# Lecture Notes in Economics and Mathematical Systems 417

Bin Ran David E. Boyce

# Dynamic Urban Transportation Network Models

## Theory and Implications for Intelligent Vehicle-Highway Systems

*Springer-Verlag*
*Berlin*
*Heidelberg*
*GmbH*

Authors

Dr. Bin Ran
University of California at Berkeley
Institute of Transportation Studies
Building 452, Richmond Field Station
1301 South 46th Street
Richmond, CA 94804, USA

Dr. David E. Boyce
University of Illinois at Chicago
Urban Transportation Center
1033 W. Van Buren Street
Suite 700 South
Chicago, IL 60607, USA

ISBN 978-3-540-58360-8 ISBN 978-3-662-00773-0 (eBook)
DOI 10.1007/978-3-662-00773-0

Library of Congress Cataloging-in-Publication Data. Ran, Bin, 1967- . Dynamic urban transportation network models: theory and implications for intelligent vehicle-highway systems / Bin Ran, David E. Boyce. p. cm. – (Lecture notes in economics and mathematical systems; 417) Includes bibliographical references.

Network analysis (Planning) 3. Intelligent Vehicle Highway Systems. I. Boyce, David E. II. Title. III. Series. HE305.R36 1994 388.4'068–dc20 94-22844



Typesetting: Camera ready by author
SPIN: 10474196 42/3140-543210 - Printed on acid-free paper

To Jing and Nani

We offer this book as a tribute to

Martin Beckmann

whose pioneering contributions to Transportation Science have inspired us and the generation of academic researchers who followed in his footsteps.

# Preface

This book seeks to provide a systematic introduction to dynamic transportation network models. Intelligent vehicle highway systems are a major new motivation for dynamic transportation network modeling; in response, this book offers important insights into the complexity and challenge of these problems and their implications for IVHS. The book is not intended, however, to review classical transportation network models and algorithms. Instead, it offers a new framework for dynamic transportation network modeling. Thus, it should serve as a benchmark for assessing future research results. Nevertheless, the models in this text are not yet fully evaluated and are subject to revision based on future research.

A summary of the necessary mathematical background, including static optimization, optimal control and variational inequalities, provides a reference for transportation engineers involved in ATMIS projects. By understanding the mathematical requirements of dynamic transportation network problems, the professional community can appreciate the intensified research effort required for elaborating this new dimension of transportation science.

The first author's PhD dissertation provided a starting point for this book. Many researchers have contributed their ideas and comments. In particular, Professor Larry LeBlanc, Vanderbilt University, collaborated with the authors in writing several of the papers on which this book is based. We wish to acknowledge fully his important contributions to this effort.

Professor Ashish Sen, University of Illinois at Chicago, offered much encouragement to complete this book and provided valuable suggestions for using TEX. Professor Nagui Rouphail, North Carolina State University, also provided valuable encouragement and advice. Professor Mark Daskin, Northwestern University, and Professor Warren Powell, Princeton University, contributed numerous valuable comments on many chapters. Professor Floyd Hanson, University of Illinois at Chicago, taught the first author numerous skills from supercomputing techniques to advanced optimal control theory. Professor Hani Mahmassani, University of Texas at Austin, and Professor Anna Nagurney, University of Massachusetts at Amherst, provided continuing encouragement for our research during the past four years. Ms. Yu-Fang Zhang, Ms. Piyushimita Thakuriah and Mr. Mark Mathes, University of Illinois at Chicago, generously helped with the preparation of the manuscript.

The research described here was begun when the first author was a graduate student at the University of Tokyo; Dr. Toshikazu Shimazaki, Dr. Masao Kuwahara and Dr. Takashi Akamatsu helped him to begin his research on one of the most challenging topics in transportation engineering. Their continuing support and constructive comments are appreciated. Appreciation is also extended to Professor Carlos Daganzo and Drs. Steven Shladover, Stein Weissenberger and Hong K. Lo, University of California at Berkeley, for their valuable advice and encouragement.

Appreciation is extended to Professors Bruce Janson, University of Colorado at Denver, Randolph Hall, University of Southern California, Moshe Ben-Akiva and David Bernstein, Massachusetts Institute of Technology, Gang-Len Chang, University of Maryland at College Park, Michael Florian, University of Montreal, Andre de Palma, University of Geneva, Malachy Carey, University of Ulster, Terry Friesz, George Mason University, B.-W. Wie, University of Hawaii, Tony Smith, University of Pennsylvania, Mr. Alberto J. Santiago, Federal Highway Administration, and many anonymous referees for their constructive and helpful comments.

Special thanks go to other friends and colleagues at the Urban Transportation Center, University of Illinois at Chicago, and the Institute of Transportation Studies, University of California at Berkeley.

A computer time grant for the CRAY Y-MP from the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, (NCSA) is gratefully acknowledged. We are indebted to the UIC Workshop Program on Scientific Supercomputing and to Dr. Steven Roy for training in the use of the CRAY Y-MP system and other supercomputers. Partial support of this research was provided by the Federal Highway Administration and the Illinois Department of Transportation through the ADVANCE Project and the California Department of Transportation through the PATH Program. Partial support of this research was also provided by the Federal Highway Administration through Project DTFH61-90-R-00074 to the Center for Transportation Research at the University of Texas at Austin.

Bin Ran
David E. Boyce

Richmond, California
March, 1994

# Contents

# VII Implications for IVHS

# Chapter 1

# Introduction

Intelligent Vehicle Highway Systems (IVHS) seek to apply advanced computer, telecommunication, and information technologies to vehicles, transportation networks and operational plans, in order to relieve traffic congestion, reduce travelers' journey times, improve safety, reduce atmospheric emissions and energy consumption, and increase the productivity of transportation investment. Using IVHS technologies, vehicles and the infrastructure will exchange vast amounts of data back and forth, making possible the warning and avoidance of congestion or hazardous conditions, the automatic collection of tolls, the efficient dispatching of trucks and buses, dramatic improvements in safety and other benefits.

Within the framework of IVHS, Advanced Traveler Information Systems (ATIS) will provide historical, real-time and predictive information to support travel decisions; this will in turn influence the travel choices of individuals and consequently improve the time and quality of travel. Successive generations of advanced route guidance systems will improve utilization of the overall capacity of highway and transit systems so as to reduce travel times, congestion and accidents. By providing early detection of incidents and congestion in the transportation network, route guidance systems will redistribute traffic among the available modes and routes when there is excess capacity in some parts of the road network or shift the departure times of travelers to avoid peak-hour congestion when no additional road capacity is available. Furthermore, route guidance systems will provide travelers with accurate, current information on both transit and road networks so that some motorists can make their own time-cost tradeoffs and shift to transit, if appropriate.

## 1.1 Requirements for Dynamic Network Modeling

Traditional static network equilibrium models were developed for long range transportation planning. They are not suitable for analyzing and evaluating dynamic route guidance systems which need capability to solve transportation problems in real time (Boyce, 1989). Friesz et al (1989) have analyzed some of the fundamental properties of dynamic models which are pertinent to such route guidance systems. At present, there exists little operational capability to solve large-scale, dynamic network equilibrium models corresponding to the above technological concepts in IVHS. The broad goal of current research in this area is to formulate dynamic models which do correspond to the above objectives, and have some reasonable prospects for solution for large urban transportation networks on the present or next generation of computers. This book seeks to contribute to this important goal.

This book aims to present a new generation of dynamic network equilibrium models, incorporating dynamic travel choice problems including traveler's destination choice, mode choice, departure/arrival time choice and route choice. These models are expected to be able to function as off-line dynamic travel forecasting and evaluation tools, and eventually as real-time on-line models of urban transportation networks. Research on dynamic transportation network models is evolving very rapidly; a rich set of new formulations and solution algorithms are presented in this book. The full evaluation of these new models and solution algorithms is a matter for future research. Moreover, extensions to problems of location choice remain to be tackled.

We first describe the general dynamic travel choice problem. Travelers seeking to travel from their current locations to their specified destinations, and to depart *or* arrive at specified times, require best modes, departure times and routings for their trips. These needs could be provided in accordance with one of several objectives, such as the following:

1. each driver seeks to be routed onto the current best route at each intersection for current traffic conditions, given a specified departure time (we refer to this type of route choice as minimizing ***instantaneous*** travel time);

2. each driver seeks to minimize his or her *actual* travel time, given a specified departure time;

3. each driver agrees to accept a route at his or her specified departure time that minimizes the travel time of all vehicles traveling during a longer time period;

4. each traveler agrees to accept a mode, destination, departure time and route that minimizes the travel time of all travelers, but makes his or her arrival time as close to a specified target as possible.

Each of these objectives corresponds conceptually to a proposed route guidance system.

In addition to ATIS, Advanced Traffic Management Systems (ATMS) will predict traffic congestion and provide real time optimal control strategies for freeways and arterials. As one important component of ATMS, dynamic traffic control systems will respond to changing traffic conditions so as to control ramp flows to improve the efficiency of freeways, to maintain priorities for high-occupancy vehicles and to coordinate signal timing strategies across regional arterial network.

In conjunction with ATIS and ATMS, Advanced Public Transportation Systems (APTS) is another important component of IVHS. APTS apply advanced navigation, information and communication technologies that most benefit public transportation. Major benefits are also expected because the application of these technologies will attract travelers to transit and ridesharing modes, thereby reducing traffic congestion, atmospheric emissions and energy consumption.

Figure 1.1 shows the framework of an integrated APTS/ATMS/ATIS system. The analysis of such systems is also summarized by Kaysi et al (1993).

