \sum_{i=1}^{U} \eta_i w_i - \sum_{i=1}^{U} \sigma_i (0.9 - w_i), \end{aligned} \tag{4.8}$$

where $\{\lambda_k\}$, $\{\underline{\mu}^k\}$, $\{\underline{\gamma}^k\}$, $\{\eta_i\}$, $\{\sigma_i\}$ are the Lagrange multipliers of constraints (4.3), (4.4), (4.5), and (4.6), respectively, with $\{\lambda_k\}$, $\{\eta_i\}$, $\{\sigma_i\}$ scalars and $\{\underline{u}^k\}$, $\{\underline{\gamma}^k\}$ vectors of dimension $|P_k|$.

Applying the results of optimization theory on the necessary conditions for the optimal solution of a constrained optimization problem (see, for example, Theorem 3.9 in Avriel [50]), we obtain the following theorem.

Theorem 4.1: Necessary Conditions of HOP

Suppose $(\underline{h},\underline{g})$ is an optimal solution of HOP. Then there exists a set of Lagrange multipliers $\{\lambda_k\}$, $\{\underline{\mu}^k\}$, $\{\underline{\gamma}^k\}$, $\{\eta_i\}$, $\{\sigma_i\}$ such that the following system of equalities and inequalities is satisfied.

$$\frac{\partial L}{\partial h_i^k} = \left\{ \frac{\partial J'}{\partial \underline{f}} + (\underline{\gamma}^{k'}\underline{u}^k/H^k)\underline{t}' + \left[\sum_{j=1}^{K} (\underline{\gamma}^{j'}\underline{u}^j/H^j)\, \underline{f}^j - \underline{B}^j\underline{\gamma}^j \right]' \frac{\partial \underline{t}}{\partial \underline{f}} \right\} \frac{\partial \underline{f}^k}{\partial h_i^k}$$

$$+ \lambda_k - \mu_i^k = 0 \quad ; \quad i \in P_k; \quad k = 1,\ldots,K\,, \qquad (4.9)$$

$$\frac{\partial L}{\partial g_i} = \frac{\partial J}{\partial g_i} + \left[\sum_{j=1}^{K} (\underline{\gamma}^{j'}\underline{u}^j/H^j)\underline{f}^j - \underline{B}^j\underline{\gamma}^j \right]' \frac{\partial \underline{t}}{\partial g_i} - \eta_i + \sigma_i = 0\,,$$

$$i = 1,\ldots,U, \qquad (4.10)$$

$$\sum_{i\epsilon P_k} h_i^k = H^k \quad ; \quad k = 1,\ldots,K\,, \qquad (4.11)$$

$$\underline{h}^k \geq 0 \quad ; \quad k = 1,\ldots,K\,, \qquad (4.12)$$

$$\underline{B}^{k'}\underline{t}(\underline{f},\underline{g}) \geq (\underline{f}^{k'}\underline{t}(\underline{f},\underline{g})/H^k)\underline{u}^k\,, \qquad k = 1,\ldots,K\,, \qquad (4.13)$$

$$0.9 \geq g_i \geq 0\,, \qquad i = 1,\ldots,U\,, \qquad (4.14)$$

$$\underline{\mu}^k \geq 0\,, \qquad k = 1,\ldots,K\,, \qquad (4.15)$$

$$\underline{\mu}^{k'}\underline{h}^k = 0\,, \qquad k = 1,\ldots,K\,, \qquad (4.16)$$

$$\underline{\gamma}^k \geq 0\,, \qquad k = 1,\ldots,K\,, \qquad (4.17)$$

$$\underline{\gamma}^{k'}\left[\underline{B}^{k'}\underline{t} - (\underline{f}^{k'}\underline{t}/H^k)\underline{u}^k\right] = 0\,, \qquad k = 1,\ldots,K\,, \qquad (4.18)$$

$$\eta_i \geq 0\,, \qquad i = 1,\ldots,U\,, \qquad (4.19)$$

$$\eta_i g_i = 0 , \qquad i = 1,\ldots,U , \tag{4.20}$$

$$\sigma_i \geq 0 , \qquad i = 1,\ldots,U , \tag{4.21}$$

$$\sigma_i (0.9 - g_i) = 0 , \qquad i = 1,\ldots,U , \tag{4.22}$$

where $\frac{\partial J}{\partial \underline{f}}$, $\frac{\partial \underline{f}^k}{\partial h_i^k}$, $\frac{\partial \underline{t}}{\partial \underline{f}}$, $\frac{\partial J}{\partial g_i}$, $\frac{\partial \underline{t}}{\partial g_i}$ in (4.9) and (4.10) are defined as follows:

$\frac{\partial J}{\partial \underline{f}}$ is a vector of dimension NL, the number of links. The i^{th} element is the partial derivative of the system cost with respect to f_i, the total flow on link i.

$\frac{\partial \underline{f}^k}{\partial h_i^k}$ is a vector of dimension NL. The j^{th} element, $\left(\frac{\partial \underline{f}^k}{\partial h_i^k}\right) = \left(\frac{\partial f_j^k}{\partial h_i^k}\right)$, which has a value of 1 if link j is on the i^{th} path of the k^{th} O-D pair, and zero otherwise, i.e., $(\partial f_j^k / \partial h_i^k) = b_{ji}^k$ from (2.2) and (2.3).

$\frac{\partial \underline{t}}{\partial \underline{f}}$ is a matrix of dimension NL by NL. The (i,j) element equals $\partial t_i / \partial f_j$. In the case of separable link time function; i.e., the travel time on link i does not depend on flows on any other links, $\frac{\partial \underline{t}}{\partial \underline{f}}$ is a diagonal matrix; i.e.,

$$\left(\frac{\partial \underline{t}}{\partial \underline{f}}\right)_{i,j} = 0 , \qquad j \neq i,\ i=1,\ldots,NL,\ j=1,\ldots,NL, \tag{4.23}$$

$$\left(\frac{\partial t}{\partial f}\right)_{i,i} = \frac{\partial t_i}{\partial f_i} , \qquad i = 1,\ldots,NL . \tag{4.24}$$

$\frac{\partial J}{\partial g_i}$ is a scalar which denotes the partial derivative of the system cost with respect to the i^{th} control variable.

$\frac{\partial \underline{t}}{\partial g_i}$ is a vector of dimension NL. The j^{th} element, $\left(\frac{\partial \underline{t}}{\partial g_i}\right)_j$, is $\frac{\partial t_j}{\partial g_i}$.

Theorem 4.1 simply states the set of Kuhn-Tucker necessary conditions that characterize the optimal solution of the Hybrid Optimization Problem. We will show that there is a more comprehensible and physical interpretation of these conditions in the next section. Equations (4.9) and (4.10) state that the partial derivatives of the Lagrangian function with respect to the decision variables, $\{h_i^k\}$, $\{g_i\}$, equal to zero. Conditions (4.11), (4.12), (4.13), and (4.14) are the constraints of the HOP which must be satisfied by any feasible solution, and hence by the optimal solution.

The group of conditions (4.12), (4.15), (4.16) is the complementarity condition of the inequality constraints (4.4). Similarly, (4.13), (4.17), (4.18) are complementarity conditions for the inequality constraints (4.2), and the same is true for (4.14), (4.19), and (4,20), as well as (4.14), (4.21), and (4.22).

4.2 Extended Equilibrium Principle for HOP

In this section we examine the necessary conditions more closely in an attempt to find a physical interpretation of these conditions. To do so, let

$$\theta^k \triangleq \underline{\gamma}^{k'} \underline{u}^k / H^k = \sum_{i \varepsilon P_k} \gamma_i^k / H^k \geq 0 \quad , \quad k = 1,\ldots,K , \tag{4.25}$$

$$\underline{\pi} = \sum_{k=1}^{K} (\underline{\gamma}^{k'} \underline{u}^k / H^k) \underline{f}^k - \underline{B}^k \underline{\gamma}^k = \sum_{k=1}^{K} (\underline{f}^k \underline{u}^{k'} / H^k - \underline{B}^k) \underline{\gamma}^k . \tag{4.26}$$

In Eqs. (4.25) and (4.26), θ^k is a scalar, and $\underline{\pi}$ is a vector of dimension NL. It is clear that both θ^k and π are linear functions of the Lagrange multiplier $\{\gamma^k\}$. It can be shown that

$$\underline{\pi}'\underline{t} = 0 . \tag{4.27}$$

Equation (4.9) can be rewritten as follows:

$$\left[\frac{\partial J'}{\partial \underline{f}} + \theta^k \underline{t}' + \underline{\pi}' \frac{\partial \underline{t}}{\partial \underline{f}}\right] \frac{\partial \underline{f}^k}{\partial h_i^k} = - \lambda_k + \mu_i^k ,$$

$$i \varepsilon P_k, \ k = 1,\ldots,K . \tag{4,28}$$

Let

$$\underline{\psi}^k \triangleq \frac{\partial J}{\partial \underline{f}} + \theta^k \underline{t} + \frac{\partial \underline{t}'}{\partial \underline{f}} \underline{\pi} \quad , \qquad k = 1,\ldots,K \; , \tag{4.29}$$

where $\underline{\psi}^k$ is a vector of dimension NL.

Equation (4.25), when written component-wise, becomes

$$\psi_i^k = \frac{\partial J}{\partial f_i} + \theta^k t_i + \sum_{j=1}^{NL} \frac{\partial t_j}{\partial f_i} \pi_j \; , \tag{4.30}$$

where the subscript, i, stands for the index for the link number, $1 \leq i \leq NL$, and the superscript, k, stands for the index of O-D pair number, $1 \leq k \leq K$.

In the case of separable link-time cost, i.e., where

$$\frac{\partial t_j}{\partial f_i} = 0 \; , \text{ for } j \neq i \; , \tag{4.31}$$

we have

$$\psi_i^k = \frac{\partial J}{\partial f_i} + \theta^k t_i + \pi_i \frac{\partial t_i}{\partial f_i} \quad . \tag{4.32}$$

Let us define

$$z_i^k = \underline{\psi}^{k'} \frac{\partial \underline{f}^k}{\partial h_i^k} \quad . \tag{4.33}$$

We can now rewrite Eq. (4.28) as

$$z_i^k = \underline{\psi}^{k'} \frac{\partial \underline{f}^k}{\partial h_i^k} \; ,$$

$$= \sum_{j=1}^{NL} \psi_j^k \frac{\partial f_j^k}{\partial h_i^k} \; ,$$

$$= \sum_{j=1}^{NL} \psi_j^k \, b_{ji}^k = - \lambda_k + \mu_i^k \; . \tag{4.34}$$

Using the definition of b_{ji}^k in Eq. (2.3), we observe that z_i^k is a summation of the quantities ψ_j^k over those link j's which are on the i^{th} path of the k^{th} O-D pair.

Equations (4.12), (4.15), (4.16) imply that:

$$\text{If } h_i^k > 0, \text{ then } \mu_i^k = 0. \tag{4.35}$$

$$\text{If } \mu_i^k > 0, \text{ then } h_i^k = 0. \tag{4.36}$$

Consequently, (4.34), (4.35), and (4.36) imply that

$$\text{If } h_i^k > 0, \text{ then } z_i^k = -\lambda_k, \tag{4.37}$$

$$\text{If } z_i^k > -\lambda_k, \text{ then } h_i^k = 0 \ . \tag{4.38}$$

In words, statements (4.37) and (4.38) say that those paths carrying positive flow have the same value of z_i^k which is equal to $-\lambda_k$. Any path having the value z_i^k greater than $-\lambda_k$ carries no flow. This is a statement of the equilibration of the quantity z_i^k over utilized paths. In addition, the optimal solution must also satisfy the Equal Travel Time Principle, as expressed in constraints (4.3), (4.4), (4.5). We summarize this result in the following theorem.

Theorem 4.2: Extended Equilibrium Principle

Suppose $(\underline{h},\underline{g})$ is an optimal solution of HOP. Then there exists some scalars $\{\theta^k, k = 1,\ldots,K\}$ and a vector $\underline{\pi}\epsilon R^{NL}$, which are related to the Lagrange multipliers $\{\underline{\gamma}^k\}$ by (4.25) and (4.26) such that the following statements are true.

1) The trip times along all used paths of the k^{th} O-D pair are the same and equal to τ_*^k. Any path belonging to P_k having a trip time greater than τ_*^k carries no flow.

2) Let z_j^k defined as in (4.34) be called a pseudo-cost along the j^{th} path of the k^{th} O-D pair. The pseudo-costs along all used paths between the k^{th} O-D pair are the same and equal to some value, say, z_*^k. Any path having a pseudo-cost greater than z_*^k carries no flow.

Further Discussion

In (4.34), the pseudo-cost on the i^{th} path of the k^{th} O-D pair is the summation of ψ_j^k's over those links along the i^{th} path of the k^{th} O-D pair. Notice that in Eq. (4.32), ψ_j^k has three parts:

$$\psi_j^k = \frac{\partial J}{\partial f_j} + \theta^k t_j + \pi_j \frac{\partial t_j}{\partial f_j} . \tag{4.39}$$

Along all utilized paths, travel times are equalized; i.e.,

$$\sum_j t_j ,$$

are equal along all utilized paths of the same O-D pair. However the Extended Equilibrium Principle also states that

$$\sum_j \psi_j^k ,$$

are equalized over all utilized paths of the same O-D pair. We deduce from (4.39) that the quantity,

$$\sum_j \frac{\partial J}{\partial f_j} + \pi_j \frac{\partial t_j}{\partial f_j} ,$$

is also equalized over all used paths of the same O-D pair.

In a special case, where $\underline{\pi}$ is zero, then the Extended Equilibrium Principle reduces to the statement that both travel time and marginal cost are equalized. In this case, the optimal control is chosen such that the resultant flow distribution is both system-optimal and user-optimal. This is rather a rare case, and we do not expect this for practical problems.

4.3 Summary

In this section, we derived the Kuhn-Tucker necessary conditions for the optimal solution for a specific Hybrid Optimization Problem which assumes user optimization as a model for flow distribution. We also provide a more physical

interpretation of these necessary conditions. This is summarized in the Extended Equilibrium Principle, which extends the conventional Equal Travel-Time Principle: not only is travel time equalized over utilized paths, but also a pseudo-cost which is related to the gradient of the system cost and the Lagrange multipliers.

5. NUMERICAL EXAMPLES

The preceding sections have been devoted to theoretical discussions of the Hybrid Optimization Problem. In this section, we divert to a discussion on some specific numerical examples. The main purposes of this section are to illustrate the concept of the Hybrid Optimization Problem and to demonstrate the validity of the mathematical formulation presented.

In Section 5.1, a general nonlinear constrained minimization problem to be used in solving the HOP is described. Section 5.2 presents a discussion of an algorithm which has been proposed and applied to traffic control problems of the same nature as the HOP [9], [14], [15]. We will refer to this algorithm as the Iterative Optimization-Assignment Algorithm in this report. Although this seems reasonable, it is shown that in the examples considered in this section, the algorithm converges to wrong solutions. In sections 5.3 to 5.5, we describe several numerical examples together with their solutions obtained by the two different methods. A comparison of the solutions is also given.

5.1 Augmented Lagrangian Method [50-54]

The HOP formulation in Section 2 can be viewed as a nonlinear constrained optimization problem expressible in the following general form.

NLP

$$\underset{\underline{x}\in R^n}{\text{Minimize}} \quad J(\underline{x})$$

subject to

$$w_i(\underline{x}) \geq 0, \qquad i=1, \ldots, M, \tag{5.1}$$

$$\phi_i(\underline{x}) = 0, \qquad i=1, \ldots, N, \tag{5.2}$$

where $\underline{x}$ is an n-dimensional vector of decision variables, and $\{w_i\}$, $\{\phi_i\}$ are the inequality and equality constraints, respectively. The notation used in this section is defined strictly for this section, and is independent of that used in other parts of this report.

In this section, we describe a general optimization algorithm to be used as a solution method for the numerical examples in Section 5.3. The particular optimization algorithm to be considered is the augmented Lagrangian method. The study of the augmented Lagrangian method is an area of extensive research in optimization theory in recent years. It is believed [50] to be one of the most flexible and efficient general algorithms available for constrainted optimization [51]. Rigorous theoretical discussions on convergence and methods of adjustment of parameters of the algorithm are published in several papers. Convergence has been proved elegantly using duality theory. The surveys by Bertsekas [52], Fletcher [53], Rockafellar [54], and Powell [51] are recommended as references for this subject. Detailed discussions on the various theoretical aspects of this method is omitted here since it is not the main theme of this report. Only a brief outline is provided in this section.

The augmented Lagrangian method resembles the classical penalty function method in the basic approach of transforming a constrained minimization problem into a sequence of unconstrained ones by incorporating the constraint functions in the objective function of the unconstrained problem. In the classical penalty function method, various penalty functions have been defined for the unconstrained problem. A commonly used [54] penalty function is

$$\Phi(\underline{x},\rho^k) = J(\underline{x}) - \rho^k \sum_{i=1}^{M} \log\,[w_i(\underline{x})] + \frac{1}{\rho^k} \sum_{i=1}^{N} [\phi_j(\underline{x})]^2 , \tag{5.3}$$

where $\rho^k > 0$, and a sequence of unconstrained minimizations of Φ is performed with respect to $\underline{x} \in R^n$ for $\{\rho^k\} \to 0$. Let $\underline{x}^{k*}$ be the unconstrained optimal solution that minimizes $\Phi(\underline{x},\rho^k)$.

The main advantage of the penalty function method is that, in constrast to other methods such as the gradient projection method [55] that explicitly treat nonlinear constraints, penalty methods avoid the time-consuming and often difficult task of moving along nonlinear boundaries of the feasible set, or trying not to cross them. The efficiency of the penalty method depends heavily on the existence of a very efficient unconstrained minimization algorithm such as the variable metric algorithms [50]. Another attractive feature of the penalty function method is the availability of the Lagrange multipliers at the optimum point. This is particularly useful for sensitivity analysis in general problems, and in Section 6, we show that this can be used to verify the global

optimality of the solution of a subproblem of the HOP. The sequences $\{\rho^k/[w_i(\underline{x}^{k*})]^2\}$ and $\{2\phi_j(\underline{x}^{k*})/\rho^k\}$ can be shown [56] to converge, under some mild assumptions, to the corresponding Lagrange multipliers ζ_i and ξ_j associated with constraints (5.1) and (5.2), respectively.