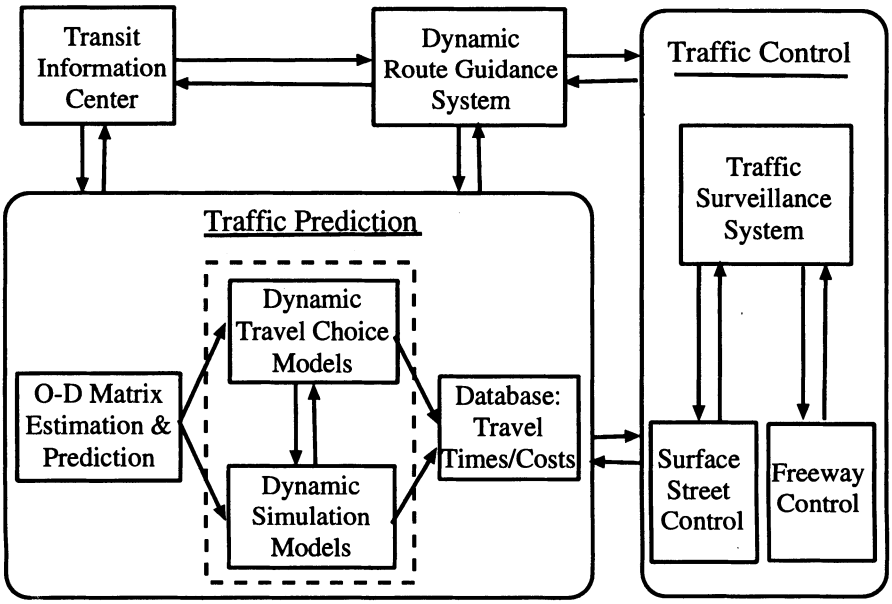



Figure 1.1: Structure of an Integrated APTS/ATMS/ATIS System

This book seeks to present our recent research findings on formulating dynamic models which bear a reasonable correspondence to the above objectives. The problems of dynamic travel choice (destination, mode, departure time, route choice, enroute diversion and parking) are described in a combined modeling framework. Using the optimal control theory approach, various dynamic travel choice models are formulated for a congested transportation network. If the analysis time period is discretized, these optimal control problems reduce to nonlinear programming problems. Solution algorithms are presented for solving some of these problems. Finally some numerical examples illustrate the properties of these models. In the following, we first analyze the transportation network models in general. Then, we review the development of dynamic network models and describe the hierarchy of models in this book.

## 1.2 Urban Transportation Network Analysis

Existing urban transportation network models or travel choice models can be classified according to the application purposes: long term and short term. Urban travel choice modeling activities to support the planning of urban freeway and transit networks began in the mid-50s. Initially, models were oriented towards the task of evaluating alternative land use patterns and transportation network proposals. Later, some efforts to consider the optimal design and extension of land use and transportation networks were undertaken with modest success. (Boyce et al, 1970).

Urban travel choice models are usually associated with network equilibrium concepts. The models for long-term planning purpose are usually static. Most of the static user equilibrium (UE) models are established to be consistent with the Wardrop's first principle (Wardrop, 1952). This principle requires *for used routes between a given origin-destination pair that the route cost equals the minimum route cost, and that no unused route has a lower cost.* This user equilibrium has been employed as the key behavioral assumption in most static urban transportation network models.

The first mathematical programming formulation for the static user equilibrium problem was proposed by Beckmann, McGuire and Winsten (1956) as an equivalent optimization problem. Their formulation considered the general case in which origin-destination flows are determined by a demand function. The fixed demand case follows immediately from their result. This formulation allows the derivation of existence, uniqueness (in terms of link flows) and optimality conditions of the solution, satisfying Wardrop's UE principle. This model was studied extensively by Dafermos and Sparrow (1969) and efficiently solved by LeBlanc et al (1975) who used the minimum-cost-route algorithm to implement the Frank-Wolfe algorithm. Sheffi (1985) gives a comprehensive account of the static UE problem. More recently, the more general problem with asymmetric link interactions had been addressed by Smith (1979) and Dafermos (1980, 1982), as have other variants of the basic formulation. The asymmetric link interactions require the model to be formulated as a variational inequality (VI) problem (Nagurney, 1993).

In the category of static user equilibrium models, there are many extensions which incorporate joint trip distribution/mode/route choice (Florian and

Nguyen, 1978) and residential location (Boyce, 1980). The above deterministic models are based on the assumption that drivers have perfect information and comply with the UE travel choice criterion.

Another important area is stochastic travel choice modeling. The initial stochastic travel choice model for stochastic network loading was proposed by Dial (1971) who described a flow-independent logit route choice model. Daganzo and Sheffi (1977) formulated a stochastic user-equilibrium (SUE) route choice model which is a generalization of the UE criterion, defined as follows: *at SUE, no traveler can improve his or her perceived travel time by unilaterally changing routes.* The SUE models are more realistic than the deterministic UE models since SUE assumes drivers to have less than perfect information when choosing routes. Detailed reviews of user-equilibrium models have been prepared by Friesz (1985), Florian (1986) and Boyce et al (1988).

In addition to models for long term analyses, there are models for short term purposes, which can be termed dynamic models. Previously, most network equilibrium models focused on the static description of traffic flows on the network, implying that flows and travel times are invariant over the duration of the peak period. In response to the untenable assumption of static traffic flow over the entire peak period, several models of dynamic route choice have been proposed. These short term or dynamic models can predict large time-dependent variations of traffic in a road network and should be able to predict the travel times of vehicles during their journey. Thus they are appropriate for the assessment of the impact of IVHS systems on highway network performance. In the long run, they should be useful in managing the real time operations of IVHS systems. A comprehensive review of dynamic models is given in the next section.

In dynamic transportation networks, the traditional traveler behavior assumption of the static models needs significant revision to consider short term variations of traffic. Thus, we are no longer considering a day-to-day traffic equilibrium. Instead, we are trying to influence or control traffic and travel patterns of travelers optimally by providing accurate traffic information (such as travel time) and effective traffic control measures. Therefore, we will no longer use the term "user-equilibrium" for dynamic traffic. Instead, we use "user-optimal" to represent the objective we are seeking to achieve.

In an ATIS system, there are two kinds of information available to travelers: current and future travel time information. Current travel time information can be obtained using the currently prevailing instantaneous link travel times. Future travel time information can be obtained using predicted link travel times. We discuss these two kinds of travel times in more detail in Chapter 4. More generally, both time and cost are considered.

For a dynamic transportation network, the types of control might be classified as follows: total control; partial control; no control (Lo et al, 1994). *Total control* mandates full compliance by travelers; that is, travelers are not given any choice as to whether they would like to comply or not. Some examples include traffic signal control at intersections and ramp metering at an entrance to a freeway. *Partial control* tries to actively influence traffic patterns, but compliance is not mandatory; for example, route guidance from changeable message signs. By providing information on best routes to travelers, traffic patterns may be altered and more travellers can be diverted to less congested routes or time periods. It is up to the travelers themselves to decide whether to follow the advice. This kind of control is rather flexible. *No control* provides real-time traffic information to the travelers, but specific routing is not proposed for individual travelers. It is entirely up to the travelers themselves to decide what to do with the information. Current radio broadcasts on traffic congestion and accident information in metropolitan areas is a simple example of no control.

Each type of control in the above classification requires a substantial modeling effort for dynamic traffic analysis and evaluation. The development of models and algorithms for optimizing flows in real time on transportation networks will be fundamental to the success of the ATMS/ATIS component of IVHS. The near term objective of ATMS/ATIS systems is to provide accurate traffic information, mainly travel time information so that travelers may adjust their travel patterns individually. In the long run, when more and more travelers are equipped with more mature ATIS systems, the coordination of route and departure time choices will become crucial. It is in this context that some of our dynamic network models are proposed.

## 1.3 Overview of Dynamic Network Models

Dynamic network models can be classified as flow-based models and vehicle-based models. Flow-based models are based on macroscopic flow equations; vehicle-based models are based on microscopic movement of vehicles. Flow-based models are more applicable for large-scale transportation networks, since mathematical representations and capabilities appropriate for the corresponding dynamic network problems are more suitable.

Vehicle-based models comprise both simulation and optimization models. INTEGRATION (Van Aerde, 1992) and DYNASMART (Mahmassani et al, 1992) are examples of simulation models. Ghali and Smith (1993) presented a set of dynamic network models using packets to represent traffic flows on links. Their models are basically vehicle-based. Lafortune et al (1991) also presented an integer-based dynamic system-optimal (DSO) traffic assignment model.

There have been two stages of development of flow-based dynamic network models. Yagar (1971), Hurdle (1974) and Merchant and Nemhauser (1978a) were among the first to consider dynamic models for congested traffic networks. But the assumptions of these models are very limiting and they are unsuitable for application to general large-scale networks. Important breakthroughs began to occur in the late 1980s when IVHS ignited the potential applicability of such models to the next generation of surface transportation systems.

The study of dynamic route choice models over a general road network was begun by Merchant and Nemhauser (1978a, 1978b) who presented a dynamic system-optimal (DSO) route choice model for a many-to-one network. Subsequently, Carey (1987) reformulated the Merchant-Nemhauser problem as a convex nonlinear program which has analytical and computational advantages over the original formulation. DSO route choice models over a multiple origin-destination (O-D) network were established by using optimal control theory (Friesz et al, 1989; Ran and Shimazaki, 1989a). Recently, many simulation-based DSO route choice models have also been proposed by various researchers, especially for freeway corridor problems (Chang et al 1993).

An important dynamic generalization of the static UE concept is called dynamic user-optimal (DUO) route choice. One dynamic user-optimal (DUO) route choice problem is to determine vehicle flows at each instant of time on each link resulting from drivers using minimal-time routes. Friesz et al (1989) proposed a DUO route choice model by considering the equilibration of instantaneous unit route costs. Furthermore, a generalized DUO route choice model over a multiple origin-destination network was presented by Wie, Friesz and Tobin (1990). In the formulation of some dynamic models (Friesz et al, 1989; Wie, 1989; Ran and Shimazaki, 1989a), only the inflow into each link at a given time is defined as a *control variable*; the exit flow from each link is considered to be a *function* of the number of vehicles on that link. If the exit flow function is nonlinear, it is impossible to establish an optimization model of DUO route choice for a network for multiple origin-destination pairs. By defining the exit flow as a *control variable*, Ran and Shimazaki (1989b) presented a DUO route choice model which considered the equilibration of instantaneous travel times. Subsequently, Ran, Boyce and LeBlanc (1993) formulated a set of new instantaneous DUO route choice models with flow propagation constraints. Some of the basic constraints for a dynamic network model were also discussed in Ran et al (1992a). Among other dynamic network models, Janson (1991) presented a set of dynamic network models using average link travel time/flow relationships and proposed a heuristic solution algorithm.

Further studies concern the extension of deterministic dynamic network models to *stochastic* dynamic network models. Vythoulkas (1990) developed a logit-type stochastic dynamic route choice model for a general network. However, some of the key constraints, such as the flow propagation, are missing, which results in unrealistic traffic flows under general network conditions. Cascetta (1991) and Cascetta et al (1993) presented a dynamic stochastic route choice model for day-to-day route choices using a stochastic process theory approach. Ran et al (1992) formulated two logit-type stochastic dynamic user-optimal (SDUO) route choice models considering both instantaneous and actual travel times for general transportation networks.

Computational issues of dynamic route choice problems have received increasing attention in recent years. Following Merchant and Nemhauser's (1978b) proposal of a conceptual algorithm for solving a single-destination

DSO route choice model, Ho (1980) solved the same model by successively optimizing a sequence of linear programs. Subsequently, Ho (1990) presented a nested decomposition algorithm for the same problem and implemented this algorithm on a hypercube computer. Ran and Shimazaki (1989) proposed a time decomposition algorithm to solve a multiple-destination DSO assignment model. Using the time-space expansion technique, Boyce et al (1991) presented a Frank-Wolfe algorithm to solve an instantaneous DUO route choice model. Codina and Barcelo (1991) also applied a time decomposition algorithm to solve a preliminary DUO route choice model. Janson (1993) solved a combined departure time/route choice problem using a heuristic based on the Frank-Wolfe algorithm.

The choice of departure time has been addressed by several researchers, including Abkowitz (1981) and Hendrickson and Plank (1984), who developed work trip scheduling models. De Palma et al (1983) and Ben-Akiva et al (1984) modeled departure time choice over a simple network with one bottleneck using the general continuous logit model. Mahmassani and Herman (1984) used a traffic flow model to derive the equilibrium joint departure time and route choice pattern over a parallel route network. Mahmassani and Chang (1987) further developed the concept of equilibrium departure time choice and presented the boundedly-rational user equilibrium concept under which all drivers in the system are satisfied with their current travel choices, and thus feel no need to improve their outcome by changing to an alternate choice. Friesz et al (1993) formulated a simultaneous departure time/route choice model using the variational inequality approach. An overview of our modeling activity in dynamic departure time, mode and route choices was presented in LeBlanc et al (1992); in addition, papers presented at various conferences formed the basis for several chapters of this book.

## 1.4 Hierarchy of Dynamic Network Models

In this section, we give an overview of the chapters of the book and their interrelationships. In Chapter 2, we summarize the basic concepts and principles of continuous time optimal control theory. This background knowledge provides a basis for the formulation of dynamic network models in this book. Discrete time optimal control problems are introduced in Chapter 3. Then, nonlinear programming (NLP) problems and their similarity to the discrete time optimal control problems are discussed. The conventional Frank-Wolfe algorithm and diagonalization technique are presented for solving NLP problems. For more advanced readers, an introduction to variational inequality theory is also provided.

We began our dynamic network model research from the dynamic route choice problem, also known as the dynamic traffic assignment problem. Based on the two types of travel times— instantaneous link travel times and actual link travel times, two route choice models can be formulated. The relation of these dynamic route choice problems is depicted in Figure 1.2. The discussion of the two types of travel times is given in Chapter 4 together with all the constraint conditions necessary for dynamic network models. In addition to flow conservation conditions, flow propagation constraints are especially emphasized. Other important constraints include link capacity and spillback constraints.

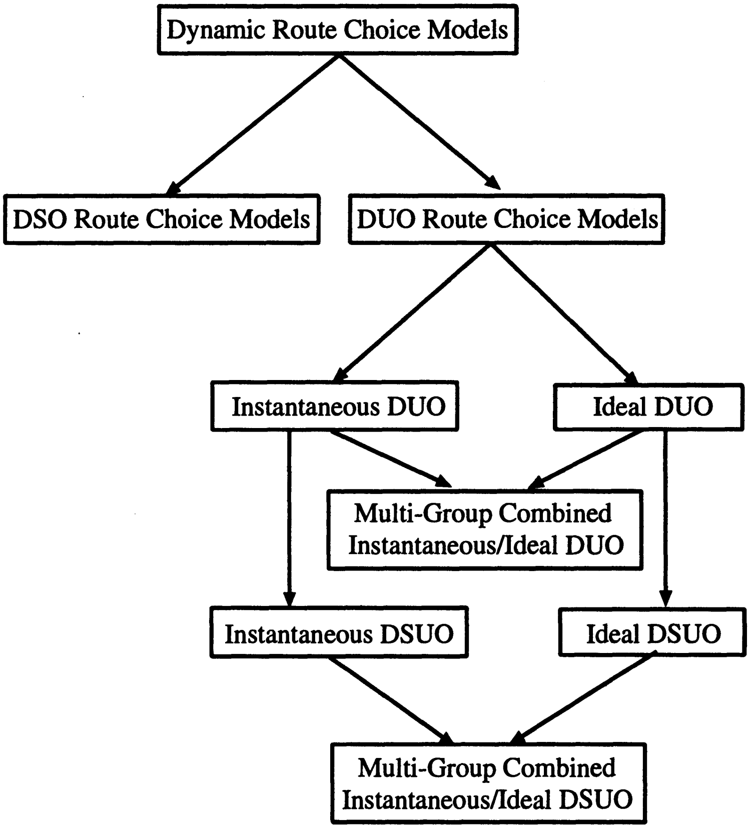



Figure 1.2: A Hierarchy of Dynamic Route Choice Models

In Ran and Shimazaki (1989b), an instantaneous DUO route choice model that avoids the use of a link exit flow function was presented. Using the same approach, three instantaneous DUO optimal control models are presented in Chapter 5. The trip pattern is assumed to be known *a priori* in the dynamic route choice problem. We define a decision node for each route between each O-D pair as any node on that route including the origin. The instantaneous route travel time between a decision node and the destination node is calculated by using the currently prevailing link times. The instantaneous DUO route choice problem is to determine vehicle flows at each instant of time on each link resulting from drivers using minimal-time routes under currently prevailing travel times. The instantaneous DUO route choice model is formulated using the optimal control theory approach. These models are fundamentally different from earlier DSO and DUO route choice models by:

1. using a different definition of DUO;

2. employing exit flows as a set of control variables rather than functions;

3. including new formulations of the objective function;

4. including a formulation of flow propagation constraints.

In Chapter 6, the continuous time formulation of the DUO route choice program is transformed into a discrete time NLP formulation. Since the model is convex, the discrete version should be efficiently solvable for the optimal solution for large networks. We present a new algorithm for solving this NLP in Chapter 6. It is solved by the Frank-Wolfe technique embedded in a diagonalization procedure. In the diagonalization procedure, the estimated link travel times are updated iteratively and the Frank-Wolfe technique is applied in each iteration to solve the resulting NLP. For the linearized Frank-Wolfe subproblem, an expanded time-space network is constructed so that the subproblem can be decomposed according to O-D pairs and can be viewed as a set of minimal-cost route problems. Flow propagation constraints representing the relationship between link flows and travel times are satisfied in the minimal-cost route search so that only flow conservation constraints for links and nodes remain. The proposed formulation has computational advantages since the gradient vector of the objective function with respect to the control and state variables is always nonnegative which allows for much more efficient minimal-cost route calculations. Preliminary computational results from applying the algorithm to a test network are reported.

In Chapter 7, we propose another concept of DUO route choice which reflects *ideal* route choice behavior of travelers. The formulation of the ideal DUO route choice problem is based on the underlying choice criterion that each traveler uses the route that minimizes his/her future (actual) travel time when departing from the origin to his/her destination. Thus, for any O-D pair, vehicles departing the origin at the same time must arrive at the destination at the same time under ideal DUO route choice conditions. In this chapter, an optimal control program of ideal DUO route choice model is presented. A solution algorithm based on a penalty and diagonalization/Frank-Wolfe algorithm is presented.

In Chapter 8, we consider two stochastic dynamic user-optimal (SDUO) route choice models which are stochastic extensions of our previous deterministic DUO route choice models. The formulation of the instantaneous SDUO route choice problem is based on the underlying choice criterion that each traveler uses the route that minimizes his/her perceived instantaneous travel time when departing from any decision node to his/her destination. The solution of this instantaneous SDUO model results in instantaneous stochastic network flows at each decision node based on a logit function of mean (measured) instantaneous travel times of alternative routes. In parallel, we present an ideal SDUO route choice model based on stochastic flows with a logit function of mean future travel times experienced by drivers over alternative routes for each O-D pair.

In Chapter 9, we first discuss solution algorithms (DYNASTOCH) for flow-independent instantaneous SDUO and ideal SDUO route choice problems. These algorithms are very similar to STOCH algorithm proposed by Dial (1971) for static logit-type assignment. Then, we use the diagonalization technique and DYNASTOCH to solve our instantaneous SDUO and ideal SDUO route choice problems. One important advantage of these algorithms is that route enumeration is avoided.

Next we study a joint dynamic departure time and route choice problem in Chapter 10. In this problem, travelers' departure times can be shifted based on origin-destination travel times corresponding to each possible departure time. This problem, then, is to determine travelers' departure times and choose their best routes at each instant of time. We present a bilevel programming formulation of this DUO departure time and route choice problem. The model extends our previous DUO model to the case where both departure time and route over a general road network must be chosen. Our lower-level program solves the DUO departure time choice problem, and our upper-level program solves the DUO route choice problem. The optimality conditions of the bilevel program demonstrate that our formulation is consistent with the desired DUO departure time and route choice properties. We suggest a heuristic algorithm for solving the bilevel program (the upper problem can be solved exactly). A numerical example illustrates that total travel time can be decreased by choosing appropriate departure times.

In Chapter 11, a combined DUO mode/departure time/route choice model with multi-class travelers is presented for a general transportation network. Our model extends the dynamic user-optimal departure time/route choice conditions to include mode choice as well. The model extends the earlier combined departure time/route choice model to the case where dynamic traffic flows by different modes affect other modes' costs. By formally stratifying travelers into different groups, an accurate analysis of the time-cost tradeoff in mode choice is possible. This model presents a two-stage non-hierarchical programming formulation of this DUO mode/departure time/route choice problem. The first-stage program solves the dynamic mode choice problem. Simultaneously, the second-stage program represents a hierarchical leader-follower program which solves the DUO departure time and route choice problem for motorists. The optimality conditions of the two-stage program demonstrate that our formulation is consistent with the DUO mode/departure time/route choice conditions. The hierarchy of dynamic travel choice models is summarized in Figure 1.3.