However, it is also known [50] that the major problem of this method is that the Hessian matrix of Φ becomes increasingly ill-conditioned as $\{\rho^k\} \to 0$. Ironically, the variable metric methods for unconstrained minimization which makes the penalty function method so very efficient is particularly vulnerable to this main disadvantage of the penalty function method.

The augmented Lagrangian method avoids this problem without losing any of the nice properties of the penalty function method by defining a new objective function called the augmented Lagrangian function, for the uncontrained problem.

Define

$$\Lambda(\cdot) \overset{\Delta}{=} \max(\cdot,0) \quad . \tag{5.4}$$

The augmented Lagrangian function for the constrained minimization problem NLP is:

$$\Omega(\underline{x},\underline{\zeta},\underline{\xi}) = J(\underline{x}) + \frac{1}{4\rho}\sum_{i=1}^{M}\{[\Lambda(\zeta_i - 2\rho w_i(\underline{x}))]^2 - (\zeta_i)^2\} + \sum_{j=1}^{N}\xi_j\phi_j(\underline{x}) + \frac{\rho}{2}\sum_{j=1}^{N}[\phi_j(\underline{x})]^2 \quad . \tag{5.5}$$

For a given $\rho > 0$ and multipliers $(\underline{\zeta}^k,\underline{\xi}^k)$, let $\underline{x}^{k*}$ be the optimal solution of the unconstrained optimization problem (UP_k)

$\underline{UP_k}$

$$\min_{\underline{x}\in R^n} \Omega(\underline{x},\underline{\zeta}^k,\underline{\xi}^k) \quad .$$

Several methods have been developed for the updating of the multipliers $(\underline{\zeta}^k,\underline{\xi}^k)$. For example, [50]:

$$\zeta_i^{k+1} = \Lambda[\zeta_i^k - 2\rho w_i(\underline{x}^{k*})] \; , \; i=1, \ldots, M \; , \tag{5.6}$$

$$\xi_j^{k+1} = \xi_j^k - \rho\phi_j(x^{k*}) \; , \qquad j=1, \ldots, N \; . \tag{5.7}$$

Several properties of the augmented Lagrangian method are known. It has been shown [50] that for sufficiently large but finite values of ρ, the sequences $\{\underline{x}^{k*}\}$, $\{\underline{\zeta}^{k}\}$, $\{\underline{\xi}^{k}\}$ tend to $\underline{x}^*$, $\underline{\zeta}^*$, $\underline{\xi}^*$, which are the optimal solution and Lagrange multipliers of NLP under some mild conditions. It is also known that the rate of convergence is linear.

5.2 Iterative Optimization-Assignment Algorithm

The Iterative Optimization-Assignment Algorithm is an iterative procedure consisting of successive alternations between a signal-optimizing program and a flow assignment program as shown in Figure 5.1. In the flow assignment phase of the procedure, an equilibrium flow or user-optimized flow is computed assuming the control parameters such as the green splits are fixed. The determination of the equilibrium flow vector, under the assumption of separable link time functions, has been shown [18], [19], [20] to be equivalent to the following minimization problem.

Equivalent Minimization Problem for Equilibrium Flow Computation

$$\underset{\{\underline{h}^k\}}{\text{Minimize}} \sum_{i=1}^{NL} \int_0^{f_i} t_i(\underline{x},\underline{g})\Big|_{\underline{g}\ \text{fixed}} dx \quad ,$$

subject to

non-negativity of path flows:

$$\underline{h}^k \geq 0 \quad , k = 1,\ldots, K \quad , \tag{5.8}$$

path flow conservation:

$$\sum_{i\in P_k} h_i^k = H^k \quad , \quad k = 1,\ldots, K \quad , \tag{5.9}$$

where the link flow, $\underline{f}$, is a known function of the path flows $\{\underline{h}^k\}$ defined in Section 2, and is given by

$$\underline{f} = \sum_{k=1}^{K} \underline{B}^k \underline{h}^k \ . \tag{5.10}$$

The signal-optimizing program computes a set of optimal signal settings with respect to some system cost, e.g., total travel time, assuming flows

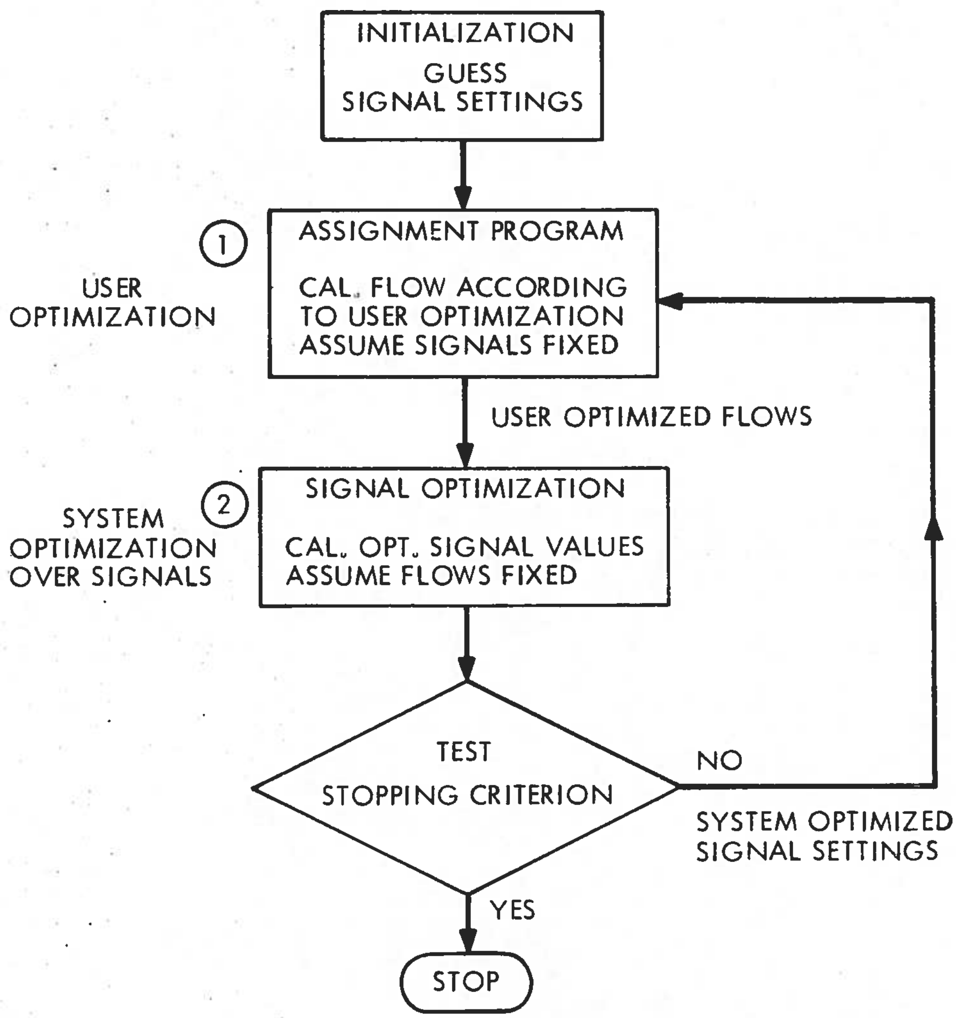


Figure 5.1 Iterative Optimization-Assignment Algorithm

to be fixed.

Signal Optimization Problem

$$\left.\begin{array}{ll}\underset{\underline{g}\in R^{U}}{\text{Minimize}} & J(\underline{h},\underline{g})\end{array}\right|_{\underline{h}\text{ fixed}}$$

subject to

feasibility of controls:

$$\underline{g} \in G \, . \tag{5.11}$$

The procedure is initiated by a guess of the optimal control parameters, $\underline{g}^*$, and continues by iterating between the two programs until a certain stopping criterion is satisfied. For example, it is reasonable to stop the procedure when the change in the magnitudes of successive control vectors is smaller than some small threshold value, ε

This algorithm has been proposed by Allsop [37] as a solution method for a traffic control problem similar to HOP. Applications of this algorithm have been documented in [14], [38], [58]. This algorithm is intuitively appealing since it simulates the real-life situation if the authority periodically updates the optimal signal-timing plan. This process of periodically reoptimizing signal-timing plans has also been considered as a control strategy to account for the changing steady-state traffic pattern due to the redistributional effects of traffic [9], [14].

As a computational algorithm, the iterative optimization assignment algorithm is potentially very attractive for large-scale network applications. This is due to the fact that both the components; i.e., the signal-optimizing program and the assignment program, have been under extensive research, and developed computer software for large-scale networks are available: for example, MITROP [11], TRANSYT [10], SIGOP [12], and TRAFFIC [59].

However, various aspects of this algorithm have not been closely examined. For instance, does the algorithm converge to a solution? If it does converge, what are the properties of the result? Most importantly, is the solution optimal in a specific sense?

It is not the purpose of this report to answer all these questions. In this section, we establish the fact that the optimization assignment algorithm does not necessarily converge to the optimal solution of the HOP by using some numerical examples.

5.3 Numerical Example 1

5.3.1 Description of Problem

Consider the network in Figure 5.2 with link travel time assumed to be a linear function of link flow:

$$t_1(f_1) = 15 + 2f_1 \quad , \tag{5.12}$$

$$t_3(f_3) = 15 + 2f_3 \quad , \tag{5.13}$$

$$t_4(f_4) = 50 + f_4 \quad , \tag{5.14}$$

$$t_5(f_5) = 50 + f_5 \quad . \tag{5.15}$$

Link 2 is under the control of the traffic authority, and for simplicity, we assume that the control is in the form of delay imposed on the traffic passing through link 2. Therefore the travel time function on link 2 is

$$t_2(f_2,w) = 10 + f_2 + w \ , \tag{5.16}$$

where w is the imposed delay. Ten units of traffic flow are required to go from node 1 to node 2. It is assumed that the traffic authority wishes to minimize the total travel time by choosing an appropriate value of w.

From Figure 5.2, there are three paths available for drivers to go from node 1 to node 2. Let h_1, h_2, h_3, be the path flows along routes (1,4,2) (1,3,2), and (1,3,4,2), respectively. The link flows are related to the path flows by the following equations:

$$f_1 = h_2 + h_3 \ , \tag{5.17}$$

$$f_2 = h_3 \quad , \tag{5.18}$$

$$f_3 = h_1 + h_3 \ , \tag{5.19}$$

$$f_4 = h_1 \quad , \tag{5.20}$$

$$f_5 = h_2 \ . \tag{5.21}$$

5.3.2 Solution Obtained by Application of General Nonlinear Constrained Algorithm

The Hybrid Optimization Problem for this example may be stated as follows:

(P1) $$\underset{\{h_1,h_2,h_3,w\},}{\text{Minimize}} \ J = \sum_{i=1}^{5} f_i t_i$$

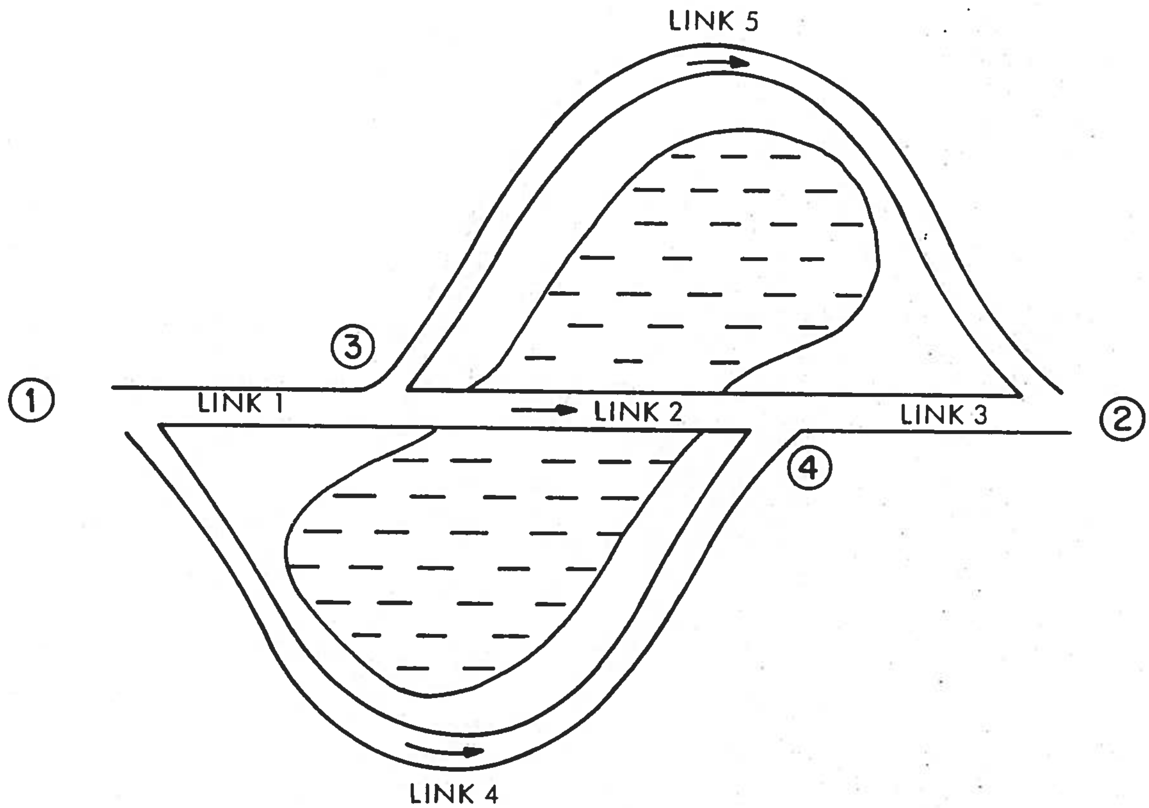


Figure 5.2 Network for Example 1

subject to

$$h_1 + h_2 + h_3 = 10, \tag{5.22}$$

$$h_1 \geq 0 , \tag{5.23}$$

$$h_2 \geq 0 , \tag{5.24}$$

$$h_3 \geq 0 , \tag{5.25}$$

$$\tau_1 = t_3 + t_4 \geq \frac{1}{10} \sum_{i=1}^{5} f_i t_i , \tag{5.26}$$

$$\tau_2 = t_1 + t_5 \geq \frac{1}{10} \sum_{i=1}^{5} f_i t_i , \tag{5.27}$$

$$\tau_3 = t_1 + t_2 + t_3 \geq \frac{1}{10} \sum_{i=1}^{5} f_i t_i , \tag{5.28}$$

$$w \geq 0 . \tag{5.29}$$

Let λ, μ_1, μ_2, μ_3, γ_1, γ_2, γ_3, η be the Lagrange multipliers to constraints (5.22) to (5.29), respectively. The Lagrange function for P1 may be written as

$$\begin{aligned} L = {} & [1 + \frac{1}{10} (\gamma_1 + \gamma_2 + \gamma_3)] \sum_{i=1}^{5} f_i t_i + \lambda (h_1 + h_2 + h_3 - 10) \\ & -\mu_1 h_1 - \mu_2 h_2 - \mu_3 h_3 - \gamma_1 (t_3 + t_4) - \gamma_2 (t_1 + t_5) \\ & -\gamma_3 (t_1 + t_2 + t_3) - \eta w . \end{aligned} \tag{5.30}$$

Listed below are the Kuhn-Tucker necessary conditions for P1.

$$\begin{aligned} \frac{\partial L}{\partial h_1} = {} & [1 + (\gamma_1 + \gamma_2 + \gamma_3)/10] \ (t_3 + 2f_3 + t_4 + f_4) + \lambda - \mu_1 \\ & - 3\gamma_1 - 2\gamma_3 = 0 , \end{aligned} \tag{5.31}$$

$$\begin{aligned} \frac{\partial L}{\partial h_2} = {} & [1 + (\gamma_1 + \gamma_2 + \gamma_3)/10] (t_1 + 2f_1 + t_5 + f_5) + \lambda - \mu_2 \\ & - 3\lambda_2 - 2\gamma_3 = 0 , \end{aligned} \tag{5.32}$$

$$\begin{aligned} \frac{\partial L}{\partial h_3} = {} & [1 + (\gamma_1 + \gamma_2 + \gamma_3)/10] (t_1 + 2f_1 + t_2 + f_2 + t_3 + 2f_3) \\ & + \lambda - \mu_3 - 2\gamma_1 - 2\gamma_2 - 5\gamma_3 = 0 , \end{aligned} \tag{5.33}$$

$$\frac{\partial L}{\partial w} = [1 + (\gamma_1 + \gamma_2 + \gamma_3)/10]\ (f_2) - \gamma_3 - \eta = 0\ , \tag{5.34}$$

$$h_1 + h_2 + h_3 = 10, \tag{5.22}$$

$$h_1 \geq 0\ , \tag{5.23}$$

$$h_2 \geq 0\ , \tag{5.24}$$

$$h_3 \geq 0\ , \tag{5.25}$$

$$\mu_i \geq 0\ , \qquad i = 1,2,3, \tag{5.35}$$

$$\mu_i h_i = 0, \tag{5.36}$$

$$g_1 = t_3 + t_4 - \left(\sum_{i=1}^{5} t_i f_i\right)/10 \geq 0 \tag{5.26}$$

$$g_2 = t_1 + t_5 - \left(\sum_{i=1}^{5} t_i f_i\right)/10 \geq 0 \tag{5.27}$$

$$g_3 = t_1 + t_2 + t_3 - \left(\sum_{i=1}^{5} t_i f_i\right)/10 \geq 0 \tag{5.28}$$

$$\gamma_i \geq 0, \qquad i = 1,2,3, \tag{5.37}$$

$$\gamma_i g_i = 0\ , \qquad i = 1,2,3, \tag{5.38}$$

$$w \geq 0, \tag{5.29}$$

$$\eta \geq 0, \tag{5.39}$$

$$\eta w = 0. \tag{5.40}$$

The Augmented Lagrangian Method has been used as a solution algorithm to the HOP P1. The following optimal solution is obtained.

$$\left.\begin{array}{l} h_1 = h_2 = 5,\quad h_3 = 0,\ w = 20 \\ \lambda = -\ 95 \\ \gamma_1 = \gamma_2 = \gamma_3 = 0 \\ \mu_1 = \mu_2 = 0,\quad \mu_3 = 5 \\ \eta = 0\quad . \end{array}\right\} \tag{5.41}$$

<u>Verification of Extended Equilibrium Principle</u>

Computed below are the trip travel times $\{\tau_i\}$ and pseudo-costs $\{z_i\}$ using

the optimal solution (5.41) along the three available paths from node 1 to node 2 in example 1.

$$\tau_1 = t_3 + t_4,$$

$$= 15 + 2f_3 + 50 + f_4 \;,$$

$$= 15 + (2)(5) + 50 + 5 = 80 \;, \tag{5.42}$$

$$\tau_2 = t_1 + t_5,$$

$$= 15 + 2f_1 + 50 + f_5 \;,$$

$$= 15 + (2)(5) + 50 + 5 = 80, \tag{5.43}$$

$$\tau_3 = t_1 + t_2 + t_3 \;,$$

$$= 15 + 2f_1 + 10 + f_2 + w + 15 + 2f_3 \;,$$

$$= 15 + (2)(5) + 10 + 0 + 20 + 15 + (2)(5) \;,$$

$$= 80 \;, \tag{5.44}$$

$$z_1 = [1 + (\gamma_1 + \gamma_2 + \gamma_3)/10]\ (t_3 + 2f_3 + t_4 + f_4)$$
$$- 3\gamma_1 - 2\gamma_3 = 95, \tag{5.45}$$

$$z_2 = [1 + (\gamma_1 + \gamma_2 + \gamma_3)/10]\ (t_1 + 2f_1 + t_5 + f_5)$$
$$- 3\gamma - 2\gamma_3 = 95 \;, \tag{5.46}$$

$$z_3 = [1 + (\gamma_1 + \gamma_2 + \gamma_3)/10](t_1 + 2f_1 + t_2 + f_2 + t_3 + 2f_3)$$
$$- 2\gamma_1 - 2\gamma_2 - 5\gamma_3 = 100 \;. \tag{5.47}$$

From Eqs. (5.42) to (5.44), the first part of the Extended Equilibrium Principle is satisfied; namely, the Equal Travel-Time Principle.

Paths 1 and 2, each carrying a positive amount of flow (five units), have equal time, $\tau_* = 80$. Path 3, which has no flow, has a travel time not less than τ_*.

The second part of the Extended Equilibrium Principle; i.e., the principle of equalization of pseudo-costs, is also satisified by (5.45) to (5.47). Paths 1 and 2, each carrying five units of flow, have the same pseudo-cost $z_* = 95$.

Path 3, which has a pseudo-cost of $100 > z_* = 95$, carries no flow.

The solution in (5.41) shows $\gamma_1 = \gamma_2 = \gamma_3 = 0$. This implies that [using (4.25) and (4.26)]

$$\theta = (\gamma_1 + \gamma_2 + \gamma_3)/10,$$
$$= 0,$$

and $\underline{\pi} = \theta \underline{f} - \underline{B}\underline{\gamma}$,

$$= 0.$$

Consequently,

$$\psi_i = \frac{\partial J}{\partial f_i} \text{ , using (4.30).}$$

That is, this is a special case where the Hybrid-Optimal solution has a flow distribution which is both system-optimal and user-optimal.

5.3.3 Solution by Iterative Optimization-Assignment Algorithm

Let

$$x = h_3 \ . \tag{5.48}$$

Because of the symmetry,

$$h_1 = h_2 = (10 - x)/2 \ . \tag{5.49}$$

Assignment Program:

Without going through all the algebraic manipulations, the minimization problem in the assignment phase of the iterative optimization reaction algorithm is

(P2)

$$\underset{x}{\text{Minimize}} \quad J(x) = 725 + (w - 20)x + 1.25x^2 \Big|_{w \text{ fixed}}$$

subject to $10 \geq x \geq 0$.

The solution of P2 can be shown to be

$$x = \begin{cases} 8 - 0.4w \ , & 0 \leq w \leq 20 \ , \\ 0 \ , & w > 20 \ . \end{cases} \tag{5.50}$$

Control Optimizing Program

The optimization problem faced by the traffic authority is one of optimizing over the control parameter, w, assuming x to be fixed. Again, we skip all intermediate steps of algebraic manipulations.

(P3).

$$\text{Minimize} \quad xw + 2.5\ [(x-3)^2 + 311]\Big|_{x\ \text{fixed}}$$

$$\text{subject to} \quad w \geq 0 \ .$$

The solution to this problem can be shown to be

$$w \begin{cases} = 0, & \text{for } x > 0 , \\ \geq 0, & \text{for } x = 0 . \end{cases} \tag{5.51}$$

The iterative optimization assignment algorithm consists of the following steps:

Initialization

Set $i = 1$.

Guess w_1 .

Assignment Program (Reaction Phase)

Perform the optimization problem P2 with $w = w_i$.

Let the solution of P2 by x_i .

Control Optimization Problem (Optimization Phase)

Perform the optimization problem P3 with $x = w_i$.

Let the solution of P3 be w_{i+1} .

Stopping Criterion

If $||w_{i+1} - w_i|| > \varepsilon$, go to step 5 ,

stop.

Update

$i \leftarrow i + 1$.

Go to Step 2.

We apply the iterative optimization assignment algorithm to this example with initial guess w = 10. The result of this procedure is shown in Table 5.1, which is constructed using Eqs. (5.50) and (5.51); the converged solution is

$$w = 0 ,$$
$$h_1 = (10-x)/2 = 1,$$
$$h_2 = (10-x)/2 = 1, \tag{5.52}$$
$$h_3 = x = 8 .$$

The solution (5.52) is not the same as that obtained by the HOP formulation!

TABLE 5.1. Result of Iterative Optimization-Assignment Algorithm Applied to Example 1

Step	System (w)	Driver (x)
Initialization	10	
Assignment		4
Control optimization	0	
Assignment		8
Control optimization	0	

Converged Solution: $w = 0$, $x = 8$

We can show that the solution (5.52) is nonoptimal by applying Theorem 4.1: Since $h_i > 0$ for $i = 1,2,3$, therefore by (5.36),

$$u_i = 0, \quad k = 1,2,3. \tag{5.53}$$

The set of necessary conditions (5.31) to (5.34) can be simplified:

$$[1 + (\gamma_1 + \gamma_2 + \gamma_3)/10]\ (103) + \lambda - 3\gamma_1 - 2\gamma_3 = 0 , \tag{5.54}$$

$$[1 + (\gamma_1 + \gamma_2 + \gamma_3)/10]\ (103) + \lambda - 3\gamma_2 - 2\gamma_3 = 0 , \tag{5.55}$$

$$[1 + (\gamma_1 + \gamma_2 + \gamma_3)/10]\ (128) + \lambda - 2\gamma_1 - 2\gamma_2 - 5\gamma_3 = 0, \tag{5.56}$$

$$[1 + (\gamma_1 + \gamma_2 + \gamma_3)/10]\ (8) - \gamma_3 - \eta = 0 . \tag{5.57}$$

Equations (5.54) and (5.55) imply

$$\gamma_1 = \gamma_2 ,$$

and with further simplifications,

$$\gamma_1 \leq 0,$$
$$\gamma_3 = 8\gamma_1 + 50 \geq 0,$$
$$\lambda = -84\gamma_1 - 518,$$
$$\eta = - 2 ,$$

which violates the necessary condition

$$\eta \geq 0 \ . \tag{5.39}$$

Therefore, we can conclude that the solution (5.52) obtained by the iterative optimization assignment algorithm is not optimal from the system's point of view. The nonoptimality of (5.52) can be more explicitly demonstrated by Figure 5.3 which shows the system cost as a function of the imposed delay, w. It is clear that the lowest cost is achieved with $w \geq 20$, and (5.5) implies that the optimal flow on link (3,4) is $x = 0$. That is, it is best not to use this link, and there should be a sufficiently large control delay imposed so that all drivers are discouraged from traveling on it. This is thus an example of Braess' paradox [20]. The optimization assignment algorithm, however, has done the opposite and converged to the worst possible solution.

5.4 Example 2

5.4.1 Problem Description

Consider the network as shown in Figure 2.1. Node 3 is a signalized intersection. We denote the green split facing link (1,3) by g. The link travel time is modeled as a fourth power polynomial of the link flow as discussed in Section 2.

The capacity is taken to be 1500 veh/hr per lane. All links except link (4,5) are assumed to be single lane. We use the Webster formula (2.22) with fixed-cycle time of 1.0 minute for the waiting time at the signalized intersection. The traffic demands are 800 veh/hr for each of the following O-D pairs: 1 to 5, 1 to 6, 2 to 5, and 2 to 6. Table 5.2 lists all the routes available between these O-D pairs. It is assumed that the cost to be minimized is the total travel time in this example.

TABLE 5.2. All Available Routes for Example 2

O-D Pair	1	2	3	4
Origin node	1	1	2	2
Destination node	5	6	5	6
Total No. of Paths	2	3	2	3
Paths	$R_1^1 = (1,4,5)$ $R_2^1 = (1,3,4,5)$	$R_1^2 = (1,4,5,6)$ $R_2^2 = (1,3,6)$ $R_3^2 = (1,3,4,5,6)$	$R_1^3 = (2,3,4,5)$ $R_2^3 = (2,5)$	$R_1^4 = (2,3,4,5,6)$ $R_2^4 = (2,5,6)$ $R_3^4 = (2,3,6)$

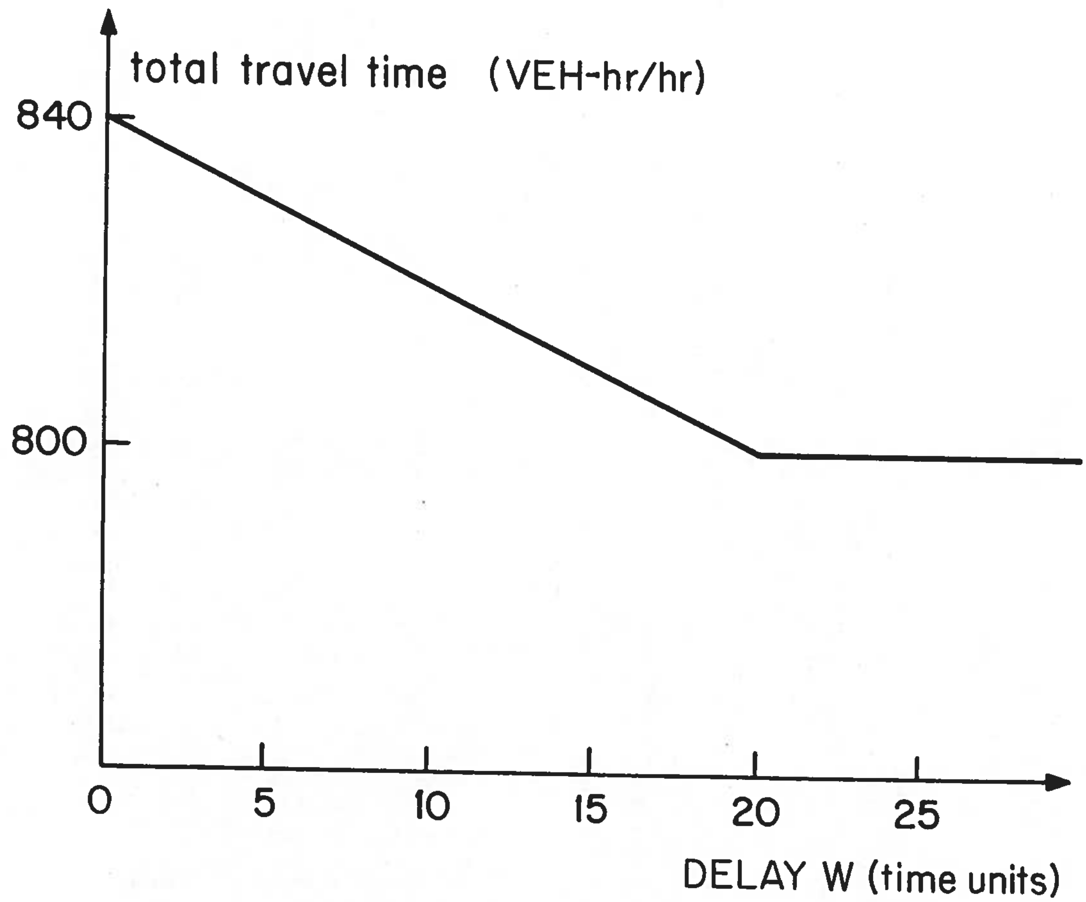


Figure 5.3 Total Travel Time (at Equilibrium) of Example 1 as Function of w

5.4.2 Solution by HOP Formulation and Augmented Lagrangian Method

As in the previous example, we have treated this network by two different approaches. First, it is solved using the HOP formulation by the Augmented Lagrangian Method. The optimal solution obtained is $g^* = 0.215$ with the minimum cost (total travel time) equal to 880.72 veh-hr/hr. The optimality of g^* is verified using the results of a series of user-optimized flow patterns computed at various values of g, the green split. The associated system cost in Figure 5.4, demonstrates the optimality of g^*.

Listed below are the optimal solution and the Lagrange Multipliers of the HOP.

$g^* = 0.215$.

O-D Pair Number 1

Flows: $h_1^1 = 800.00$, $h_2^1 = 0.00$.

Lagrange Multipliers:

$\lambda_1 = -0.255205$,

$\gamma_1^1 = 0.056212$, $\gamma_2^1 = 0.0$,

$\mu_1^1 = 0.0$, $\mu_2^1 = 0.0181421$.

O-D Pair Number 2

Flows: $h_1^2 = 505.72$, $h_2^2 = 294.28$, $h_3^2 = 0.0$.

Lagrange Multipliers:

$\lambda_2 = -0.44$,

$\gamma_1^2 = 58.0864$, $\gamma_2^2 = 47.5195$, $\gamma_3^2 = 0.0$,

$\mu_1^2 = 0.0$, $\mu_2^2 = 0.0$, $\mu_3^2 = 0.041376$.

O-D Pair Number 3

Flows: $h_1^3 = 222.30$, $h_2^3 = 577.70$,

Lagrange Multiplier:

$\lambda_3 = -218.417$,

$\gamma_1^3 = 1.5203 \times 10^5$, $\gamma_2^3 = 3.9139 \times 10^5$

$\mu_1^3 = 0.0$, $\mu_2^3 = 0.0$.

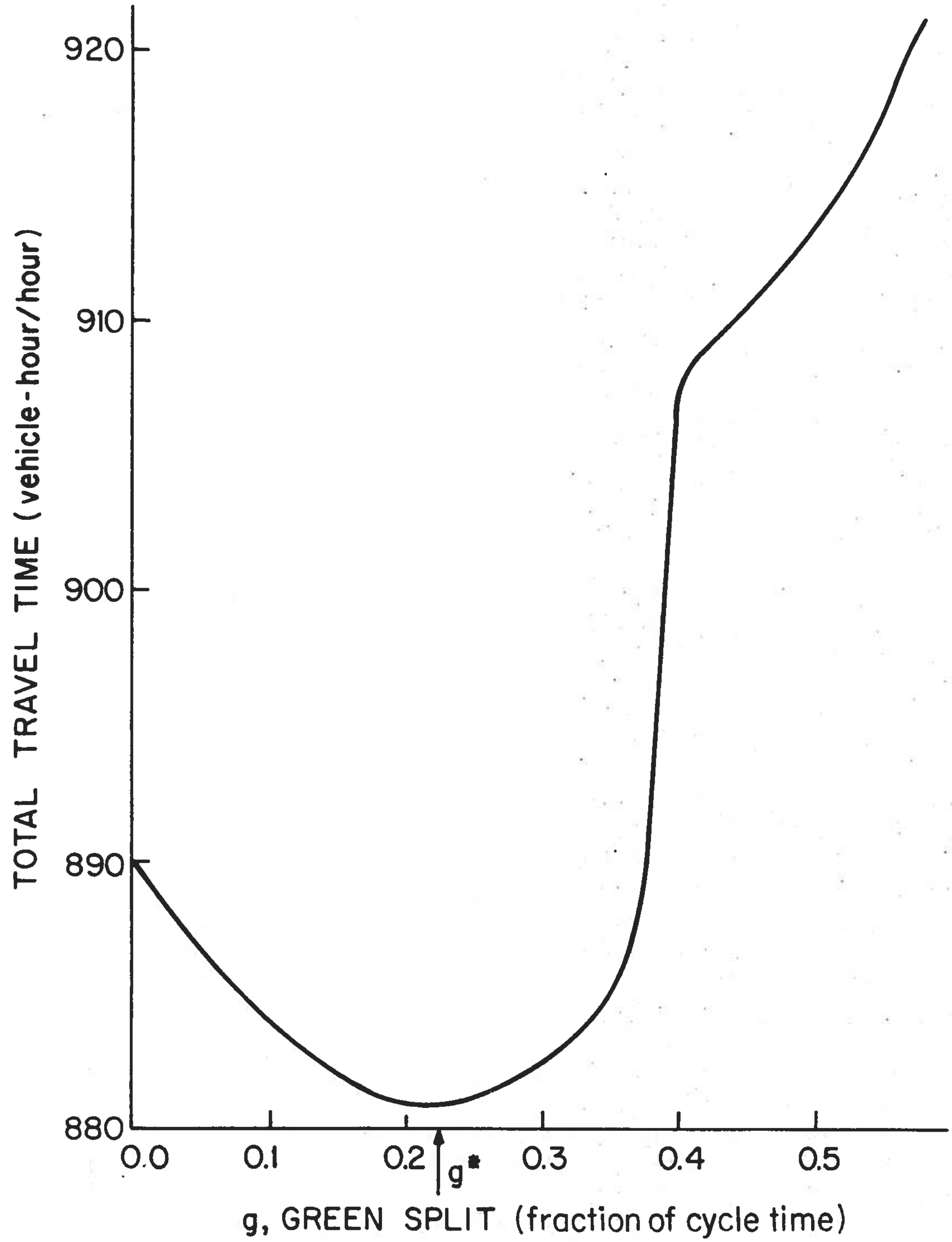


Figure 5.4 Total Travel Time (at Equilibrium) of Example 2 as Function of g

O-D Pair Number 4

Flows: $h_1^4 = 0.0$, $h_2^4 = 0.0$ $h_3^4 = 800.00$,

Lagrange Multipliers:

$$\lambda_4 = -0.3349 ,$$

$$\gamma_1^4 = 0.0, \quad \gamma_2^4 = 0.0, \quad \gamma_3^4 = 2.346 \times 10^{-9} ,$$

$$\mu_1^4 = 0.01814, \quad \mu_2^4 = 0.01814, \quad \mu_3^4 = 0.0 .$$

The link travel times, t_i, and the link pseudo-costs, ψ_i^k, for each O-D pair are computed and summarized in Table 5.3 below:

TABLE 5.3. Link Times and Pseudo-Costs for Example 2

Link	Travel Time	Pseudo-Cost			
(i,j)	t_{ij} (min)	$\psi^1_{i,j}$	$\psi^2_{i,j}$	$\psi^3_{i,j}$	$\psi^4_{i,j}$
(1,3)	4.80	8.84187×10^{-2}	1.41757×10^{-1}	5.44463×10^{1}	8.84131×10^{-2}
(2,3)	8.37	1.47846×10^{-1}	2.40787×10^{-1}	9.48651×10^{1}	1.47837×10^{-1}
(1,4)	3.91	9.02578×10^{-2}	1.33694×10^{-1}	4.43567×10^{1}	9.02532×10^{-2}
(3,4)	1.20	1.99820×10^{-2}	3.33136×10^{-1}	1.36064×10^{1}	1.99806×10^{-2}
(4,5)	9.70	1.64946×10^{-1}	2.72668×10^{-1}	1.09946×10^{2}	1.64935×10^{-1}
(2,5)	19.26	3.32775×10^{-1}	5.46769×10^{-1}	2.18417×10^{2}	3.32752×10^{-1}
(3,6)	10.01	1.87067×10^{-1}	2.98243×10^{-1}	1.13488×10^{2}	1.87055×10^{-1}
(5,6)	1.20	2.02808×10^{-2}	3.36372×10^{-2}	1.36321×10^{1}	2.02793×10^{-2}

In Table 5.4, we show the total travel time and pseudo-cost along every path for all O-D pairs.