In Chapter 12, we extend our dynamic network models to a broader framework using the variational inequality (VI) approach. We first discuss

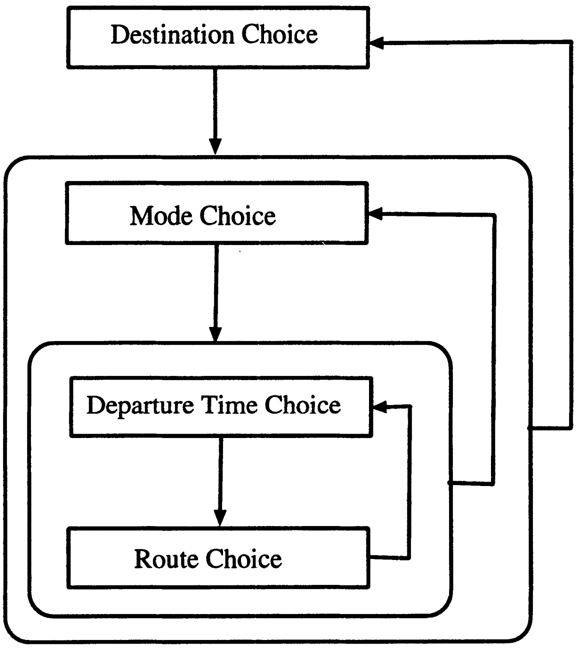



Figure 1.3: A Hierarchy of Dynamic Travel Choice Models

instantaneous DUO route choice problems and formulate route-based and link-based VI models for single group and multi-group travelers. The link-based models are computationally more tractable than the route-based models. The relationship of VI models and optimal control models is also investigated. We show that the optimal control models in Chapter 5 are special cases of general VI models.

In Chapter 13, both route-based and link-based VI models for ideal DUO route choice problems are formulated, and their relationship to optimal control models is investigated. We show that VI models can be reformulated as optimal control models under relaxation and solved using the diagonalization/Frank-Wolfe algorithm. General VI formulations for joint departure time/route choice problems are presented in Chapter 14; both route-based and link-based VI models are discussed for this joint dynamic travel choice problem. We show that VI models can be reformulated as optimal control models under relaxation.

In Chapter 15, a set of DSO route choice models are formulated. For comparison, we discuss in detail a DSO route choice model which minimizes total travel time under the same set of constraints as the instantaneous DUO route choice models in Chapter 5. Subsequently, a set of DSO route choice models with elastic departure times are discussed. Specifically, the time-minimizing problem is emphasized in the context of emergency evacuation. Furthermore, two types of dynamic congestion pricing schemes are presented.

In order to implement the above dynamic network models in realistic urban transportation networks, we investigate time-dependent travel time functions for signalized arterial and freeway segment links in Chapter 16. Dynamic link travel times are first classified according to various applications. Subsequently, travel time functions for arterial links with longer and shorter time horizons are discussed separately, and two sets of functions are recommended for dynamic transportation network problems. Implications of those functional forms are analyzed and some modifications for dynamic network models are suggested. In addition, based on dynamic link travel time functions, we discuss how many independent variables are necessary to describe the temporal traffic flow and properly estimate the time-dependent travel time over an arterial link. As a result, six link flow variables and corresponding link state equations are proposed as the basis for formulating dynamic transportation network models. Finally, time-dependent travel time functions for freeway segment links are recommended.

Various implementation issues are discussed in Chapter 17. Among those issues, we focus on the following items: dynamic traffic prediction; dynamic traffic control; incident management; dynamic congestion pricing; operations and control for automated highway systems (AHS); dynamic transportation planning. The application of dynamic network models to these tasks is discussed specifically. In addition, we analyze how these models can serve operating and evaluation functions in IVHS systems. The detailed requirements of IVHS systems for dynamic network models are also identified. Subsequently, we discuss data needs of dynamic network models to accomplish the above implementation tasks. The following items are identified as the most important: time-dependent O-D matrices; network geometry data and intersection/ramp control data; traffic flow data for calibrating various dynamic link travel time functions; traveler information for stratifying travelers into multiple groups and calibrating travel disutility functions.

The flowchart of the logical sequence of this book is shown in Figure 1.4. Throughout this book, the main emphasis is on dynamic travel choice models formulated using the optimal control theory approach. For dynamic route choice problems, we extend deterministic models to stochastic models (Chapters 8 and 9). Then, we discuss more general VI formulations for both route choice and departure time choice problems in Chapters 12–14. Thus, for readers who are not familiar with dynamic transportation network models, optimal control models (Chapters 4-7) may be the most interesting. For advanced readers, Chapters 8 and 9 as well as Chapters 12–14 provide a broader picture of models and a good starting point for future research. In general, Chapters 1–7 can be used as a classroom reference for advanced transportation network courses, especially for those with a strong IVHS orientation.

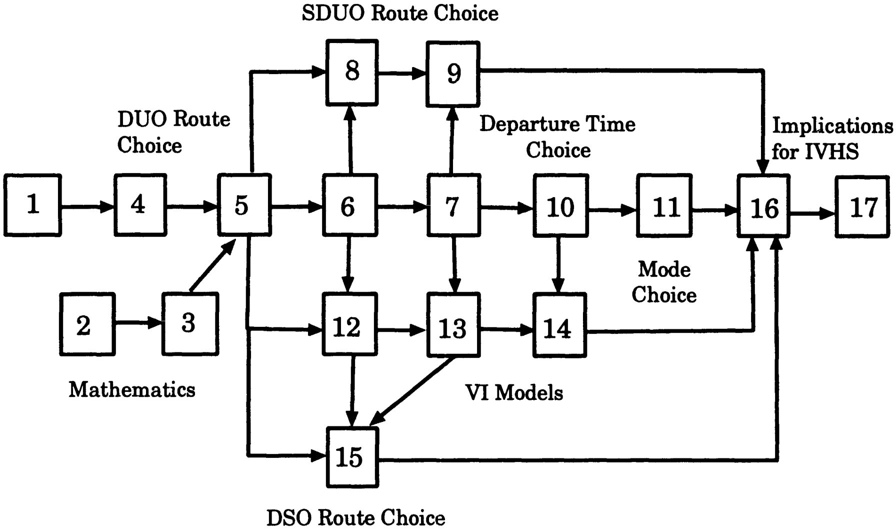



Figure 1.4: Logical Flowchart of the Chapters

## 1.5 Notes

The success of ATMS/ATIS systems depends on development and successful applications of advanced transportation models and control models. This field will need to develop models and algorithms that use real-time data to determine optimal control strategies of traffic. They must enable real-time management of traffic while accommodating both pre-trip planning and en route travel plan modification. They must also provide the means of evaluating the benefits of various aspects of IVHS. Technologies that are likely to be useful include artificial intelligence, expert systems and parallel computing. Traffic modeling for ATMS/ATIS includes:

1. travel forecasting models;
2. optimal routing methods;
3. support systems for traffic management centers;
4. dynamic route choice models;
5. traffic simulation models;
6. network-wide optimization programs;
7. driver/traveler behavior models (human factors).

In addition, developments in traffic modeling should take into account multi-modal travel requirements, such as HOV or public transit, and should also include the development of the following capabilities:

1. real-time, traffic-adaptive logic for signal control;
2. real-time, traffic-adaptive logic for freeway control, including ramp and possibly mainline metering;
3. real-time integration of freeway and surface street control;
4. transit and emergency vehicle priorities.

Research into each of the above tasks is a long term effort. This book is intended to provide a framework and some background knowledge to study and investigate the above complicated application problems. We developed those dynamic network models bearing the above tasks in mind. Nevertheless, application of dynamic network models in IVHS is still premature. The main concerns for dynamic network models which prevent their large-scale applications include:

1. accurate representation of traffic propagation and travel time functions for links;
2. accurate representation of travel choice behavior;
3. validation of models using real-time data.

However, extensive potential applications of dynamic network models can be expected in future ATIS/ATMS systems, including dynamic route guidance, freeway ramp control, arterial signal optimization/coordination and automated highway, etc. We expect to summarize those applications in a subsequent book.

# Chapter 2

# Continuous Optimal Control Problems

Optimal control theory has been extensively used in solving many engineering problems, such as in mechanical engineering and aeronautics engineering. Its application in transportation engineering has been limited to traffic signal control on surface streets and ramp metering control on freeways. Recently, with the rapid advance of supercomputing facilities and techniques, solution of optimal control theory formulations of large-scale problems has become feasible. Therefore, the application of this approach to dynamic transportation network modeling is attractive. The objective of conventional optimal control theory is to determine optimal control strategies that will cause a process to satisfy the physical constraints and at the same time minimize or maximize some performance criterion. In this book, we will use optimal control theory to formulate and analyze time-dependent transportation network problems. Those optimal control models have many similarities with the optimization models for solving static counterparts of these problems which are formulated and solved using nonlinear programming theory.

In this chapter, we review some basic concepts of optimal control theory, which are sufficient to provide a basis for formulating dynamic transportation network models and analyzing the conditions necessary for the existence of an optimal solution. In Section 2.1, we present definitions which are associated with any optimal control problem. In Section 2.2, we discuss optimal control problems assuming that the admissible controls are not bounded (no constraints) and derive the corresponding necessary conditions. These necessary conditions are then employed to find the optimal control law for the important linear regulator problem. Furthermore, Pontryagin's minimum principle is introduced as a generalization of the fundamental theorem of the calculus of variations.

In Section 2.3, optimal control problems with general bounded control and state variables are discussed. For problems with bounded control and state variables, we discuss a generalization of Pontryagin's Minimum Principle and the corresponding partial differential equations of the necessary (optimality) conditions derived using a general Hamiltonian function. In Section 2.4, we discuss some optimal control problems with simple constraints on control and state variables. These optimal control problems will be used in the analysis of dynamic transportation network models in later chapters. In Section 2.5, bilevel optimal control problems are discussed; here we are concerned with interrelationships between two interdependent optimal control problems.