The results in Table 5.4 clearly demonstrate the validity of the Extended Equilibrium Principle.

In obtaining the results in Figure 5.4, we also have computed the associated total fuel consumption for each equilibrium flow pattern at different values of g. This is shown in Figure 5.5. It is clear that the optimal value of g is 0.0 if fuel consumption is the system cost to be minimized. A comparison of Figures 5.4 and 5.5 also suggests the conclusion that a set of control parameters optimal with respect to a system cost of total travel time may not be optimal with respect to a system cost of total fuel consumption, and vice versa.

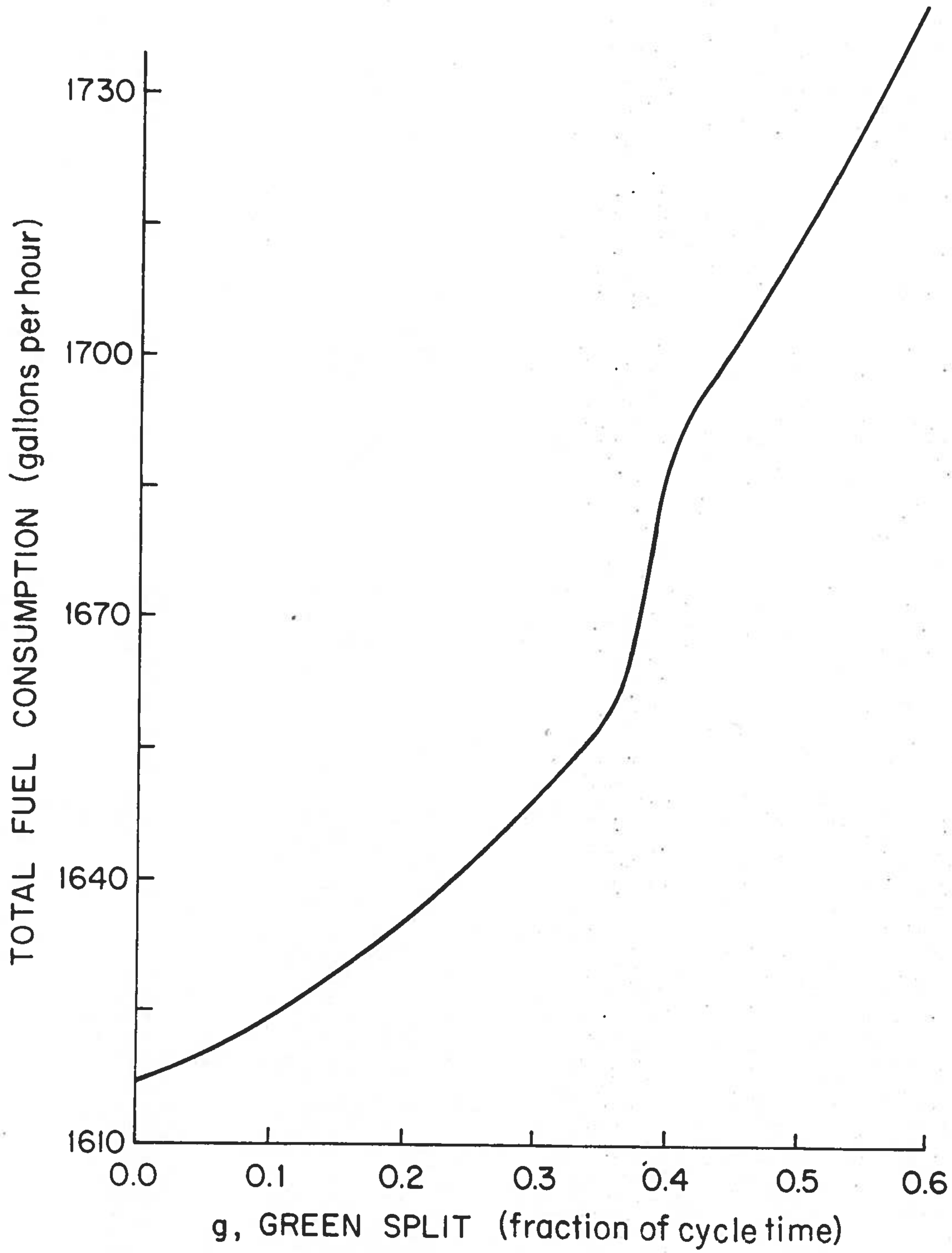


Figure 5.5 Total Fuel Consumption (at Equilibrium) of Example 2 as Function of g

TABLE 5.4. Path Times and Pseudo-Costs for Example 2

O-D Pair	Path	Flow	Time (min)	Pseudo-Cost
1	1	800.00	13.61	2.55204×10^{-1}
1	2	0.0	15.70	2.73347×10^{-1}
2	1	505.72	14.81	4.39999×10^{-1}
2	2	294.28	14.81	4.40000×10^{-1}
2	3	0.0	16.90	4.81376×10^{-1}
3	1	222.30	19.26	2.18417×10^{2}
3	2	577.70	19.26	2.18417×10^{2}
4	1	0	20.47	3.53031×10^{-1}
4	2	0	20.47	3.53032×10^{-1}
4	3	800.00	18.37	3.34892×10^{-1}

Figures 5.4 and 5.5 exhibit several interesting and illuminating points. Valuable insights into the behavior of system cost of a function of the control parameter can be gained from a close examination of these results. Tables 5.5 to 5.8 show the equilibrium flow patterns and system cost (total travel time) at various values of g. Figure 5.4 exhibits two interesting features: the rapid increase and nondifferentiality at around g = 0.36. We divide the g axis into three regions for the purpose of discussion. In region 1, for g = 0.0 to g = 0.35, we notice that the set of active paths (i.e., paths carrying positive flows) consists of R_1^1, R_1^2, R_2^2, R_1^3, R_2^3, where R_j^k denotes the j^{th} path of the k^{th} O-D pair. In region 2, $(0.35 \leq g \leq .365)$, path R_1^3, disappears, and path R_2^4 is introduced. In region 3 $(0.365 \leq g \leq 0.6)$, the active paths are R_1^1, R_1^2, R_2^2, R_2^3, R_2^4, and R_3^4. In each region, the system cost is a function of the flows on the active paths, and depends solely on the path cost characteristics of these paths. Consequently, the system cost is not the same function in different regions, since the sets of active paths are not the same in different regions. Therefore, it should be no surprise that the system cost is not differentiable at the transition points where some active paths disappear and new paths are introduced.

The rapid increase in the system cost at around g = 0.36 can also be explained if we examine closely the equilibrium flow patterns in this interval. Tables 5.5 to 5.8 show the equilibrium flow patterns in detail from g = 0.36 to g = 0.3644. The equilibrium flow pattern is shown to remain almost constant

TABLE 5.5. Equilibrium Flows for Example 2 (O-D Pair Number 1)

Green Split	Path 1 (1, 4, 5)		Path 2 (1, 3, 4, 5)		System Cost
	Flow (veh/hour)	Time (min)	Flow (veh/hour)	Time (min)	
0.10	800.00	13.76	0.00	16.03	881.59
0.20	800.00	13.64	0.00	15.77	880.77
0.30	800.00	13.47	0.00	15.31	882.22
0.36	800.00	13.39	0.00	15.08	885.03
0.3637	800.00	13.40	0.00	15.06	886.16
0.3638	800.00	13.40	0.00	15.06	888.92
0.3639	800.00	13.40	0.00	15.06	891.94
0.3640	800.00	13.40	0.00	15.06	895.19
0.3641	800.00	13.40	0.00	15.06	898.74
0.3642	800.00	13.40	0.00	15.06	902.52
0.3640	800.00	13.40	0.00	15.06	906.62
0.3644	800.00	13.40	0.00	15.06	907.43
0.40	800.00	13.36	0.00	15.03	908.49
0.45	800.00	13.32	0.00	14.99	910.56
0.50	800.00	13.92	0.00	14.96	913.82
0.55	800.00	13.27	0.00	14.94	918.27
0.60	800.00	13.25	0.00	14.85	922.53

TABLE 5.6. Equilibrium Flows for Example 2 (O-D Pair Number 2)

Green Split	Path 1 (1,4,5,6)		Path 2 (1,3,6)		Path (1,3,4,5,6)	
	Flow (veh/hour)	Time (min)	Flow (veh/hour)	Time (min)	Flow (veh/hour)	Time (min)
0.1000	597.10	14.97	202.90	14.97	0.00	17.24
0.2000	526.94	14.84	273.06	14.84	0.00	16.98
0.3000	393.64	14.67	406.36	14.67	0.00	16.51
0.35	334.85	14.61	465.15	14.61	0.00	16.31
0.36	324.18	14.60	475.82	14.60	0.00	16.28
0.3637	320.47	14.60	479.54	14.60	0.00	16.26
0.3638	320.46	14.60	479.54	14.60	0.00	16.26
0.3639	320.46	14.60	479.54	14.60	0.00	16.26
0.3640	320.46	14.60	479.54	14.60	0.00	16.26
0.3641	319.88	14.60	480.12	14.60	0.00	16.26
0.3642	319.88	14.60	480.12	14.60	0.00	16.26
0.3643	319.88	14.60	480.12	14.60	0.00	16.26
0.3644	319.88	14.60	480.12	14.60	0.00	16.26
0.4000	269.03	14.56	530.97	14.56	0.00	16.23
0.45	197.48	14.52	602.52	14.52	0.00	16.19
0.50	125.37	14.49	674.63	14.49	0.00	16.16
0.55	52.84	14.47	747.16	14.47	0.00	16.14
0.60	00.00	14.45	800.00	14.33	0.00	16.05

TABLE 5.7. Equilibrium Flows for Example 2 (O-D Pair Number 3)

Green Split	Path 1 (2, 3, 4, 5)		Path 2 (2, 5)	
	Flow (veh/hour)	Time (min)	Flow (veh/hour)	Time (min)
0.1000	320.24	19.23	479.76	19.23
0.2000	245.31	19.25	554.69	19.25
0.3000	95.42	19.34	704.58	19.34
0.3500	20.48	19.41	779.52	19.41
0.3600	5.50	19.43	794.50	19.43
0.3637	0	19.50	800.00	19.43
0.3638	0	19.50	800.00	19.43
0.3639	0	19.50	800.00	19.43
0.3640	0	19.50	800.00	19.43
0.3641	0	19.50	800.00	19.43
0.3640	0	19.50	800.00	19.43
0.3643	0	19.50	800.00	19.43
0.3644	0	19.50	800.00	19.43
0.40	0	21.18	800.00	19.50
0.45	0	21.29	800.00	19.62
0.50	0	21.45	800.00	19.78
0.55	0	21.64	800.00	19.97
0.60	0	21.93	800.00	20.21

TABLE 5.8. Equilibrium Flow for Example 2 (O-D Pair Number 4)

Green Split	P a t h (2,3,4,5,6)		P a t h 2 (2,5,6)		P a t h 3 (2,3,6)	
	Flow (veh/hour)	Time (min)	Flow (veh/hour)	Time (min)	Flow (veh/hour)	Time (min)
0.1000	0.00	20.43	0.00	20.43	800.00	18.16
0.2000	0.00	20.46	0.00	20.46	800.00	18.32
0.3000	0.00	20.54	0.00	20.54	800.00	18.67
0.3500	0.00	20.61	0.00	20.61	800.00	18.91
0.36	0.00	20.63	0.00	20.63	800.00	19.95
0.3637	0.00	20.70	0.00	20.63	800.00	19.33
0.3638	0.00	20.91	0.00	20.63	800.00	19.24
0.3639	0.00	21.13	0.00	20.63	800.00	19.47
0.3640	0.00	21.38	0.00	20.63	800.00	19.72
0.3641	0.00	21.64	0.00	20.63	800.00	19.98
0.3642	0.00	21.92	0.00	20.63	800.00	20.26
0.3643	0.00	22.24	0.00	20.63	800.00	20.57
0.3644	0.00	22.30	0.12	20.63	799.88	20.64
0.4000	0.00	22.38	53.49	20.70	746.51	20.70
0.45	0.00	22.49	128.44	20.82	671.56	20.82
0.50	0.00	22.65	203.37	20.98	596.63	20.98
0.55	0.00	22.84	278.30	21.17	521.70	21.17
0.60	0.00	23.13	353.19	21.41	446.81	21.41

in this interval. However, the path time on R_3^4 rises rapidly from 18.95 to 20.64 minutes in this very short interval, resulting in a rapid increase in system cost. The reason behind this lies in the fact that link (2,3), which faces the signal, is heavily congested. The effective capacity on this link around g = 0.36 is (1500) (0.9-g) = 810, while the link flow is 800. The degree of saturation is 800/810 = 0.988. With this level of flow so close to saturation, the waiting time in the Webster formula, which has a term,

$$\frac{1}{1500\ (0.9 - g) - f} \quad ,$$

is in a region of rapid increase in magnitude. This explains the behavior of the system cost at g = 0.36.

5.4.3 Solution by the Iterative Optimization Assignment Algorithm

In the second method, we have applied the iterative optimization assignment algorithm. An initial value of 0.6 is assumed for the green split, g. Figure 5.6 shows the values of g from iteration to iteration. The final converged solution is g = 0.0, which is not the optimal solution of the Hybrid Optimization Problem; this is clearly shown in Figure 5.4.

5.5 Example 3

5.5.1 Problem Description

In this example, we consider the same traffic network used in example 2. The traffic demands for O-D pairs 1 to 5 and 1 to 6 are increased from 800 to 1200 veh/hr. All other parameters remain the same.

Both the Energy optimization and travel-time optimization problems are solved using two different approaches, as in the previous examples.

5.5.2 Solution by HOP Formulation and Augmented Lagrangian Method

Table 5.9 shows the optimal solution of the HOP using total travel time as the system cost. The optimal solution of the HOP with total fuel consumption as the system cost is displayed in Table 5.10.

The Optimality of the results shown in Tables 5.9 and 5.10 is verified by Figures 5.7 and 5.8, where the total travel time and total fuel consumption are evaluated at each user-optimized flow at various values of g.

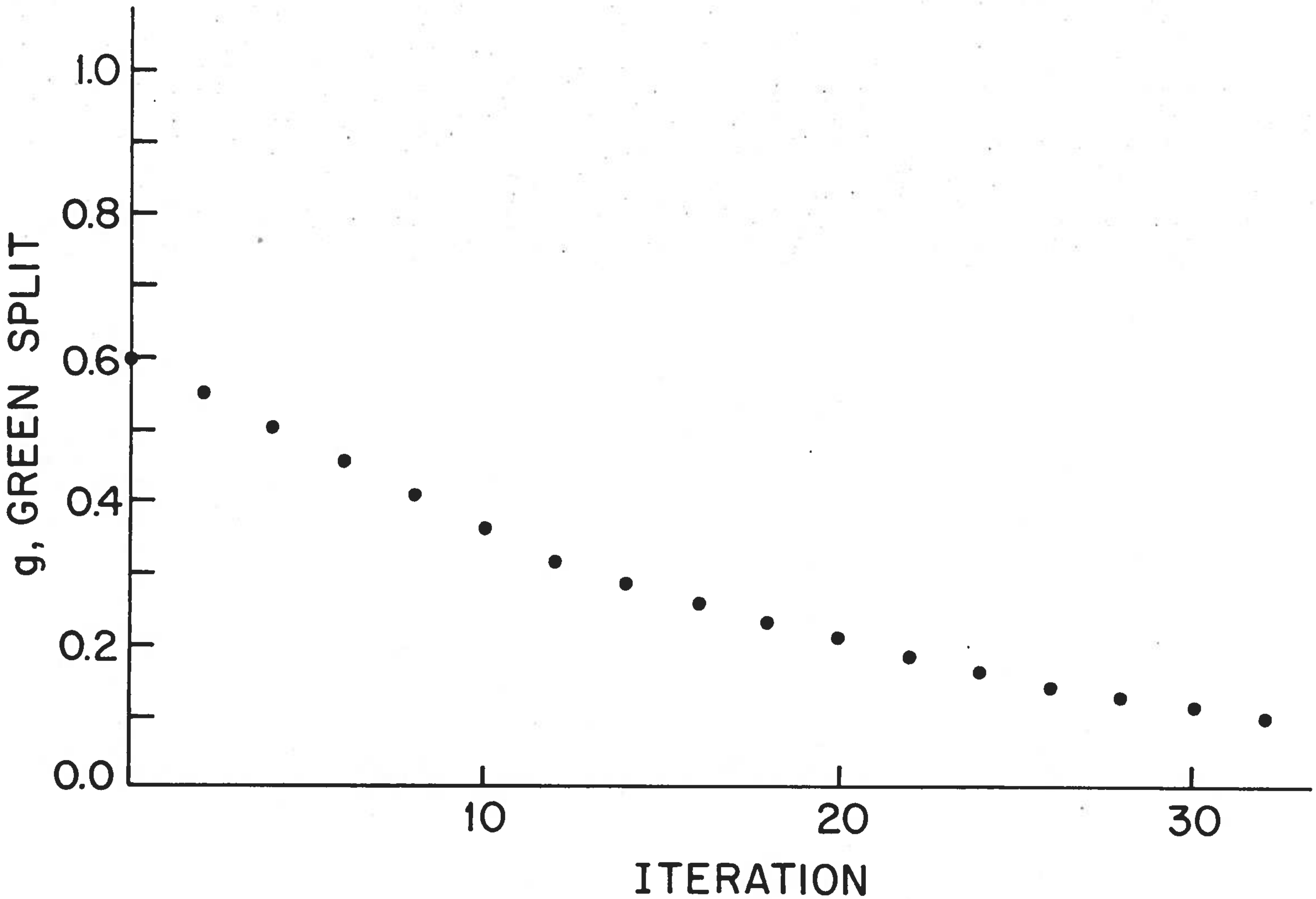


Figure 5.6 Convergence Behavior of Iterative Optimization-Assignment Algorithm (Non-Optimal) for Example 2

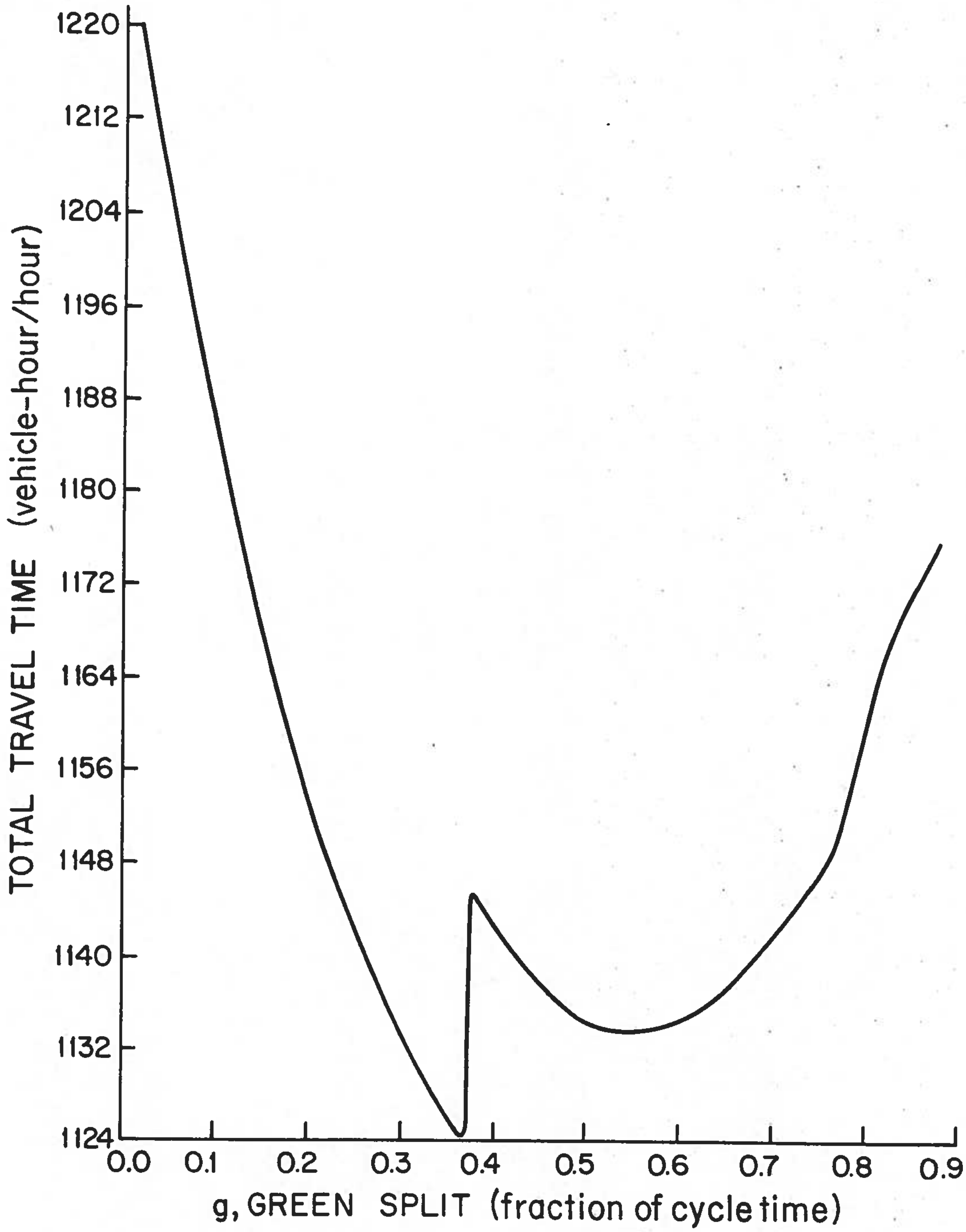


Figure 5.7 Dependence of Total Travel Time on Green Split for Example 3

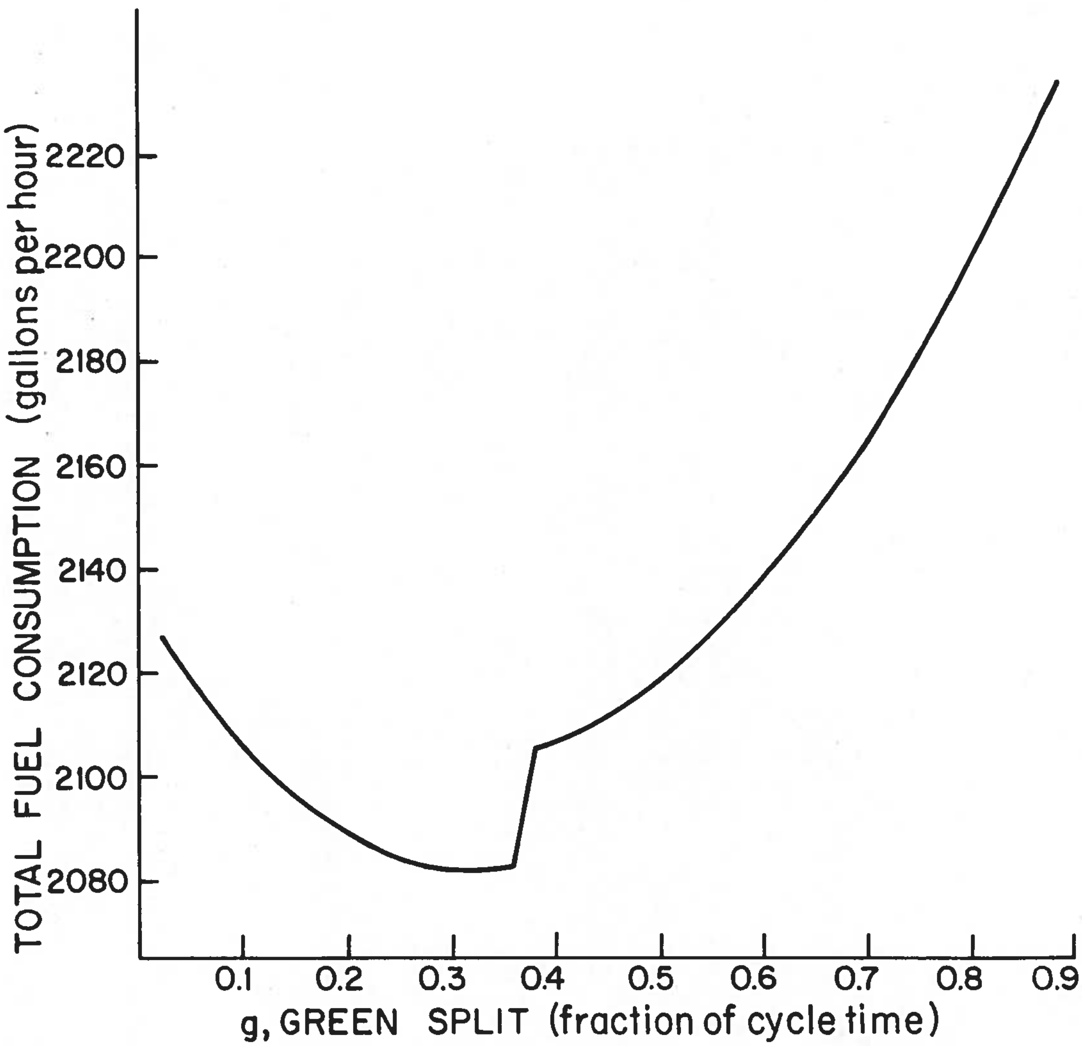


Figure 5. 8 Dependence of Fuel Consumption on Green Split for Example 3

TABLE 5.9. Optimal Solution of HOP of Example 3 with Total Travel Time as System Cost

O-D Pair	Path	Flow	Time (min)	Pseudo-Cost
1	1	1200	14.73	0.36688
	2	0	16.47	0.37079
2	1	671.444	15.93	0.57675
		528.556	15.93	0.57675
		0	17.67	0.60127
3	1	0	19.44	0.74265
	2	800	19.43	0.67709
4	1	0	20.64	1.17241
	2	0	20.64	1.17240
	3	800	18.90	1.10690

Optimal Total Travel Time = 1124.3244 vehicle-hour/hour.

Optimal Green Split, g* =0.3636.

TABLE 5.10. Optimal Solution of HOP of Example 3 with Total Energy as System Cost

O-D Pair	Path	Flow	Time (min)	Pseudo-Cost
1	1	1200	14.96	0.59554
	2	0	16.92	0.65313
2	1	731.164	16.17	0.77979
	2	468.836	16.17	0.77979
	3	0	18.13	0.85453
3	1	62.83	19.37	1.01776
	2	737.17	19.37	1.01776
4	1	0	20.58	0.71078
	2	0	20.58	0.71080
	3	800	18.62	0.65326

Optimal Energy Cost = 2082.229 gallons/hour.

Optimal Green Split, g* =0.3216.

Figures 5.7 and 5.8 further strengthen our previous observations on the behavior of the cost as a function of the control parameters, namely, the non-differentiability and rapid increase. Figure 5.7 reveals yet another feature of the behavior of the cost: there may be more than one local minimum. Figure 5.7 shows that there are two local minima: at $g \cong 0.36$ and $g \cong 0.55$.

5.5.3 Solution by Iterative Optimization-Assignment Algorithm

In the second method, we have applied the iterative optimization assignment algorithm. Note that the problem is independent of whether total travel time or total fuel consumption is used as the system cost.

This is due to the fact that the system cost function only appears in the phase of control parameter optimization of the algorithm. We recall from Section 2, for the cost of total travel time optimization, the cost can be written as

$$J_t = \sum_{i=1}^{NL} f_i t_i(\underline{f},\underline{g}). \tag{5.40}$$

For the cost of total fuel consumption, the cost is

$$J_e = \sum_{i=1}^{NL} f_i (k_1 d_i + k_2 t_i(\underline{f},\underline{g})),$$

$$= k_1 \sum_{i=1}^{NL} d_i f_i + f_2 \sum_{i=1}^{NL} f_i t_i(\underline{f},\underline{g}) . \tag{5.41}$$

However in the control optimization phase of the Iterative Optimization-Assignment Algorithm, link flows, $\{f_i\}$, are assumed fixed. Hence, the system cost of total fuel consumption, J_e, is an affine function of J_t, i.e.,

$$J_e = (\text{constant}_1) + (\text{constant}_2) J_t .$$

Hence in the control optimization phase, minimizing J_e is exactly the same as minimizing J_t.

We have applied the iterative optimization assignment algorithm for example 3. An initial value of 0.6 is assumed for the green split, g. Figure 5.9 shows the value of g from iteration to iteration. The final converged solution is g = 0.0, which is not the optimal solution of the HOP with either total travel time or total fuel consumption as the cost; this is clearly shown in Figures 5.7 and 5.8.

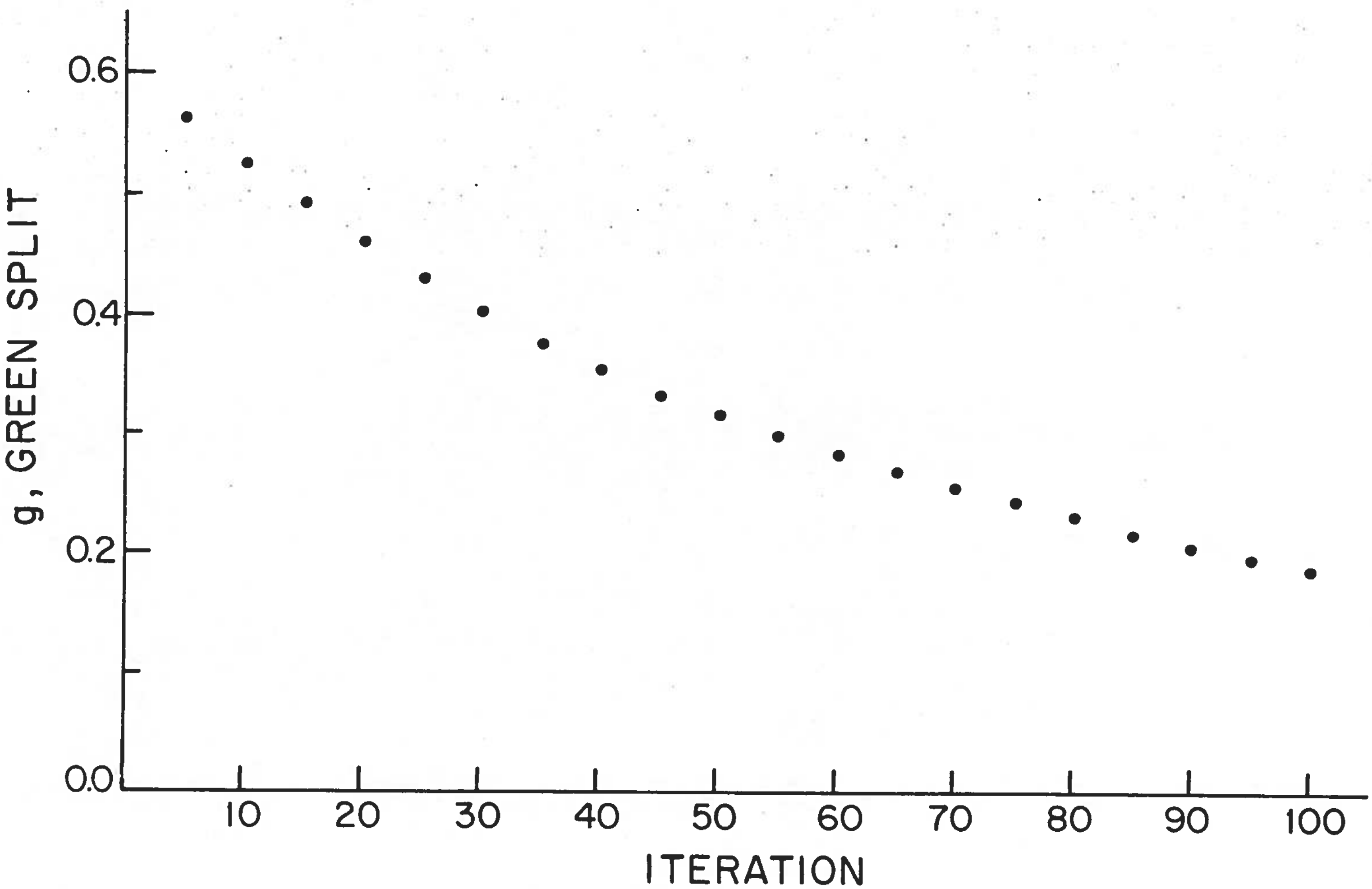


Figure 5.9 Behavior of Green Split Using Non-Optimal Iterative Optimization-Assignment Algorithm for Example 3

5.6 Conclusions

In this section, we have employed three simple numerical examples to illustrate the concept of hybrid optimization in transportation systems. We also have studied an intuitively appealing algorithm, the optimization-assignment algorithm, as a possible algorithm that may be applicable to solution of Hybrid Optimization Problems for large networks. From the results of these examples, it is shown that this algorithm leads to very wrong solutions in all cases.

A practical implication is that if this procedure is actually implemented as mentioned in [9], [14], and [15], the amount of excess cost accrued over a long period of operation can be substantial. The main problem with the optimization assignment algorithm is that, in the control optimization phase, controls are chosen without anticipating the reaction of the drivers. More generally, this section demonstrates that proper evaluation of any control policy cannot be made without taking into consideration the drivers' reactions.

The results of the three numerical examples also demonstrate that the formulation of the Hybrid Optimization Problem is indeed well defined. However, it must be cautioned that the optimization method used for HOP in this section is not practically applicable to large networks. Although the augmented Lagrange multiplier method has been shown to be an efficient algorithm, problems of higher dimension have seldom been solved successfully by straightforward applications of this method to date. In the next section, we consider a special structure of the HOP that may be exploited in developing algorithms for large networks.

6. PROPOSED ALGORITHMS FOR LARGE-SCALE NETWORKS

In section 5, we have presented three numerical examples and their solutions obtained by two different approaches. We have shown that the Iterative Optimization Assignment Algorithm leads to wrong solutions in both cases. On the other hand, the approach of using the HOP formulation and the Augmented Lagrangian Method is shown to yield the correct solutions.

However, the examples considered so far are relatively small in size. It is not apparent at this stage that such an approach can be applied to problems involving larger networks. There are several reasons that such an approach cannot be directly applied to these cases. In the HOP formulation, it is assumed that all paths between all O-D pairs are available. Therefore, it is required that all paths be generated before the problem can be solved by the Augmented Lagrangian Method. However, this step in itself is nontrivial even for small networks, and prohibitive for large networks, since the number of paths grows rapidly with the size of the network.