## 2.1 Definitions for Optimal Control Theory

The basic theory of optimal control is derived from the calculus of variations. Compared with the calculus of variations, optimal control theory has many advantages in solving time-dependent optimization problems. It also has many similarities with static optimization problems, such as nonlinear programming (NLP) theory. We will discuss these relationships in detail in Chapter 3.

Before presenting the optimality conditions for the optimal control problem, we introduce some basic definitions. The formulation of an optimal control problem is associated with a time-dependent process or system and requires the following:

1. a mathematical representation of the process to be controlled or to be optimized;
2. a statement of the physical constraints;
3. specification of a performance criterion or an objective function.

### Dynamic Process/Dynamic System

The modeling of a dynamic process seeks to obtain the simplest mathematical description that adequately predicts the response of the physical system to all anticipated inputs. We restrict our discussion to systems described by ordinary differential equations. We define

$$x_1(t), x_2(t), \cdots, x_n(t)$$

as the *state variables* (or simply the *states*) of the process at time $t$, and

$$u_1(t), u_2(t), \cdots, u_m(t)$$

as the *control variables* (or the *controls*) to the process at time $t$. The system or process may be described by $n$ first-order differential equations, where $\dot{x}_i(t) = dx_i(t)/dt$:

$$\dot{x}_1(t) = f_1[x_1(t), x_2(t), \cdots, x_n(t), u_1(t), u_2(t), \cdots, u_m(t), t]$$

$$\dot{x}_2(t) = f_2[x_1(t), x_2(t), \cdots, x_n(t), u_1(t), u_2(t), \cdots, u_m(t), t]$$

$$\vdots \tag{2.1}$$

$$\dot{x}_n(t) = f_n[x_1(t), x_2(t), \cdots, x_n(t), u_1(t), u_2(t), \cdots, u_m(t), t]$$

We define the *state vector* of the system as

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ \vdots \\ x_n(t) \end{bmatrix}$$

and the *control vector* as

$$\mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ \vdots \\ u_m(t) \end{bmatrix}$$

The state equations can then be written in vector notation as

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \tag{2.2}$$

where $\mathbf{f}$ is also a vector.

### State Variables

The starting point for optimal control investigations is a mathematical model in state variable form. Why are state variables used in control problems? Formulating a mathematical model in state variable form is convenient because:

1. the concept of state has a strong physical motivation;
2. the state variable form is easy to use in theoretical investigations and the resulting differential equations are suitable for digital or analog solution;
3. the state form provides a unified framework for the study of nonlinear and linear systems.

Referring to the state of a system, we have the following definition:

**Definition 2.1** *The state of a system is a set of quantities $x_1(t)$, $x_2(t)$, $\cdots$, $x_n(t)$ which, if known at $t = 0$, are determined for $t \geq 0$ by specifying the inputs to the system for $t \geq 0$.*

### System Classification

Systems are described by the terms linear, nonlinear, time-invariant and time-variant (explicit function of time). We shall classify systems according to the form of their state equations. For example, if a system is nonlinear and time-variant, the state equations are written as

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \tag{2.3}$$

When the system does not depend on time $t$ explicitly, a special form of the above nonlinear dynamic systems is represented by equations of the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t)] \tag{2.4}$$

which is time-invariant. Many dynamic transportation network problems have time-invariant state equations. If a system is linear and time-invariant, its state equations are

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \tag{2.5}$$

where $\mathbf{A}(t)$ and $\mathbf{B}(t)$ are $n \times n$ and $n \times m$ matrices with time-dependent elements. State equations for one special type of linear dynamic systems have the form

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{2.6}$$

where $\mathbf{A}$ and $\mathbf{B}$ are constant matrices.

### Performance Measure/Objective Function

In order to evaluate the performance of a system quantitatively, the system controller needs to select a performance measure or an objective function. An optimal control is defined as one that minimizes (or maximizes) this objective function. In certain cases, the objective function may be self-explanatory, whereas in other problems the objective function may be artificial. For example, the statement, "To minimize the total travel time over a traffic network," clearly indicates that the objective function is defined from a system point of view. On the other hand, the statement, "To achieve equal minimal travel times over each used route for each O-D pair," does not directly suggest an objective function. In such a problem, the modeler may need to design an artificial objective function so that the resulting optimization conditions of the optimal control program represent desired physical properties. However, such an objective function may not have any direct physical interpretation. We shall discuss the definition of an objective function in more detail in the following chapters for various dynamic network problems.

In the following, we assume that the *performance* of a system can be evaluated by a measure $J$ of the form

$$J = \int_0^T F[\mathbf{x}(t), \mathbf{u}(t), t]\, dt + S[\mathbf{x}(T), T] \tag{2.7}$$

where 0 and $T$ are initial and final times; $F$ and $S$ are scalar functions. $T$ may be pre-specified or changeable according to the problem requirement. If $T$ is changeable, it is termed "free". $F$ is a cost term that depends on the state and control variables at any time $t$. $S$ is associated with final state $\mathbf{x}(T)$ and the final time $T$. It is sometimes termed a salvage cost in economics.

Starting from the initial state $x(0) = x_0$, assigning values to control variables $u(t)$ for $t \in [0, T]$ will cause a system to achieve some value for each state variable. The sequence of values achieved by each state variable through time is termed a *state trajectory*. Minimization of the objective function with respect to the control variables assigns unique real values to each trajectory of the system so that some optimal controls can be found.

We now present an explicit statement of the *optimal control problem* (OCP). First, some definitions of various forms of optimal controls are presented.

**Definition 2.2** *If a functional relationship of the form*

$$\mathbf{u}^*(t) = \mathbf{f}[\mathbf{x}(t), t] \tag{2.8}$$

*can be found for the optimal control at each instant of time t, then the function* $\mathbf{f}$ *is called an optimal control law.*

Notice that equation (2.8) implies that $\mathbf{f}$ is a rule which determines the optimal control at time $t$ for any admissible state value at time $t$. For example, if

$$\mathbf{u}^*(t) = \mathbf{G}\,\mathbf{x}(t) \tag{2.9}$$

where $\mathbf{G}$ is an $m \times n$ matrix of real constants, then we say that the optimal control law is linear and time-invariant.

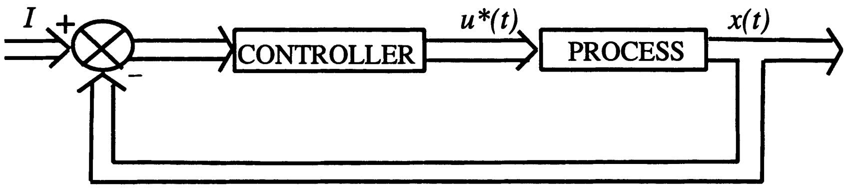
**Definition 2.3** *If the optimal control is determined as a function of time for a specified initial state value, i.e.,*

$$\mathbf{u}^*(t) = \mathbf{e}[\mathbf{x}(0), t] \tag{2.10}$$

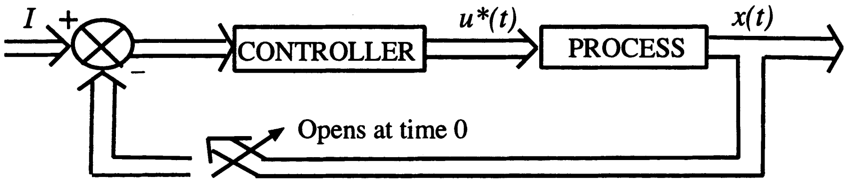
*then the optimal control is said to be in open-loop form.*

Thus, the optimal open-loop control is optimal only for a particular initial state value. In other words, if the optimal control law is known, the optimal control history starting from any initial state value can be generated.

Conceptually, it is helpful to imagine the difference between a closed-loop optimal control law and an open-loop optimal control law as shown in Figure 2.1. Although engineers normally prefer closed-loop solutions to optimal control problems, there are cases when an open-loop control may be desirable. For example, in radar tracking of a satellite, once the orbit is set, very little can happen to cause an undesired change in the trajectory parameters. In this situation a pre-programmed control for the radar antenna might well be used. However, in dynamic transportation network problem, it is hard to find an open-loop control in general. Thus, open-loop optimal control is not our major focus.




Closed-Loop Optimal Control




Open-Loop Optimal Control

Figure 2.1: Closed-Loop and Open-Loop Optimal Controls

**Definition 2.4** *If the optimal control is determined as a function of the state variable only, i.e.,*

$$\mathbf{u}^*(t) = \mathbf{f}[\mathbf{x}(t)] \tag{2.11}$$

*then the optimal control is an optimal feedback control.*

A typical example of feedback control is the classic servomechanism problem where actual and desired outputs are compared and any deviation produces a control signal that attempts to reduce the discrepancy to zero. In dynamic transportation network problems, there are many applications of such a control law.

With the above basic definitions in mind, we begin our discussion of optimal control problems in next three sections. First, we present the simplest form of optimal control problem which has no constraints on state and control variables.

## 2.2 Continuous Problems with No Constraints

In this section, we mainly present two types of optimal control problems which differ from each other by the specification of either a fixed end time or a free end time.

### 2.2.1 Fixed Beginning and Fixed End Times

We now consider a dynamic system having a cost function $F[\mathbf{x}(t), \mathbf{u}(t), t]$ for a fixed time period $[0, T]$. The optimal control problem is to seek a control function $\mathbf{u}$ which is feasible, or admissible, in order to minimize the objective function as follows:

$$\min_{\mathbf{x},\mathbf{u}} \quad J = \int_0^T F[\mathbf{x}(t), \mathbf{u}(t), t]\, dt + S[\mathbf{x}(T), T] \tag{2.12}$$

s.t.

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \tag{2.13}$$

$$T \quad \text{and} \quad \mathbf{x}(0) = \mathbf{x}_o \text{ fixed;} \quad \mathbf{x}(T) \text{ free} \tag{2.14}$$

where $F[\mathbf{x}(t), \mathbf{u}(t), t]$ has continuous partial derivatives with respect to $\mathbf{x}(t)$ and $S[\mathbf{x}(T), T]$ has continuous partial derivatives with respect to $\mathbf{x}(T)$.