Furthermore, even if all paths are available, the resulting HOP would, most likely, be intractable because of the large number of variables. Therefore, for algorithms to be applicable to large networks, it is necessary that no a priori generation of all paths joining each origin-destination pair is required. Furthermore, the resulting optimization problem must be relatively small in size.

In this section, we propose several algorithms which satisfy these requirements by exploiting the special structures of the Hybrid Optimization Problem. In section 6.1, we consider an algorithm which iteratively generates paths by making use of the Extended Equilibrium Principle; in each iteration, the minimization of the HOP is carried out over a subset of paths only. In section 6.2, we consider a solution method similar to that used by Abdulaal, et al. [42] in solving the Equilibrium Network Design Problem. In this method, only the traffic control parameters are considered as decision variables. The flow vector, which is constrained to be an equilibrium flow, is considered to

be a function of the control variables. Using this approach, the dimensionality of the problem is greatly reduced to the number of control parameters.

In section 6.3, we provide a procedure for computing the upper and lower bounds of the optimal cost of the Hybrid Optimization Problem. We also conjecture that the optimal control parameters in the solution to the Generalized System Optimization Problem can be used as an approximate solution to the HOP.

6.1 Algorithm That Makes Use of Extended Equilibrium Principle

This algorithm is motivated by the observation that in large-scale networks, the number of active paths; i.e., paths carrying positive flow, in an equilibrium flow pattern is quite small relative to the total number of paths. See, for example, Gershwin, et al. [15]. This observation is also corroborated by Leventhal, et al. [33], on equilibrium flow patterns in large-scale networks. In that study, it is shown that for a network of 16 nodes, 48 links, and 10 O-D pairs, the number of active paths is less than 0.1 percent of the total number of paths; and for another network with 64 nodes, 224 links, and 5 O-D pairs, the number of active paths is less than 0.01 percent of the total number of paths.

Let Q_k^* be the set of active paths for the k^{th} O-D pair in an optimal solution of the HOP. That is,

$$Q_k^* \triangleq \{i | i \varepsilon P_k,\ h_i^k > 0 \text{ in a given optimal solution of the HOP}\},$$
$$k = 1,\ldots,K. \tag{6.1}$$

Let $\overline{Q_k^*}$ be the complement of Q_k^* relative to P_k. Clearly, if $\{Q_k^*\}$ is known, it would be sufficient to solve the HOP over paths in $\{Q_k^*\}$ only. In the algorithm proposed in this section, at every iteration i, the minimization of HOP is carried out over all paths in Q_k^i, where Q_k^i is a subset of P_k. At the end of iteration i, we make use of the Extended Equilibrium Principle to verify if $Q_k^i \supset Q_k^*$. If this is not true, new paths are generated by making use of the Extended Equilibrium Principle. This is described in detail as follows.

The Algorithm

STEP 1. Initialization

Start with an initial guess $\{Q_k^\circ\}$. Let $i = 0$.

STEP 2. Restricted HOP

For the i^{th} iteration, solve a restricted HOP.

HOP^i

$$\underset{\underline{h},\underline{g}}{\text{Minimize}}\ J(\underline{f},\underline{g})$$

subject to

$$\sum_{j \varepsilon Q_k^i} h_j^k = H^k , \qquad k = 1,\ldots,K, \qquad (6.2)$$

$$h_j^k \geq 0 \quad j \varepsilon Q_k^i , \qquad k = 1,\ldots,K, \qquad (6.3)$$

$$\underline{B}^{k'}\underline{t}(\underline{f},\underline{g}) \geq [\underline{f}^{k'}\underline{t}(\underline{f},\underline{g})/H^k]\underline{u}^k , \qquad k = 1,\ldots,K, \qquad (6.4)$$

$$0.9 \geq g_j \geq 0 , \qquad j = 1,\ldots,U. \qquad (6.5)$$

Let $(\underline{h},\underline{g})$ be an optimal solution and $\underline{\gamma}^k$ be the Lagrange multipliers of (6.4).

STEP 3. Verification of Optimality

3.1 Compute the shortest path for every O-D pair using the link travel time, $t_j(\underline{f},\underline{g})$, as the length of arc j, for every j. Let r_k^i be the shortest path for the k^{th} O-D pair.

3.2 Compute the shortest path for every O-D pair using the link pseudo-cost, $\psi_j^k\ (\underline{f},\underline{\gamma}^k,\underline{g})$, as the length of arc j, for every j and k. Let s_k^i be the shortest path for the k^{th} O-D pair.

3.3 If $r_k^i \varepsilon Q_k^i$ and $s_k^i \varepsilon Q_k^i$ for all k, then the necessary conditions are satisfied. Stop. Otherwise, proceed to Step 4.

STEP 4. Updating of Active Path Set

If $r_k^i \varepsilon \overline{Q}_k^i$, then $Q_k^{i+1} \leftarrow Q_k^i \cup r_k^i$,

$$\text{Otherwise,} \qquad Q_k^{i+1} \leftarrow Q_k^i, \qquad k = 1,\ldots,K. \tag{6.6}$$

If $s_k^i \varepsilon \overline{Q}_k^i$, then $Q_k^{i+1} \leftarrow Q_k^{i+1} \cup s_k^i$.

$$\text{Otherwise,} \qquad Q_k^{i+1} \leftarrow Q_k^{i+1}, \qquad k = 1,\ldots,K, \tag{6.7}$$

$i \leftarrow i + 1$. Go to Step 2.

At the end of Step 2, since $(\underline{h},\underline{g})$ is an optimal solution of HOP^i, then the Extended Equilibrium Principle is satisfied for all paths in Q_k^i. In Step 3, we check to see if the Extended Equilibrium Principle is satisfied for all paths P_k. If both r_k^i and s_k^i are in Q_k^i, then we observe that the Extended Equilibrium Principle is satisfied for all paths in P_k. This can be be shown as follows.

Since $(\underline{h},\underline{g})$ is an optimal solution of HOP^i, therefore the Extended Equilibrium Principle is satisfied for all paths in $\{Q_k^i\}$. That is,

$h_j^k > 0$ implies that

$$\tau_j^k(\underline{h},\underline{g}) = x^k \overset{\Delta}{=} \min_{n\varepsilon Q_k^i} \tau_n^k(\underline{h},\underline{g}), \tag{6.8}$$

$$z_j^k(\underline{h},\underline{g},\underline{\gamma}^k) = y^k \overset{\Delta}{=} \min_{n\varepsilon Q_k^i} z_n^k(\underline{h},\underline{g},\underline{\gamma}^k), \tag{6.9}$$

and

$$\tau_j^k(\underline{h},\underline{g}) > x^k \text{ implies that } h_j^k = 0, \tag{6.10}$$

$$z_j^k(\underline{h},\underline{g},\underline{\gamma}^k) > y^k \text{ implies that } h_j^k = 0, \tag{6.11}$$

for all $j \varepsilon Q_k^i$, $k = 1,\ldots,K$.

We recall that τ_j^k and z_j^k are, respectively, the time and pseudo-cost along path j of the k^{th} O-D pair. But r_k and s_k are in Q_k^i . This implies that

$$x^k = \tau_{min}^k \triangleq \min_{n \in P_k} \tau_n^k(\underline{h},\underline{g}) \quad , \tag{6.12}$$

and

$$y^k = z_{min}^k \triangleq \min_{n \in P_k} z_n^k(\underline{h},\underline{g}) . \tag{6.13}$$

Now let

$$h_j^k = 0 \text{ for all } j \in \overline{Q}_k^i \; , \quad k = 1,\ldots,K . \tag{6.14}$$

Equations (6.12) and (6.13) imply that the Extended Equilibrium Principle holds for all paths in P_k, $k = 1,\ldots,K$. Therefore, it may be concluded that the solution obtained by this algorithm satisfies the necessary conditions of the HOP.

In addition, it can also be shown that this algorithm terminates in a finite number of major iterations.

This is explained as follows. It is known that the total number of paths in a network is finite. Also, in the method of updating of the active path set (6.6), (6.7), no path is ever thrown out. Therefore, the number of elements in the active path set increases as i increases, and there exists some finite i* such that $Q_k^* \subset Q_k^{i*}$ for all k. Since the solution obtained minimizes the restricted Hybrid Optimization Problem, HOP^{i*}, it is at least a local minimum over the restricted path sets Q_k^{i*} , $k = 1,\ldots,K$.

However, there are still some practical problems that have to be resolved before this algorithm can be implemented. The step of verifying the optimality condition requires both the solution as well as the multipliers to be fairly accurate. However, our numerical experience shows that the solution of the HOP obtained by the Augmented Lagrangian Method is sufficiently accurate, but the multiplier obtained may not be as accurate. Special actions which require more computations have to be taken to obtain more accurate multipliers. Moreover, the restricted HOPs at each major iteration may grow too large as i approaches i*.

The dimensionality of the restricted HOP at i* is at least as big as

$U + \sum_{k=1}^{K} |Q_k^*|$. For a network of 10 O-D pairs, 10 control parameters, and assuming there are 2 active paths in Q_k^* for each k, this dimensionality is 30, which is still large. The restricted HOP at every iteration is a nontrivial problem to be solved; even if we can make use of the solution of the previous iteration as an initial guess solution. Note that this guess solution is not feasible; however, it satisfies most of the constraints.

6.2 Algorithm with Reduced Dimensionality

In this algorithm, we make use of the fact that the equilibrium flow pattern is a function of $\underline{g}$, the traffic control parameters, and under some assumptions on the link time function, it is uniquely determined by $\underline{g}$. This is so because, under the assumptions of separability in link travel-time function, the computation of equilibrium flow pattern has been shown [32] to be equivalent to solving the following optimization problem, the Equilibrium Assignment Program.

EAP (Equilibrium Assignment Program)

$$\underset{\{\underline{f}^k\}}{\text{Minimize}} \quad \sum_{a=1}^{NL} \int_0^{f_a} t_a(x,\underline{g})dx ,$$

subject to

$$\underline{A}\ \underline{f}^k = \underline{w}^k, \tag{6.15}$$

$$\underline{f}^k \geq 0 , \tag{6.16}$$

$$\underline{f} = \sum_{k=1}^{K} \underline{f}^k . \tag{6.17}$$

If the link time function is further assumed to be an increasing function of the link flow at a fixed value of $\underline{g}$; i.e.,

$$f''_a > f'_a \Rightarrow t_a(f''_a,\underline{g}) > t_a(\underline{f}'_a,\underline{g}) , \tag{6.18}$$

then the cost of EAP is strictly convex in $\underline{f}$. The feasible region defined by constraints (6.15) and (6.16) is a convex polytope. Therefore, EAP is a

strictly convex program. It has been proved that the solution to this type of problem exists and is unique (see, for example, [50]).

This observation allows us to consider only the traffic control variables as decision variables in the HOP. The flow variables are then considered as dependent variables which are uniquely determined by $\underline{g}$. The advantage of taking this approach is that the number of decision variables in the HOP is greatly reduced to the number of traffic control variables, and no enumeration of paths is required at all.

The HOP is now written as

HOP2 $$\min_{\underline{g} \in G} J[\underline{f}(\underline{g}),\underline{g}] ,$$

where <u>f(g) is an equilibrium flow</u> for controls $\underline{g}$.

Note that, when solving HOP2, for each functional evaluation of $J[\underline{f}(\underline{g}),\underline{g}]$, we must solve an Equilibrium Assignment Problem (EAP) to find $\underline{f}(\underline{g})$. Moreover, because the functional form of $f(\underline{g})$ is unknown, the derivatives of $f(\underline{g})$ with respect to $\underline{g}$ are not explicitly available in closed form. This implies that optimization algorithms that require gradient information such as the Augmented Lagrangian Method cannot be applied directly. Consequently, we must resort to one of the two alternatives available: approximate the gradient by finite differences, or make use of solution techniques which do not require derivatives.

However, optimization algorithms which require derivative information are still not applicable even if we attempt to compute the derivatives by finite difference approximations. This is due to the fact that these algorithms assume at least first-order differentiability. However, we have observed that $J[\underline{f}(\underline{g}),\underline{g}]$ may not be differentiable in $\underline{g}$ at some points as shown in example 2 in section 5.4.

There exist in the literature several minimization algorithms that do not require derivatives. Powell [60] provides an excellent survey and discussion. Abdulaal, <u>et al</u>. [41] have successfully applied two different algorithms of this class, Powell's Method [61] and Hooke and Jeeves' Method [62], to the Equilibrium Network Design Problem. It seems such an approach is worth further in-

vestigation for solving large-scale HOPs. However, it should be noted that because of the nondifferentiability of the cost, algorithms which approximate gradients should not be used.

This approach enables us to reduce greatly the dimensionality of the problem, and hence, large-scale problems can be treated. However, this advantage is not without cost. The disadvantage is that the algorithms are comparatively slow and functional evaluation of $\underline{f}(\underline{g})$, which involves solving an Equilibrium Assignment Program, is required; hence, it is time consuming.

6.3 Approximate Solution

In this section, we outline a method of computing the upper and lower bounds on the optimal cost of the HOP. Using this result and previous results in the literature, we conjecture that the differences between these bounds are small, and propose an approximate solution to the Hybrid Optimization Problem. Suppose we solve the following sequence of Generalized System Optimization Problems and Equilibrium Assignment Problems:

GSOP

$$\underset{\underline{f},\underline{g}}{\text{Minimize}} \quad J(\underline{f},\underline{g})$$

$$\underline{A}\,\underline{f}^k = \underline{w}^k \;, \; k = 1,\ldots,K \;, \tag{6.19}$$

$$\underline{f}^k \geq 0 \;, \quad k = 1,\ldots,K \;, \tag{6.20}$$

$$\underline{g} \in G \;, \tag{6.21}$$

where

$$\underline{f} = \sum_{k=1}^{K} \underline{f}^k \;. \tag{6.22}$$

Let the optimal solution be $(\underline{f}^*,\underline{g}^*)$ and

$$J^L = J(\underline{f}^*,\underline{g}^*) \;. \tag{6.23}$$

Note that $\underline{f}^*$ is also a system optimized flow for $\underline{g} = \underline{g}^*$. Now, solve the following Equilibrium Assignment Problem at $\underline{g} = \underline{g}^*$.

EAP ($\underline{g}$*)

$$\underset{\{\underline{f}^k\}}{\text{Minimize}} \sum_{i=1}^{NL} \int_o^{f_i} t_i(x,\underline{g}^*)dx,$$

subject to

$$\underline{A}\ \underline{f}^k = \underline{w}^k \ , \quad k = 1,\ldots,K \ , \tag{6.24}$$

$$\underline{f}^k \geq \underline{0} \ , \qquad k = 1,\ldots,K \ , \tag{6.25}$$

where

$$\underline{f} = \sum_{k=1}^{K} \underline{f}^k . \tag{6.26}$$

Let the optimal solution be $\underline{f}^{**}$ and

$$J^U \triangleq J(\underline{f}^{**},\underline{g}^*) . \tag{6.27}$$

We have the following lemma

Lemma 6.1

Let J^H be the optimal cost of the HOP. Then,

$$J^L \leq J^H \leq J^U \ . \tag{6.28}$$

Proof

The lower bound has already been proved in Section 2, where we have shown that the optimal cost of GSOP is the best cost achievable if the system has control over both link flows and traffic control parameters. Now, we show the upper bound.

Since $\underline{f}^{**}$ solves EAP($\underline{g}$*), then by definition, $\underline{f}^{**}$ is an equilibrium flow vector at $\underline{g} = \underline{g}^*$. It follows that

($\underline{f}^{**},\underline{g}^*$) is a feasible solution of HOP.

Hence,

$$J^H \leq J(\underline{f}^{**},\underline{g}^*) = J^U ,$$ Q.E.D.

Computationally, the determination of these bounds involves solving a GSOP and an EAP. The computation of GSOP is not as difficult as solving the HOP because it is possible to decompose the GSOP into two studied problems. If we fix $\underline{g}$, the GSOP becomes the conventional system optimization problem which is a min cost multicommodity network flow problem.

If flows are fixed in the GSOP, we obtain a control optimization problem. Since both of these problems have the same system cost, the GSOP can be solved by an iterative procedure consisting of successive alternations between the system optimization and control optimization problems. Note that the computation of the bounds now involves three problems, (a) the conventional system optimization problem, (b) the control optimization problem, and (c) the Equilibrium Assignment Problem, which have been under intensive research and are understood. Efficient algorithms and software are available for networks of relatively large scale. Therefore, the computation of the bounds is not as difficult as the problem of solving the HOP, especially in cases involving large-scale networks.

The bounds allow us to have an approximate solution to the HOP if their difference, $J^U - J^L$ is not too large. Practical experience by previous researchers [40], [63], seems to indicate that this is indeed the case. Note that $\underline{f}^*$ and $\underline{f}^{**}$ are, respectively, the system optimized and user optimized flow at $\underline{g} = \underline{g}^*$. Numerical experiments on flow patterns according to system and user optimization have been conducted by different researchers; e.g., Yagar [63] and LeBlanc _et al_. [40]. Both of these studies have arrived at the same conclusion: the difference between the system optimized and the user optimized flow patterns is not significant. Both the studies assume total travel time as the cost in the system otpimization problem. Since energy cost is a linear function of the travel time, we suspect that the same is true for the energy optimization case. In addition, our limited computation experience seems to indicate that flow distributions which are energy- and time-optimal, are not very different. In the work of LeBlanc _et al_. [40] a network of 76 arcs and 24 nodes with 528 O-D pairs is considered. It has been found that the average percentage difference in link flows between the system and user optimized flow patterns is 6 percent and that of the total travel time (system cost) is 2 percent. Based on these numerical results, we may conjecture

that, in most practical cases, $J^U - J^L$ is small; therefore ($\underline{f}^{**}$, $\underline{g}^*$) can be used as an approximate solution to the HOP. This approximation has also been suggested by Murchland [64].

6.4 Conclusion

In this section, we have pointed out two practical difficulties which must be treated by any algorithm to be applicable for the solution of Hybrid Optimization Problems in large networks. Since enumeration of all paths in a large network is nontrivial, it must be avoided. In addition, the dimensionality of the resulting optimization problem must not be too large to be solvable on computers within reasonable time limits. Two algorithms are proposed to satisfy these requirements. We also have outlined a procedure for computing the upper and lower bounds of the optimal cost of the Hybrid Optimization Problem. It is shown that this procedure involves solving understood problems with developed computer software, and is not as difficult as solving the HOP. It is conjectured that the difference between the bounds is small, and approximate solution of the HOP is proposed.

7. SUMMARY, CONCLUSIONS, AND SUGGESTIONS FOR FUTURE RESEARCH

7.1 Summary and Conclusions

In this report, we address a class of traffic control problems which takes into account the role of individual drivers as independent decisionmakers. The Hybrid Optimization Problem is formulated to provide a general theoretical framework for the analysis of this problem. The formulation of the Hybrid Optimization Problem (HOP) can be viewed as a unified approach which combines the system optimization, control optimization, and assignment problems in traffic engineering.

System optimization and signal control optimization contribute to the concept of a system-wide cost function, such as total travel time (traveler-hours per hour) or total energy consumption (gallons of gasoline per hour). The HOP has the selection of values for control parameters (such as green splits) in common with the traffic signal control optimization problem. The third element is the concept of assignment which predicts the distribution of vehicles in the network.

In Section 3, two behavioral models on the route selection process and the resulting overall network flow distribution principles are discussed. We devote a major part of this report to the study of a specific HOP which assumes user optimization as a flow distribution principle.

Conventionally, user-optimized flow is mathematically characterized by a set of logical statements of the form, "If a, then b," and in the case where the link travel-time function is separable, it can also be characterized as an optimal solution of a certain minimization problem. These mathematical characterizations are not in a form convenient for use as constraints to the HOP. In Section 3, we present a new mathematical characterization of the user-optimized flow which consists of equalities and inequalities. They can then be incorporated as constraints.

In Section 4,we derive the necessary conditions for the optimal solution of the HOP. We show that it is possible to have a physical interpretation of these conditions. This is summarized in the Extended Equilibrium Principle which generalizes the notion of the conventional Equilibrium Principle in user-optimization: not only are travel times equalized over utilized paths, but also some quantities

involving the marginal system cost and marginal link travel times. It is interesting to note that the Extended Equilibrium Principle bears a remarkable similarity to the Equal Travel-Time Principle of user-optimization and the Equal Marginal Cost Principle of system optimization in traffic engineering. This is intuitive since the HOP combines features of user optimization and system optimization.

A Generalized System Optimization Problem (GSOP) is defined to extend the notion of system optimization. It is shown that the GSOP is useful in two ways. In Section 2, we show that the optimal cost of the GSOP is a lower bound for any traffic control strategy, independent of the behavioral assumptions on route choice among the drivers. In Section 6, we conjecture that the optimal traffic control parameters of the GSOP can be used as an approximate solution to the HOP.

Three numerical examples are presented in Section 5 to demonstrate and verify the concepts and formulation of the HOP and the Extended Equilibrium Principle. Several important conclusions are made from these examples. The Iterative Optimization Assignment Algorithm is studied for two important reasons. First, it has been proposed [14], [15], [37], as a solution method for traffic control problems (similar to those the HOP is intended for). Second, it simulates the real-life process if optimal traffic controls are periodically updated, based on new information about the flow distribution in the network. This process has also been described [14] as a control strategy to take into account the redistributional effects of traffic after implementation of some optimal signal timing plans. The numerical results in Section 5 show that the Iterative Optimization-Assignment Algorithms converge to wrong solutions, and in a contrived example, the worst possible solution.

The approach of using the Hybrid Optimization formulation with the Augmented Lagrangian Method as a solution method is shown to lead to the correct control and assignment. The optimality of the results obtained in this way is verified by direct evaluation of the system cost at different values of the control parameters. The Extended Equilibrium Principle is also shown to be valid for the examples considered.

These results point to the following conclusions. The HOP formulation is a useful approach for treating the class of traffic control problems considered

here, while the Iterative Optimization-Assignment Algorithm may lead to wrong solutions, and therefore, is not recommended.

In addition, valuable insights are also gained from these simple examples. A close examination of the variation of the system cost as a function of the control parameters shows that the system cost may not be differentiable. This is due to the fact that the set of active paths i.e., paths carrying positive flows, is not the same in different regions in the space of traffic control parameters. The points at which the cost is nondifferentiable are the transition points where some active paths disappear and some new active paths are introduced. This is attributable to the assumption of user optimization as the flow distribution principle.

In Section 6, we present some practical considerations for solving large-scale Hybrid Optimization Problems. It is pointed out that the enumeration of all paths in large networks is a difficult task, hence it should be avoided. So that the computation be performed in reasonable time, the dimensionality of the optimization problem must not be too large.

Two algorithms that exploit some of the special structures of the HOP and the Extended Equilibrium Principle are proposed to satisfy these requirements. However, there are still some difficulties associated with these algorithms. In one algorithm, it is required to have the solution and multipliers of the restricted HOP sufficiently accurate. Moreover, the dimensionality of the restricted HOP is still quite large despite all the efforts taken to reduce the dimensionality. In the second method, evaluation of functions is extremely time-consuming.

Nevertheless, they can be viewed as possible directions for the development of algorithms for large-scale HOPs. In Section 6, we also present a procedure for computing the upper and lower bounds of the optimal cost of the HOP. This procedure is shown to be decomposable into system optimization, control optimization, and user optimization problems, which are relatively easy to solve compared with the HOP. Using previous results on traffic flow patterns [40], [63], we conjecture that the difference between these bounds is small, and that an approximate solution of the HOP can be obtained from the procedure for computing the bounds.

7.2 Future Research

The purposes of the research reported here have been to formulate the Hybrid Optimization Problem and to demonstrate its importance. Future research should be directed toward (1) understanding the consequences of this concept for traffic control and network planning, (2) further clarifying its technical aspects, (3) developing practical algorithms for larger networks than those described here, and (4) studying the game theoretic aspects of the problem. In the remainder of this section, these research goals are discussed at greater length.

7.2.1 Traffic Control and Network Planning

To extend the concept of Hybrid Optimization to network planning, the vector of control parameters must be augmented to include the capacities of links that are under consideration for construction or improvement. The criterion function J must be modified to include the construction cost. The solution to the resulting problem will yield the optimal link capacities as well as the signal control strategy and the distribution of flow in the resulting network.

It should be emphasized that this is not the existing network planning formulation [20], [41], since we include signal parameters. Since only a small number of links are likely to be considered for improvement at any one time, this is a small extension to the HOP discussed in this report.

Before the hybrid concept is useful for network planning, and possibly even for traffic control, there are two kinds of issues that should be understood.

Macroscopic phenomena include choices made by travelers or potential travelers throughout the network, such as whether or not to travel, when, to and from what points, and by what modes. In this report, we have assumed that these choices have been made. We restrict our attention to those that have chosen the automobile, and we focus on the last remaining choice: what route to take. We have demonstrated that when control policies favor one part of the network over another, there will be important consequences for that last choice, and thus, for the distribution of vehicles.

There may also be important consequences on the other choices as well. In particular, if an important commuting route is made substantially more attractive

by a change in control policy or by link expansion, it may attract additional traffic during rush hour, not just from other parallel routes, but from other modes, and even from other hours of the day.

These considerations would appear to be important, at least for large changes in the network. If so, their inclusion would require some additional degrees of freedom. However, they might also require that the whole urban region be treated simultaneously. This implies that networks of great complexity must be treated, or that a reliable method of network aggregation be used.

Microscopic phenomena require more detailed, more accurate models of delays on links or at intersections. For example, the best open-loop traffic signal control strategy involves values of cycle times, green splits, and offsets at each intersection. Existing techniques for calculating these quantities take network flows as fixed and specified. To extend the Hybrid Optimization concept to this kind of control strategy, models of link delays must be incorporated into the problem formulation. A difficulty that this will create is that signal offsets are constrained by equations involving integers. That is, the HOP becomes a mixed integer-continuous variable nonlinear programming problem.

Improved vehicle detection hardware and software and improved on-line microprocessor-based computational facilities are making closed-loop traffic control strategies increasingly attractive. Typically such devices can have considerable local autonomy. That is, the value of the signal (red or green facing a given link) at each intersection depends largely on local conditions (e.g., the current estimate of the number of vehicles in queues on each link facing the signal). A limited amount of information may be available from a central computer such as when to switch to an open-loop strategy.

Here again, if a part of the network is favored by the control policy, it will tend to attract drivers. Further, if the closed loop policy is such that more heavily used links are favored, then such a policy is likely to result in the undesirable behavior observed with the heuristic iterative assignment optimization in Section 5. Therefore, this is a potential application of the hybrid concept.

It is also important to investigate methods of extending the HOP formulation to include multiple vehicle classes such as cars, carpools, buses. One important

point is that buses are scheduled to travel along fixed routes, and thus, must be treated differently.

The Electronic Route Guidance System [65] is a proposed traffic control system in which every vehicle follows the routing instructions given by the traffic authority in traveling in a network. A traffic system completely controlled by the Electronic Route Guidance System may be a bit far-fetched. A traffic system where only some of the vehicles are under the control of the Electronic Route Guidance System seem more probable. This kind of situation seems to have some important ingredients of the HOP. Research should be carried out to investigate how the HOP formulation can be extended to this situation.

7.2.2 Technical Aspects

Sensitivity Analysis

The numerical examples in Section 5 reveal the fact that the system cost of the HOP may not be differentiable at some points in the space of traffic control parameters. These numerical results and intuition seem to indicate that the cost is continuous. More theoretical analysis on the behavior of equilibrium flow and system cost is required. Sensitivity analysis of equilibrium flows with respect to input flows is reported in Hall [36]. The sensitivity analysis of equilibrium flow with respect to the traffic control parameters is still an unexplored area. It is important for theoretical network flow analysis, evaluation of alternative control strategies, and it can possibly be used in algorithms for large-scale Hybrid Optimization Problems.

Hybrid Optimization Problem with Alternative Flow Distribution Prinicples

The nondifferentiability of system cost has been attributed to the assumption of user optimization as the flow optimization model. Therefore, it is important to study Hybrid Optimization with alternative flow distribution principles such as probabilistic assignment models. It is of interest to investigate the sensitivities of cost, controls, and flows with respect to different behavioral assumptions.

A study on the HOPs with the probabilistic assignment model as the flow distribution can provide some answer to the following questions. How sensitive are the optimal cost and controls of the HOP with respect to the quality of the drivers' knowledge of the traffic conditions? How are the system cost, flow

distribution, and controls influenced by the sensitivity of the drivers with respect to the time differences among different routes?

Theoretical Study on the Bounds of the Optimal Cost of HOP

It has been conjectured that the difference between the upper and lower bounds of the optimal cost of HOP is small enough to justify using the optimal values of the control parameters in the GSOP as an approximate solution to the HOP. More theoretical study is needed to investigate the conditions under which this conjecture is true. This knowledge is important because, under these conditions, the more difficult task of performing the HOP can be avoided.

7.2.3 More Detailed Developments on Proposed Algorithms

Two algorithms are outlined in Section 6 to alleviate some of the problems in solving large-scale HOPs. There are still some practical difficulties associated with them. For example, research on the first algorithm should be directed at the methods of updating the active path set to prevent the restricted HOP from growing too large. In the second algorithm, functional evaluations (complete network assignment calculations) are time-consuming, and some method of approximation must be devised to avoid functional evaluations at every intermediate step.

7.2.4 Study of Game Implications of HOP

It has been pointed out [66] that the equilibrium flow is a Nash equilibrium point in the sense that no single driver can improve his trip cost by unilaterally changing his route. Miller et al. [67] suggests the name "supervised noncooperative game" for a traffic control problem which is similar to the HOP. The drivers are considered to be engaged in a noncooperative game, each trying to reach his/her destination as fast as possible by making a choice among different, available routes. The "supervisor" in this case is the traffic authority who uses the traffic control parameters to manipulate the noncooperative game.

The HOP can also be viewed as a Stackelberg game [68], [69] in the following sense. The traffic authority is the leader who tries to minimize a system cost, using available traffic control parameters. The follower in this game

is an aggregated player who represents the drivers. The minimization problem faced by the follower is the Equilibrium Assignment problem, where the cost function is the sum of the integrals of link time (Section 6.3), and the decision variables are the flow variables. The traffic authority (the leader) minimizes his cost with the knowledge of the reaction of the drivers (the follower) to his action.

In addition, if the Iterative Optimization Assignment Algorithm converges, it does so to a Nash equilibrium point. In this case, the two players in this noncooperative game are the leader and follower of the Stackelberg game just described.

The solution, if it exists, has the property that neither of the plays can improve his cost by unilaterally deviating from his current strategy. An investigation using game theory methodologies in greater depth is appropriate to study the interactions between the various players.

In conclusion, we view the research results presented in this report as being very important contributions to improved operation of urban traffic networks. We have demonstrated that an intuitively appealing method, the Iterative Optimiation Assignment Algorithm, can lead to erroneous results. On a more constructive note we have developed in detail a new Hybrid Optimization method that overcomes the shortcoming of not fully taking into account driver behavior. The research reported in this report is a necessary first step; long-range theoretical, algorithmic, and simulation research is needed before this method can have a <u>practical</u> impact upon important problems in traffic engineering, both with respect to planning and traffic control.

REFERENCES

1. D. Tabak, "Application of Modern Control and Optimization Techniques to Transportation Systems," in Control and Dynamic Systems, edited by C.T. Leondes, Vol. 10, 1973, Academic Press, New York, p. 346-435.

2. P.K. Houpt, M. Athans, "Dynamic Stochastic Control of Freeway Corridor Systems. Volume I: Summary," U.S. Department of Transportation Report DOT-TSC-RSPA-78-12, August 1975.

3. D.P. Looze, P.K. Houpt, N.R. Sandell, Jr., M. Athans, "On Decentralized Estimation and Control with Application to Freeway Ramp Metering," IEEE Trans. Automatic Control, Vol. AC-23, April 1978, p. 268-275.

4. P. Dersin, S.B. Gershwin, M. Athans, "Perturbation Strategies for Optimal Traffic Reassignment," IEEE Trans. on Automatic Control, Vol. AC-24, February 1979, p. 3-12.

5. S. Guberinic, G. Senborn, "Optimal Sequences of Fixed-Time Signal Plans," Transportation Science, Volume 12, No. 1, February 1978, p. 48-57.

6. D.I. Robertson, R.D. Bretherton, "Optimum Control of an intersection for any known sequence of vehicle arrivals," in Traffic Control and Transportation Systems, Proc. of Second IFAC/IFIP/IFORS Symposium, North Holland Publishing, Amsterdam, 1974, p. 3-19.

7. A.D. May, "Some fixed time signal control programs," in Traffic Control and Transportation Systems, Proc. of Second IFAC/IFIP/IFORS Symposium, North Holland Publishing, Amsterdam, 1974, p. 209-236.

8. J. Garcia Ramon, A. Lopez Montejano, D. Sanchez Hernandez, "New area traffic control system in the City of Barcelona," in Traffic Control and Transportation Systems, Proc. of Second IFAC/IFIP/IFORS Symposium, North Holland Publishing, Amsterdam, 1974, p. 83-94.

9. K.W. Huddart, D. Overton, M. Chandler, E. Turner, "London traffic signals - Their practical management and the use of computers and detectors," in Traffic Control and Transportation Systems, Proc. of Second IFAC/IFIP/IFORS Symposium, North Holland Publishing, Amsterdam, 1974, p. 69-82.

10. D.I. Robertson, "TRANSYT: A Traffic Network Study Tool," Road Research Laboratory Report LR 253, Crowthorne, England, 1969.

11. N.H. Gartner, J.D.C. Little, H. Gabbay, "Optimization of Traffic Signal Settings by Mixed-Integer Linear Programming -- Part I: The Network Coordination Problem," and "Part II: The Network Synchronization Problem," Transportation Science, Vol. 9, p. 321-363, 1975.

12. Traffic Research Corporation, "SIGOP: Traffic Signal Optimization Program - A Computer Program to Calculate Optimum Coordination in a Grid Network of Synchronized Traffic Signals." Distributed by U.S. Clearinghouse, Report PB-173-738, September 1966.

13. M.J. Beckmann, "Equilibrium versus Optimum in Public Transportation Systems," in Traffic Equilibrium Methods, edited by M.A. Florian, Springer-Verlag, Berlin, 1976, p. 119-131.

14. R. Akcelik, M.J. Maher, "The redistributional effects of area traffic control policy," Traffic Engineering and Control, Vol. 16, No. 9, p 383-385, September 1975.

15. S.B. Gershwin, P. Ross, J.D.C. Little, N. Gartner, "Optimization of Large Traffic Systems," presented at 57th Annual Meeting, Transportation Research Board, January 16-20, 1978, Washington DC.

16. J.G. Wardrop, "Some Theoretical Aspects of Road Traffic Research," Proc. Inst. Civil Engineers, Part 2, 1952, p. 325-378.

17. Cosmis Corporation, Traffic Assignment, Wheaton, Maryland, prepared for the Urban Planning Div., Office of Highway Planning, U.S. Department of Transportation, Federal Highway Administration, August 1973.

18. S.C. Dafermos, F.T. Sparrow, "The Traffic Assignment Problem for a General Network," J. Research of National Bureau of Standards - B Mathematical Sciences, Vol. 73B, No. 2, April-June 1969, p. 91-118.

19. M. Beckmann, et al., Studies in the Economics of Transportation, Yale University Press, New Haven, CT, 1956.

20. P.A. Steenbrink, Optimization of Transportation Network, Wiley, New York, 1974.

21. A.A. Assad, "Multicommodity Network Flows -- A Survey," Networks, Vol. 8, No. 1, Spring 1978, p. 37-91.

22. D.G. Cantor, M. Gerla, "Optimal Routing in a Packet-Switched Computer Network," IEEE Trans. on Computers, Vol. C-23, 10, p. 1062-1068, October 1974.

23. J.E. Defenderfer, "Comparative Analysis of Routing Algorithms for Computer Network," Ph.D. Thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1977.

24. M. Schwartz, C. Cheung, "The Gradient Projection Method for Multiple Routing in Message-Switched Networks," IEEE Trans. Communication, Vol. COM-24, No. 4, April 1976, p. 449-455.