We use the method of Lagrange multipliers to adjoin the system differential state equations to the objective function, which gives

$$\begin{aligned} I &= \int_0^T \{F[\mathbf{x}(t), \mathbf{u}(t), t] + \boldsymbol{\lambda}(t)\, [\mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] - \dot{\mathbf{x}}(t)]\}\, dt \\ &+ S[\mathbf{x}(T), T] \end{aligned} \tag{2.15}$$

where $\boldsymbol{\lambda}(t)$ is a vector of Lagrange multipliers associated with the dynamic state equations. It is standard in optimal control theory to designate the major part of the integrand of equation (2.15) as the Hamiltonian,

$$\mathcal{H} = F[\mathbf{x}(t), \mathbf{u}(t), t] + \boldsymbol{\lambda}(t)\, \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \tag{2.16}$$

Then, the first-order necessary conditions for the optimal control problem may be summarized as follows:

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}(t)} = \frac{\partial F[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{u}(t)} + \boldsymbol{\lambda}(t) \frac{\partial \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{u}(t)} = \mathbf{0} \tag{2.17}$$

$$-\dot{\boldsymbol{\lambda}}(t) = \frac{\partial \mathcal{H}}{\partial \mathbf{x}(t)} = \frac{\partial F[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{x}(t)} + \boldsymbol{\lambda}(t) \frac{\partial \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{x}(t)} \tag{2.18}$$

$$\dot{\mathbf{x}}(t) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(t)} = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \tag{2.19}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \tag{2.20}$$

$$\boldsymbol{\lambda}(T) = \frac{\partial S[\mathbf{x}(T), T]}{\partial \mathbf{x}(T)} \tag{2.21}$$

Equations (2.19) and (2.20) are simply the restatement of state equation (2.13) and boundary condition (2.14). Equations (2.20) and (2.21) are termed the transversality conditions which constitute two-point boundary conditions for the set of differential equations. The above first-order necessary conditions are derived using Pontryagin's Minimum (or Maximum) Principle or the Hamilton-Jacobi equation. The Pontryagin's Minimum Principle can simply be stated as follows. The inequality

$$\mathcal{H}[\mathbf{u}^*(t)] \leq \mathcal{H}[\mathbf{u}(t)] \tag{2.22}$$

is valid for all admissible $\mathbf{u}(t)$, where $*$ represents that the solution is optimal. Sometimes, we directly use this principle to analyze complicated optimal control problems. It is also interesting to compute the total derivative with respect to time $t$ as

$$\begin{aligned} \frac{d\mathcal{H}}{dt} &= \frac{\partial F}{\partial t} + \dot{\mathbf{x}}(t)\left[\frac{\partial F}{\partial \mathbf{x}(t)} + \boldsymbol{\lambda}(t)\frac{\partial \mathbf{f}}{\partial \mathbf{x}(t)}\right] + \dot{\mathbf{u}}(t)\left[\frac{\partial F}{\partial \mathbf{u}(t)} + \boldsymbol{\lambda}(t)\frac{\partial \mathbf{f}}{\partial \mathbf{u}(t)}\right] \\ &+ \dot{\boldsymbol{\lambda}}(t)\,\mathbf{f} + \boldsymbol{\lambda}(t)\frac{\partial \mathbf{f}}{\partial t} \end{aligned} \tag{2.23}$$

Now, we consider that equation (2.17) may or may not hold. Substituting equations (2.17) and (2.19) into equation (2.23), we obtain

$$\frac{d\mathcal{H}}{dt} = \frac{\partial F}{\partial t} + \dot{\mathbf{u}}(t)\frac{\partial \mathcal{H}}{\partial \mathbf{u}(t)} + \boldsymbol{\lambda}(t)\frac{\partial \mathbf{f}}{\partial t} \tag{2.24}$$

Thus, if $F$ and $f$ are not explicit functions of time $t$, the Hamiltonian is constant along an optimal trajectory where $\partial\mathcal{H}/\partial\mathbf{u} = 0$. It can be shown that this is true along an optimal trajectory, even if we cannot require $\partial\mathcal{H}/\partial\mathbf{u} = 0$. Therefore, the following additional condition holds

$$\frac{d\mathcal{H}}{dt} = 0 \tag{2.25}$$

when $F$ and $f$ are not explicit functions of time. This is an important result which we will use in later developments. In the following, we discuss a special problem using the above first-order necessary conditions.

**The Linear Regulator**

We now study a particular optimal control problem which has its solution as a linear feedback control law. The typical problem is stated as follows. We have a linear differential system

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\,\mathbf{x}(t) + \mathbf{B}(t)\,\mathbf{u}(t), \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.26}$$

and wish to find the control which minimizes the objective function

$$\mathbf{J} = \frac{1}{2}\int_0^T [\mathbf{x}(t)\,\mathbf{Q}(t)\,\mathbf{x}(t) + \mathbf{u}(t)\,\mathbf{R}(t)\,\mathbf{u}(t)]\,dt + \frac{1}{2}\mathbf{x}(T)\,\mathbf{W}\,\mathbf{x}(T) \tag{2.27}$$

Without loss of generality, we assume $\mathbf{Q}(t)$, $\mathbf{R}(t)$ and $\mathbf{W}$ are symmetric matrices. We may solve this problem using the first-order necessary conditions or the Minimum Principle. The Hamiltonian is

$$\begin{aligned}\mathcal{H}[\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\lambda}(t),t] &= \frac{1}{2}\mathbf{x}(t)\,\mathbf{Q}(t)\,\mathbf{x}(t)+\frac{1}{2}\mathbf{u}(t)\,\mathbf{R}(t)\,\mathbf{u}(t)\\ &+ \boldsymbol{\lambda}(t)\,\mathbf{A}(t)\,\mathbf{x}(t)+\boldsymbol{\lambda}(t)\,\mathbf{B}(t)\,\mathbf{u}(t) \qquad (2.28)\end{aligned}$$

Application of the first-order necessary conditions requires, for an optimal control, that

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}(t)} = \mathbf{0} = \mathbf{R}(t)\,\mathbf{u}(t)+\boldsymbol{\lambda}(t)\,\mathbf{B}(t) \qquad (2.29)$$

and

$$\frac{\partial \mathcal{H}}{\partial \mathbf{x}(t)} = -\dot{\boldsymbol{\lambda}}(t) = \mathbf{Q}(t)\,\mathbf{x}(t)+\boldsymbol{\lambda}(t)\,\mathbf{A}(t) \qquad (2.30)$$

with the boundary condition

$$\boldsymbol{\lambda}(T) = \frac{\partial \mathbf{S}[\mathbf{x}(T)]}{\partial \mathbf{x}(T)} = \mathbf{W}\,\mathbf{x}(T) \qquad (2.31)$$

Thus, from equation (2.29), we require that

$$\mathbf{u}(t) = -\mathbf{R}^{-1}(t)\,\mathbf{B}(t)\,\boldsymbol{\lambda}(t) \qquad (2.32)$$

We shall inquire whether we can convert this to a closed-loop control by assuming that the solution for the Lagrange multiplier is similar to equation (2.31)

$$\boldsymbol{\lambda}(t) = \mathbf{P}(t)\,\mathbf{x}(t) \qquad (2.33)$$

where $\mathbf{P}(t)$ is a symmetric matrix, with $n \times n$ time-dependent elements. If we substitute relation (2.33) into equations (2.26) and (2.32), we must require

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\,\mathbf{x}(t)-\mathbf{B}(t)\,\mathbf{R}^{-1}(t)\,\mathbf{B}(t)\,\mathbf{P}(t)\,\mathbf{x}(t) \qquad (2.34)$$

Also, from equations (2.33) and (2.30), we require that

$$\dot{\boldsymbol{\lambda}}(t) = \dot{\mathbf{P}}(t)\,\mathbf{x}(t)+\mathbf{P}(t)\,\dot{\mathbf{x}}(t) = -\mathbf{Q}(t)\,\mathbf{x}(t)-\mathbf{A}(t)\,\mathbf{P}(t)\,\mathbf{x}(t) \qquad (2.35)$$

Combining equations (2.34) and (2.35), we obtain

$$\begin{aligned}&\Big[\dot{\mathbf{P}}(t)+\mathbf{P}(t)\,\mathbf{A}(t)+\mathbf{A}(t)\,\mathbf{P}(t)\\ -\;&\mathbf{P}(t)\,\mathbf{B}(t)\,\mathbf{R}^{-1}(t)\,\mathbf{B}(t)\,\mathbf{P}(t)+\mathbf{Q}(t)\Big]\;\mathbf{x}(t) = \mathbf{0} \qquad (2.36)\end{aligned}$$

Since this equation must hold for all nonzero $\mathbf{x}(t)$, the term premultiplying $\mathbf{x}(t)$ must be zero. Thus, the $\mathbf{P}(t)$ matrix, which must satisfy equation (2.36), is called the Riccati matrix equation, and must be positive definite, so that

$$\dot{\mathbf{P}}(t) = -\mathbf{P}(t)\mathbf{A}(t)-\mathbf{A}(t)\mathbf{P}(t)+\mathbf{P}(t)\mathbf{B}(t)\mathbf{R}^{-1}(t)\mathbf{B}(t)\mathbf{P}(t)-\mathbf{Q}(t) \qquad (2.37)$$

with a boundary condition given by equations (2.31) and (2.33), stated as

$$\mathbf{P}(T) = \mathbf{W} \tag{2.38}$$

Thus, we may solve the Riccati matrix equation backward in time from $T$ to 0 and obtain $\mathbf{P}(t)$. The solution for $\mathbf{P}(t)$ will be illustrated in Example 2.2.1 below. Define the matrix $\mathbf{K}(t)$ as

$$\mathbf{K}(t) = -\mathbf{R}^{-1}(t)\,\mathbf{B}(t)\,\mathbf{P}(t) \tag{2.39}$$

Combining equations (2.32), (2.33) and (2.39), we obtain a closed-loop control of the form

$$\mathbf{u}(t) = \mathbf{K}(t)\,\mathbf{x}(t) \tag{2.40}$$

$\mathbf{Q}(t)$, $\mathbf{R}(t)$, and $\mathbf{W}$ must be at least positive semidefinite in order to establish the sufficient condition for a minimum. In addition, we know from equation (2.32) that $\mathbf{R}(t)$ must have an inverse; therefore, it is sufficient that $\mathbf{R}(t)$ be positive definite and the $\mathbf{Q}(t)$ and $\mathbf{W}$ be at least positive semidefinite. A detailed discussion on sufficient conditions is not given in this book. Interested readers should refer to the notes at the end of this chapter.