25. S. Yagar, "Emulation of Dynamic Equilibrium in Traffic Networks," in Traffic Equilibrium Methods, edited by M.A. Florian, Springer-Verlag, Berlin, 1976, p. 240-264.

26. M.H.Ueberschaer, "Choice of Routes on Urban Network for the Journey to Work," Highway Research Record No. 369, p. 228-238.

27. B.V. Martin, "Minimum Path Algorithms for Transportation Planning," Report 63-52, Department of Civil Engineering, M.I.T., December 1963.

28. R.R. Snell, et al., "Traffic Assignment with a Nonlinear Travel Time Function," Transportation Science, Vol. 2, p. 146-159, 1968.

29. E.R. Ruiter, "Implementation of Operational Network Equilbirium Procedures," Transportation Research Record No. 491, p. 40-51.

30. J.B. Dennis, Mathematical Programming and Electrical Networks, M.I.T. Press, Cambridge MA, 1959.

31. H.J. Payne, S.D. Miller, W.A. Thompson, "An Algorithm for Traffic Assignment on Capacity Constraint Transportation Network with Queue," presented at Johns Hopkins Conference on Information Sciences and Systems, April 2-4, 1975.

32. S. Nguyen, "A Unified Approach to Equilibrium Methods for Traffic Assignment," presented at International Symposium on Traffic Equilibrium Methods, Montreal, PQ, Canada, November 1974.

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

34. L.J. LeBlanc, et al., "An Accurate and Efficient Approach to Equilibrium Traffic Assignment on Congested Networks," Transportation Research Record 491, p. 12-33, 1974.

35. H. Aashtiani, "Multi-modal Traffic Assignment," Ph.D. Thesis, Operations Research Center, M.I.T. (in preparation).

36. M.A. Hall, "Properties of Equilibrium State in Transportation Networks," Transportation Science, Vol. 12, No. 3, August 1978, p. 208-216.

37. R.E. Allsop, "Some Possibilities for Using Traffic Control to Influence Trip Distribution and Route Choice," Transportation and Traffic Theory, Proc. of Sixth Internatl. Symposium on Transportation and Traffic Theory, edited by D.J. Buckley, p. 345-375, Elsevier, New York, 1974.

38. R.E. Allsop, J.A. Charlesworth, "Traffic in a Signal-Controlled Road Network: An Example of Different Signal Timings Inducing Different Routings," Traffic Engineering and Control, p. 262-264, Vol. 18, No. 5, May 1977.

39. N. Gartner, "Area Traffic Control and Network Equilibrium," International Symposium on Traffic Equilibrium Methods, University of Montreal, PQ Canada, November 1974.

40. L.J. LeBlanc, E.K. Morlok, "An Analysis and Comparison of Behavioral Assumptions in Traffic Assignment," in Traffic Equilibrium Methods, edited by M.A. Florian, Springer-Verlag, Berlin, 1976, p. 413-425.

41. M. Abdulaal, L.J. LeBlanc, "Continuous Equilibrium Network Design Models," Tech. Report OREM 77022, Department of Operations Research and Engineering Management, Southern Methodist University, Dallas TX, April 1978.

42. L. Evans, R. Herman, "A Simplified Approach To Calculations of Fuel Consumption in Urban Traffic Systems," Traffic Engineering and Control, August/September 1976, p. 352-354.

43. L. Evans, R. Herman, "Automobile Fueld Economy on Fixed Urban Driving Schedules," Transportation Science, Vol. 12, No. 2, May 1978, p. 137-152.

44. Highway Capacity Manual, Highway Research Board Special Report 87, National Academy of Sciences, National Research Council Publication 1328, 1965.

45. T.L. Magnanti, B.L. Golden, "Transportation Planning: Network Models and Their Implementation," Operations Research Center, M.I.T., December 1977.

46. K.M. Anantharamaiah, "Equilibrium Conditions in Traffic Assignment," Transportation and Traffic Theory, edited by D.J. Buckley, Elsevier NY, 1974, p. 483-494.

47. H. Beilner, F. Jacobs, "Probabilistic Aspects of Traffic Assignment," Proc. of Fifth Internatl. Symposium on the Theory of Traffic Flow and Transportation, edited by G.F. Newell, American Elsevier, New York, 1972, p. 183-194.

48. R.B. Dial, "A Probabilistic Multipath Traffic Assignment Model which Obviates Path Enumeration," Transportation Research, Vol. 5, 1971, p. 83-111.

49. M. Florian, S. Nguyen, "An Application and Validation of Equilibrium Trip Assignment Methods," Transportation Science, Vol. 10, No. 4, p. 374-390, November 1976.

50. M. Avriel, Nonlinear Programming, Analysis and Methods, Prentice-Hall, Englewood Cliffs NJ, 1976.

51. M.J.D. Powell, "Algorithms for Nonlinear Constraints that use Lagrangian Functions," Mathematical Programming, Vol. 14, No. 2, p. 224-248, March 1978.

52. D.P. Bertsekas, "Multiplier methods: a survey," Automatica, Vol. 12, 1976, p. 133-145.

53. R. Fletcher, "Methods related to Lagrangian functions," in Numerical methods for constrained optimization, edited by P.E. Gill and W. Murray, Academic Press, London, 1974, p. 219-239.

54. R.T. Rockafellar, "A Dual Approach to Solving Nonlinear Programming Problems by Unconstrained Optimization," Mathematical Programming, Vol. 5, 1973, p. 354-373.

55. R.W.H. Sargent, "Reduced-gradient and Projection Methods for Nonlinear Programming," in Numerical Methods for Constrained Optimization, edited by P.E. Gill and W. Murray, Academic Press, London, 1974, p. 149-174.

56. A.V. Fiacco, G.P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, New York, 1968.

57. D.P. Bertsekas, "Convergence Rate of Penalty and Multiplier Methods," Proc. of 1973 IEEE Conference on Decision and Control, p. 260-264.

58. S.B. Gershwin, J.D.C. Little, N. Gartner, "Computer-assisted Traffic Engineering using Assignment, Optimal Signal Setting and Modal Split," Department of Transportation, Final Report No. DOT-TSC-RSPA-78-10, Laboratory for Information and Decision Systems, M.I.T., May 1978.

59. S. Nguyen, L. James, "TRAFFIC: An Equilibrium Traffic Assignment Program," Publication #17, Center of Research on Transportation, University of Montreal, PQ, Canada, March 1975.

60. M.J.D. Powell, "A View of Unconstrained Minimization Algorithms That Do Not Require Derivatives," ACM Trans. on Math. Software, Vol. 1, No. 2, June 1975, p. 97-108.

61. M.J.D. Powell, "An Efficient Method for Finding the Minimum of a Function of Several Variables Without Using Derivatives," Brit. Computer Journal, Vol. 9, p. 155-162, 1964.

62. R. Hooke, T.A. Jeeves, "Direct Search Solution of Numerical and Statistical Problems," J. Assoc. Comp. Mach., Vol. 8, p. 212-229, 1961.

63. S. Yagar, "A Study of the Travel Patterns in a Corridor with Respect to the Assignment Principles of Wardrop," in Traffic Flow and Transportation, G.F. Newell, Editor, American Elsevier, 1972, p. 165-182.

64. J.D. Murchland, private communication with N. Gartner, 1977.

65. Electronic Route Guidance System, Highway Research Record No. 265, Highway Research Board, 1968.

66. A. Charnes, W.W. Cooper, "Multicopy Traffic Network Models," in Theory of Traffic Flow, Proc. of Symposium on Theory of Traffic Flow, edited by R. Herman, American Elsevier, New York, 1961, p. 85-96.

67. S.D. Miller, H.J. Payne, W.A. Thompson, "An Algorithm for Traffic Assignment on Capacity Constrained Transportation Networks with Queues," presented at Johns Hopkins Conference on Information Sciences and Systems, April 2-4, 1975.

68. J.B. Cruz, Jr., "Leader-Follower Strategies for Multilevel Systems," Decision and Control Laboratory of the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign IL.

69. H. von Stackelberg, The Theory of the Market Economy, Oxford University Press, Oxford, England, 1952.

APPENDIX A: LIST OF SYMBOLS

Symbol		Section
$\underline{A}$	the node-arc incidence matrix of a network, of dimension NN by NL	[2.1]*
a_{ij}	the (i,j) element of $\underline{A}$	[2.1]
$\alpha(i)$	the initial node of link i	[2.1]
$\underline{B}^k$	the link-path incidence matrix, of dimension NL by $\|P_k\|$, for the k^{th} O-D pair	[2.1]
b_{ij}^k	the (i,j) element of $\underline{B}^k$	[2.1]
$\beta(i)$	the terminal node of link i	[2.1]
$\underline{c}$	the link capacity vector, of dimension NL	[2.1.1]
c_i	the i^{th} element of $\underline{c}$; represents the capacity of link i	[2.1.3]
CYC	the cycle time on a traffic signal	[2.1.3]
D^k	the destination node of the k^{th} O-D pair	[2.1]
d_j	distance of link j	[2.1.3]
$\underline{e}$	the link fuel consumption vector, of dimension NL	[2.1]
e_i	the i^{th} element of $\underline{e}$; represents the total amount of fuel each vehicle consumes in traveling through link i	[2.1]
EAP	Equilibrium Assignment Program	[6.2]
η_i	the Lagrange multiplier (scalar) associated with inequality constraints (4.5)	[4.1.2]
$\underline{f}$	the link flow vector of dimension NL	[2.1]
f_i	the i^{th} element of $\underline{f}$; represents the amount of traffic flow on link i	[2.1]
$\underline{f}^k$	the link flow vector of the k^{th} O-D pair	[2.1]

* number in brackets is the section number where the symbol is first defined.

Symbol		Section
f_i^k	the i^{th} element of $\underline{f}^k$; represents the amount of traffic flow of the k^{th} O-D pair on link i	[2.1]
$\underline{f}^*$	optimal link flow vector of GSOP	[6.3]
$\underline{f}^{**}$	solution of EAP with $\underline{g} = \underline{g}^*$	[6.3]
G	the set of feasible controls	[2.1.2]
$\underline{g}$	the vector of traffic control parameters, of dimension U	[2.1.2]
$\mathcal{G}$	a directed graph	[2.1]
$\underline{\gamma}^k$	the Lagrange multipliers (of dimension $\|P_k\|$) of the constraints (4.2)	[4.1.2]
GSOP	Generalized System Optimization Problem	[2.2]
$\underline{g}^*$	optimal control parameters of GSOP	[6.3]
$\underline{h}$	the augmented path flow vector of dimension $\sum_{k=1}^{K} \|P_k\|$, formed from all $\underline{h}^k$, $k = 1,\ldots,K$	[2.1]
H^k	the total traffic requirement of the k^{th} O-D pair	[2.1]
$\underline{h}^k$	the path flow vector, of dimension $\|P_k\|$, of the k^{th} O-D pair	[2.1]
h_i^k	the i^{th} element of $\underline{h}^k$ and representing the traffic flow of the k^{th} O-D pair on the i^{th} path between the k^{th} O-D pair	[2.1]
HOP	Hybrid Optimization Problem	[2.4]
J	the system-wide cost to be minimized	[2.1.3]
J_e	total fuel consumption	[5.5.3]
J^H	the optimal system cost of the HOP	[6.3]
J^L	a lower bound on system cost J, defined in (6.22)	[6.3]
J_t	total travel time	[5.5.3]
J^U	an upper bound on system cost J, defined in (6.26)	[6.3]
K	the total number of origin-destination (O-D) pairs	[2.1]
L	the Lagrange function of the HOP	[4.1.2]

Symbol		Section
$\mathcal{L}$	a collection of links in a network	[2.1]
Λ	non-negative maximum function defined by (5.4)	[5.1]
λ_k	the Lagrange multiplier (a scalar) of the constraint (4.4)	[4.1.2]
ℓ_k	parameter used in the probabilistic flow distribution models in section (3.3)	[2.2]
ℓ_s	parameter used in the probabilistic flow distribution models in section (3.3)	[3.3]
ℓ_t	parameters used in the probabilistic flow distribution models in section (3.3)	[3.3]
M	total number of inequality constraints in NLP	[5.1]
MFP	multicommodity minimum cost network flow problem	[1.3.1]
$\underline{\mu}^k$	the Lagrange multipliers (of dimension $\|P_k\|$) of the constraints (4.3)	[4.1.2]
N	total number of equality constraints in NLP	[5.1]
n	total number of decision variables in NLP	[5.1]
$\mathcal{N}$	a collection of nodes in a network	[2.1]
NEDP	Network Equilibrium Design Problem	[2.4]
NL	the total number of links in a network	[2.1]
NLP	a general nonlinear constrained optimization problem	[5.1]
NN	the total number of nodes in a network	[2.1]
o^k	the origin node of the k^{th} O-D pair	[2.1]
Ω	the augmented Lagrangian function defined in (5.5)	[5.1]
w_i	inequality constraint of a general nonlinear minimization problem (NLP)	[5.1]
P_k	the set of all loop-free paths connecting the k^{th} O-D pair	[2.1]
$\|P_k\|$	the total number of available loop-free paths connecting the k^{th} O-D pair	[2.1]
Φ	a penalty function defined in (5.3)	[5.1]

Symbol		Section
ϕ_i	equality constraint of a general nonlinear minimization problem (NLP).	[5.1]
$\underline{\pi}$	a vector of dimension NL defined in (4.25)	[4.2]
$\underline{\psi}^k$	a vector of dimension N (defined in (4.28)) denoting the link pseudo-cost vector of the k^{th} O-D pair	[4.2]
ψ_i^k	the i^{th} element of $\underline{\psi}^k$, denoting the pseudo cost of the k^{th} O-D pair on link i	[4.2]
Q_k^i	set of active paths used in the i^{th} iteration of the proposed solution procedure for HOP in section (6.1)	[6.1]
Q_k^*	set of active paths for the k^{th} O-D pair in an optimal solution of the HOP, defined in (6.1)	[6.1]
$\overline{Q}_k^*$	complement of Q_k^* relative to P_k	[6.1]
R_i^k	the set of links on the i^{th} path of the k^{th} O-D pair	[3.3]
r_k^i	the shortest travel time path for the k^{th} O-D pair in the i^{th} iteration of the proposed solution algorithm for HOP in section (6.1)	[6.1]
ρ	penalty weight used in Ω	[5.1]
ρ^k	penalty weight used in the k^{th} iteration of the penalty function method as a solution procedure for NLP	[5.1]
s_k^i	the cheapest path (in terms of the link cost defined as the pseudo link cost, ψ_j^k) for the k^{th} O-D pair, in the i^{th} iteration of the proposed solution procedure for the HOP in section (6.1)	[6.1]
σ_i	the Lagrange multipliers (scalar) of the constraints (4.5)	[4.1.2]
$\underline{t}$	the link time vector, of dimension NL	[2.1]
t_i	the total amount of time each vehicle spends in traveling through link i	[2.1]
t_i^0	the average free flow speed on link i	[2.1.3]
t_i^s	the average transit time on link i	[2.1.3]
t_i^w	the average waiting time on link i due to queuing at the traffic signal on link i	[2.1.3]

Symbol		Section
$\underline{\tau}^k$	the path time vector, of dimension $\lvert P_k \rvert$ of the k^{th} O-D pair	[2.1]
τ_i^k	the i^{th} element of τ^k and representing the path time on the i^{th} path of the k^{th} O-D pair	[2.1]
τ_*^k	the travel time along the fastest path connecting the k^{th} O-D pair	[3.2.1]
θ^k	a scalar defined in (4.24)	[4.2.4]
U	the total number of traffic control parameters	[2.1.2]
$\underline{u}^k$	a vector of dimension $\lvert P_k \rvert$ with all elements having a value of 1	[4.1.1]
UP_k	the transformed unconstrained problem in the k^{th} iteration of the augmented Lagrangian method as a solution algorithm for NLP	[5.1]
$\underline{v}^k$	the minimum node time vector (of dimension NN) for the k^{th} O-D pair	[3.2.2]
v_i^k	the i^{th} element of $\underline{v}^k$; represents the minimum time taken to go from O^k, the origin node of the k^{th} O-D pair, to node i	[3.2.2]
$\underline{w}^k$	the traffic requirement vector, of dimension NN, for the k^{th} O-D pair	[2.1.1]
$\underline{x}$	an n-dimensional vector of decision variables of a general nonlinear minimization problem (NLP)	[5.1]
$\underline{x}^{k*}$	an n-dimensional vector which represents the optimal solution of UP_k, the transformed unconstrained problem in the k^{th} iteration of the augmented Lagrangian method	[5.1]
$\underline{\xi}$	an N-dimensional vector of Lagrange multipliers associated with the equality constraints, $\{\phi_i\}$	[5.1]
ξ_j	the Lagrange multiplier associated with the equality constraint ϕ_j of NLP	[5.1]
$\underline{\xi}^k$	the N-dimensional vector of Lagrange multipliers associated with the equality constraints, $\{\phi_i\}$, used in the k^{th} iteration of the augmented Lagrangian method	[5.1]
z_i^k	the path pseudo cost along the i^{th} path of the k O-D pair	[4.2]

Symbol		Section
z_*^k	the minimum path pseudo cost for the k^{th} O-D pair	[4.2]
$\underline{\zeta}$	an M-dimensional vector of Lagrange multipliers associated with the inequality constraints, $\{w_i\}$	[5.1]
ζ_i	the Lagrange multiplier associated with the inequality constraint w_i of NLP	[5.1]
$\underline{\zeta}^k$	the M-dimensional vector of Lagrange multipliers associated with the inequality constraints, $\{w_i\}$, used in the k^{th} iteration of the augmented Lagrangian method as a solution procedure for NLP.	[5.1]

APPENDIX B: REPORT OF NEW TECHNOLOGY

There are no inventions or other patentable items in this work. However, certain advances in the technology of traffic-signal-setting are reported here. In particular, a new formulation and analysis of the optimal-signal-setting problems are presented in Sections 2 and 4. Also two new algorithms are developed for the numerical solution of this problem. These algorithms as well as a preliminary discussion of their practicality are detailed in Section 6.

110 copies

U.S. DEPARTMENT OF TRANSPORTATION
RESEARCH AND SPECIAL PROGRAMS ADMINISTRATION
TRANSPORTATION SYSTEMS CENTER
KENDALL SQUARE, CAMBRIDGE, MA. 02142

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF TRANSPORTATION
613

U.S. MAIL