**Example 2.2.1.** Find the feedback optimal control law for the scalar system

$$\dot{x} = x(t) + u(t), \qquad x(0) = x_0 \tag{2.41}$$

to minimize the objective function

$$J = \frac{1}{2}\int_0^T u^2(t)dt + 2\,x^2(T) \tag{2.42}$$

The state and control variables are unconstrained. The final time $T$ is specified and $x(T)$ is free.

Raccati equation (2.37) and boundary condition (2.38) become

$$\dot{p}(t) = -2p(t) + p^2(t), \qquad p(T) = 4 \tag{2.43}$$

Solving the above differential equation, we obtain the solution

$$p(t) = \frac{2}{1 - e^{(2t-2T-\ln 2)}} \tag{2.44}$$

Thus, using equation (2.39), we have

$$K(t) = -R^{-1}(t)\;B(t)\;P(t) = -\frac{2}{1 - e^{(2t-2T-\ln 2)}} \tag{2.45}$$

The optimal feedback control law is

$$u(t) = K(t)\;x(t) \tag{2.46}$$

### 2.2.2 Fixed Beginning and Free End Times

We consider an optimal control problem with a fixed beginning $\mathbf{x}_0$ and a free end time $T$. But we have additional constraints for the final state $\mathbf{x}(T)$ and final time $T$ as follows

$$\mathrm{N}[\mathbf{x}(T), T] = \mathbf{0} \tag{2.47}$$

This kind of free end time conditions has several practical implications. Some system-optimal problems fall into this category of problems. One typical example is a system-optimal problem which has an objective function of reducing the total travel time subject to a fixed amount of traffic within an area at end time $T$. The optimal control problem is presented as follows:

$$\min_{\mathbf{x},\mathbf{u}} \quad J = \int_0^T F[\mathbf{x}(t), \mathbf{u}(t), t]\, dt + S[\mathbf{x}(T), T] \tag{2.48}$$

s.t.

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \tag{2.49}$$

$$\mathbf{x}(0) = \mathbf{x}_0, \qquad \mathrm{N}[\mathbf{x}(T), T] = \mathbf{0} \tag{2.50}$$

$$T \text{ free} \tag{2.51}$$

where $F[\mathbf{x}(t), \mathbf{u}(t), t]$ has continuous partial derivatives with respect to $\mathbf{x}(t)$ and $\mathbf{u}(t)$, and $S[\mathbf{x}(T), T]$ and $\mathrm{N}[\mathbf{x}(T), T]$ possess continuous partial derivatives with respect to $\mathbf{x}(T)$.

As before, we use the method of Lagrange multipliers to adjoin the system differential state equations to the objective function, which gives

$$\begin{aligned} I \;=\; & \int_0^T \{F[\mathbf{x}(t), \mathbf{u}(t), t] + \boldsymbol{\lambda}(t)\, [\mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] - \dot{\mathbf{x}}(t)]\}\, dt \\ + \; & S[\mathbf{x}(T), T] + \boldsymbol{\nu}(T)\, \mathrm{N}[\mathbf{x}(T), T] \end{aligned} \tag{2.52}$$

where $\boldsymbol{\lambda}(t)$ and $\boldsymbol{\nu}(T)$ are vectors of Lagrangian multipliers associated with the dynamic state equations and constraints for states at end time $T$, respectively. As before, we define a scalar function, the Hamiltonian, as

$$\mathcal{H} = F[\mathbf{x}(t), \mathbf{u}(t), T] + \boldsymbol{\lambda}(t)\, \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \tag{2.53}$$

We also define another scalar function as

$$\Theta[\mathbf{x}(T), \boldsymbol{\nu}(T), T] = S[\mathbf{x}(T), T] + \boldsymbol{\nu}(T)\, \mathrm{N}[\mathbf{x}(T), T] \tag{2.54}$$

Then, the first-order necessary conditions for the optimal control problem may be summarized as follows:

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}(t)} = 0 = \frac{\partial F[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{u}(t)} + \boldsymbol{\lambda}(t) \frac{\partial \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{u}(t)} \tag{2.55}$$

$$-\dot{\boldsymbol{\lambda}}(t) = \frac{\partial \mathcal{H}}{\partial \mathbf{x}(t)} = \frac{\partial F[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{x}(t)} + \boldsymbol{\lambda}(t) \frac{\partial \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{x}(t)} \tag{2.56}$$

$$\dot{\mathbf{x}}(t) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(t)} = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \tag{2.57}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \tag{2.58}$$

$$\boldsymbol{\lambda}(T) = \frac{\partial \Theta}{\partial \mathbf{x}(T)} = \frac{\partial \mathrm{S}[\mathbf{x}(T), T]}{\partial \mathbf{x}(T)} + \boldsymbol{\nu}(T)\frac{\partial \mathrm{N}[\mathbf{x}(T), T]}{\partial \mathbf{x}(T)} \tag{2.59}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\nu}(T)} = \mathbf{0} = \mathrm{N}[\mathbf{x}(T), T] \tag{2.60}$$

$$\mathcal{H}[\mathbf{x}(T), \mathbf{u}(T), \boldsymbol{\lambda}(T), T] + \frac{\partial \mathrm{S}[\mathbf{x}(T), T]}{\partial T} + \boldsymbol{\nu}(T)\frac{\partial \mathrm{N}[\mathbf{x}(T), T]}{\partial T} = \mathbf{0} \tag{2.61}$$

These represent $2n$ differential equations for the two-point boundary value problems. Equations (2.58)-(2.61) are termed transversality conditions. Suppose boundary condition (2.60) has $q$ equations and there are $n$ state equations. Then, equation (2.59) provides $n$ conditions with $q$ Lagrange multipliers $\boldsymbol{\nu}(T)$ to be determined. Equation (2.60) provides $q$ equations to eliminate the Lagrange multipliers $\boldsymbol{\nu}(T)$, and equation (2.61) provides one additional equation which is used to determine the unspecified end time $T$.

## 2.3 Continuous Problems with Equality and Inequality Constraints

In this section, we consider optimal control problems with nonlinear equality and inequality constraints on state and control variables. Those constraints can represent most practical constraints in realistic applications. The set of equality constraints for control and state variables are denoted as $\mathbf{G}[\mathbf{u}(t), \mathbf{x}(t), t] = \mathbf{0}$ and the set of inequality constraints are denoted as $\mathbf{K}[\mathbf{u}(t), \mathbf{x}(t), t] \geq \mathbf{0}$. We assume $\mathbf{G}$ and $\mathbf{K}$ are continuous and differentiable with respect to $\mathbf{u}$, $\mathbf{x}$ and $t$.

### 2.3.1 Fixed Beginning and Fixed End Times

We first consider an optimal control problem with a fixed beginning $\mathbf{x}_0$ and fixed end time $T$. The optimal control problem is formulated as follows

$$\min_{\mathbf{x},\mathbf{u}} \; J = \int_0^T F[\mathbf{x}(t), \mathbf{u}(t), t]dt + S[\mathbf{x}(T), T] \tag{2.62}$$

s.t.

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}[\mathbf{u}(t), \mathbf{x}(t), t] \tag{2.63}$$

$$\mathbf{G}[\mathbf{u}(t), \mathbf{x}(t), t] = \mathbf{0} \tag{2.64}$$

$$\mathbf{K}[\mathbf{u}(t), \mathbf{x}(t), t] \leq \mathbf{0} \tag{2.65}$$

$$\mathbf{x}(0) = \mathbf{x}_0, \qquad T \text{ fixed.} \tag{2.66}$$

Denote the Lagrangian multipliers associated with equations (2.63)-(2.65) as $\boldsymbol{\lambda}(t)$, $\boldsymbol{\sigma}(t)$ and $\boldsymbol{\eta}(t)$, respectively. We construct the augmented Hamiltonian $\mathcal{H}$ for the above optimal control problem as

$$\begin{aligned} \mathcal{H} &= F[\mathbf{x}(t),\mathbf{u}(t),t] + \boldsymbol{\lambda}(t)\,\mathbf{f}[\mathbf{u}(t),\mathbf{x}(t),t] \\ &+ \boldsymbol{\sigma}(t)\,\mathbf{G}[\mathbf{u}(t),\mathbf{x}(t),t] + \boldsymbol{\eta}(t)\,\mathbf{K}[\mathbf{u}(t),\mathbf{x}(t),t] \end{aligned} \tag{2.67}$$

where

$$\boldsymbol{\eta}(t)\begin{cases} >\mathbf{0} & \text{if } \mathbf{K}=\mathbf{0} \\ =\mathbf{0} & \text{if } \mathbf{K}<\mathbf{0} \end{cases}$$

The first order necessary conditions for the optimal control program are

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial \mathbf{u}(t)} = 0 &= \frac{\partial F[\mathbf{x}(t),\mathbf{u}(t),t]}{\partial \mathbf{u}(t)} + \boldsymbol{\lambda}(t)\frac{\partial \mathbf{f}[\mathbf{u}(t),\mathbf{x}(t),t]}{\partial \mathbf{u}(t)} \\ &+ \boldsymbol{\sigma}(t)\frac{\partial \mathbf{G}[\mathbf{u}(t),\mathbf{x}(t),t]}{\partial \mathbf{u}(t)} + \boldsymbol{\eta}(t)\frac{\partial \mathbf{K}[\mathbf{u}(t),\mathbf{x}(t),t]}{\partial \mathbf{u}(t)} \end{aligned} \tag{2.68}$$

$$\begin{aligned} -\dot{\boldsymbol{\lambda}}(t) = \frac{\partial \mathcal{H}}{\partial \mathbf{x}(t)} &= \frac{\partial F[\mathbf{x}(t),\mathbf{u}(t),t]}{\partial \mathbf{x}(t)} + \boldsymbol{\lambda}(t)\frac{\partial \mathbf{f}[\mathbf{x}(t),\mathbf{u}(t),t]}{\partial \mathbf{x}(t)} \\ &+ \boldsymbol{\sigma}(t)\frac{\partial \mathbf{G}[\mathbf{u}(t),\mathbf{x}(t),t]}{\partial \mathbf{x}(t)} + \boldsymbol{\eta}(t)\frac{\partial \mathbf{K}[\mathbf{u}(t),\mathbf{x}(t),t]}{\partial \mathbf{x}(t)} \end{aligned} \tag{2.69}$$

$$\dot{\mathbf{x}}(t) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(t)} = \mathbf{f}[\mathbf{x}(t),\mathbf{u}(t),t] \tag{2.70}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\sigma}(t)} = \mathbf{0} = \mathbf{G}[\mathbf{u}(t),\mathbf{x}(t),t] \tag{2.71}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\eta}(t)} = \mathbf{K}[\mathbf{u}(t),\mathbf{x}(t),t] \geq \mathbf{0} \tag{2.72}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \tag{2.73}$$

$$\boldsymbol{\lambda}(T) = \frac{\partial S[\mathbf{x}(T),T]}{\partial \mathbf{x}(T)} \tag{2.74}$$

Equations (2.68)-(2.69) are similar to equations (2.17) and (2.18) except the additional terms resulting from equality and inequality constraints on the control and state variables. Equation (2.70) is a restatement of the state equation (2.63). Equations (2.71)-(2.72) are restatements of the equality and inequality constraints. Equations (2.73)-(2.74) are two-point boundary conditions.

### 2.3.2 Fixed Beginning and Free End Times

We now consider an optimal control problem for a time period $[0, T]$ where the end time $T$ is free. This optimal control problem is presented as follows:

$$\min_{\mathbf{x},\mathbf{u}} \quad J = \int_0^T F[\mathbf{x}(t), \mathbf{u}(t), t]\, dt + S[\mathbf{x}(T), T] \tag{2.75}$$

s.t.

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \tag{2.76}$$

$$\mathbf{G}[\mathbf{u}(t), \mathbf{x}(t), t] = \mathbf{0} \tag{2.77}$$

$$\mathbf{K}[\mathbf{u}(t), \mathbf{x}(t), t] \leq \mathbf{0} \tag{2.78}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \tag{2.79}$$

$$\mathbf{N}[\mathbf{x}(T), T] = \mathbf{0} \tag{2.80}$$

$$T \ \text{ free} \tag{2.81}$$

where $\mathbf{N}[\mathbf{x}(T), T]$ possess continuous partial derivatives with respect to $\mathbf{x}(T)$. We construct the augmented Hamiltonian $\mathcal{H}$ for the above optimal control problem as

$$\begin{aligned} \mathcal{H} &= F[\mathbf{x}(t), \mathbf{u}(t), t] + \boldsymbol{\lambda}(t)\, \mathbf{f}[\mathbf{u}(t), \mathbf{x}(t), t] \\ &+ \boldsymbol{\sigma}(t)\, \mathbf{G}[\mathbf{u}(t), \mathbf{x}(t), t] + \boldsymbol{\eta}(t)\, \mathbf{K}[\mathbf{u}(t), \mathbf{x}(t), t] \end{aligned} \tag{2.82}$$

where

$$\boldsymbol{\eta}(t) \begin{cases} > \mathbf{0} & \text{if } \mathbf{K} = \mathbf{0} \\ = \mathbf{0} & \text{if } \mathbf{K} < \mathbf{0} \end{cases}$$

and $\boldsymbol{\lambda}(t)$, $\boldsymbol{\sigma}(t)$ and $\boldsymbol{\eta}(t)$ are vectors of Lagrangian multipliers associated with the dynamic state equations and constraints for state and control variables. We define another scalar function as

$$\Theta[\mathbf{x}(T), \boldsymbol{\nu}(T), T] = S[\mathbf{x}(T), T] + \boldsymbol{\nu}(T)\, \mathbf{N}[\mathbf{x}(T), T] \tag{2.83}$$

where $\boldsymbol{\nu}(T)$ is a vector of Lagrange multipliers associated with constraints for states at end time $T$. Then, the first-order necessary conditions for the optimal control problem may be summarized as follows:

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial \mathbf{u}(t)} = 0 &= \frac{\partial F[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{u}(t)} + \boldsymbol{\lambda}(t) \frac{\partial \mathbf{f}[\mathbf{u}(t), \mathbf{x}(t), t]}{\partial \mathbf{u}(t)} \\ &+ \boldsymbol{\sigma}(t) \frac{\partial \mathbf{G}[\mathbf{u}(t), \mathbf{x}(t), t]}{\partial \mathbf{u}(t)} + \boldsymbol{\eta}(t) \frac{\partial \mathbf{K}[\mathbf{u}(t), \mathbf{x}(t), t]}{\partial \mathbf{u}(t)} \end{aligned} \tag{2.84}$$

$$\begin{aligned} -\dot{\boldsymbol{\lambda}}(t) = \frac{\partial \mathcal{H}}{\partial \mathbf{x}(t)} &= \frac{\partial F[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{x}(t)} + \boldsymbol{\lambda}(t) \frac{\partial \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{x}(t)} \\ &+ \boldsymbol{\sigma}(t) \frac{\partial \mathbf{G}[\mathbf{u}(t), \mathbf{x}(t), t]}{\partial \mathbf{x}(t)} + \boldsymbol{\eta}(t) \frac{\partial \mathbf{K}[\mathbf{u}(t), \mathbf{x}(t), t]}{\partial \mathbf{x}(t)} \end{aligned} \tag{2.85}$$

$$\dot{\mathbf{x}}(t) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(t)} = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \tag{2.86}$$

$$\frac{\partial \mathcal{H}}{\partial \sigma(t)} = \mathbf{0} = \mathbf{G}[\mathbf{u}(t), \mathbf{x}(t), t] \tag{2.87}$$

$$\frac{\partial \mathcal{H}}{\partial \eta(t)} = \mathbf{K}[\mathbf{u}(t), \mathbf{x}(t), t] \leq \mathbf{0} \tag{2.88}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \tag{2.89}$$

$$\boldsymbol{\lambda}(T) = \frac{\partial \Theta}{\partial \mathbf{x}(T)} = \frac{\partial \mathbf{S}[\mathbf{x}(T), T]}{\partial \mathbf{x}(T)} + \boldsymbol{\nu}(T)\frac{\partial \mathbf{N}[\mathbf{x}(T), T]}{\partial \mathbf{x}(T)} \tag{2.90}$$

$$\mathbf{N}[\mathbf{x}(T), T] = \mathbf{0} \tag{2.91}$$

$$\mathcal{H}[\mathbf{x}(T), \mathbf{u}(T), \boldsymbol{\lambda}(T), T] + \frac{\partial \mathbf{S}[\mathbf{x}(T), T]}{\partial T} + \boldsymbol{\nu}(T)\frac{\partial \mathbf{N}[\mathbf{x}(T), T]}{\partial T} = 0 \tag{2.92}$$

Equations (2.84)-(2.85) are $2n$ differential equations for the two-point boundary value problems. Equations (2.86)-(2.88) are simply the restatement of state equation (2.76) and constraints (2.77)-(2.78). Equations (2.89)-(2.92) are boundary or transversality conditions. Equation (2.90) provides $n$ conditions with $q$ Lagrange multipliers to be determined. Equation (2.91) provides $q$ equations to eliminate the Lagrange multipliers, and equation (2.92) provides one additional equation which is used to determine the unspecified end time $T$.

## Bang-Bang Control and Minimum Time Problem

We now discuss a special case of the free end time problem — the Bang-Bang control and minimum time problem. This problem has potential applications in evacuation purposes, which are important for managing traffic congestion in emergencies.

In a variety of applications, maximum effort control problems have become increasingly important. It is natural to ask under what circumstances optimal controls will always be at maximum effort, or ***Bang-Bang***. To do this, we restrict each component of the control vector, $u(t)$, to some bounded interval as follows:

$$a_i \leq u_i(t) \leq b_i, \qquad \forall\ i \tag{2.93}$$

where $a_i$ and $b_i$ are lower and upper bounds, respectively. We also consider a nonlinear differential system where the control enters in a linear fashion as follows:

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), t] + \mathbf{g}[\mathbf{x}(t), t]\,\mathbf{u}(t), \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.94}$$

We consider an objective function which contains only linear terms in control variables. The objective function is as follows

$$J = \int_0^T \{F[\mathbf{x}(t), t] + \mathbf{h}[\mathbf{x}(t), t]\,\mathbf{u}(t)\}\, dt + S[\mathbf{x}(T), T] \tag{2.95}$$

Thus, the Hamiltonian will also be linear in $\mathbf{u}(t)$. It follows that

$$\begin{aligned}\mathcal{H}[\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\lambda}(t), t] &= F[\mathbf{x}(t), t] + \mathbf{h}[\mathbf{x}(t), t]\,\mathbf{u}(t) \\ &+ \boldsymbol{\lambda}(t)\,\{\mathbf{f}[\mathbf{x}(t), t] + \mathbf{g}[\mathbf{x}(t), t]\,\mathbf{u}(t)\} \\ &= F[\mathbf{x}(t), t] + \boldsymbol{\lambda}(t)\mathbf{f}[\mathbf{x}(t), t] \\ &+ \{\mathbf{h}[\mathbf{x}(t), t] + \boldsymbol{\lambda}(t)\mathbf{g}[\mathbf{x}(t), t]\}\,\mathbf{u}(t) \end{aligned} \tag{2.96}$$

Since the Hamiltonian is linear in the control vector, $\mathbf{u}(t)$, minimization of the Hamiltonian with respect to $\mathbf{u}(t)$ requires that

$$u_i(t) = \begin{cases} a_i & \text{if } \{\, \mathbf{h}[\mathbf{x}(t), t] + \boldsymbol{\lambda}(t)\mathbf{g}[\mathbf{x}(t), t]\}_i > 0 \\ b_i & \text{if } \{\, \mathbf{h}[\mathbf{x}(t), t] + \boldsymbol{\lambda}(t)\mathbf{g}[\mathbf{x}(t), t]\}_i < 0 \end{cases}$$

Thus, when the control vector appears linearly in both the state equation and the objective function, and each component of the control vector is bounded, the optimal control is Bang-Bang. We call the above criterion the Bang-Bang control rule. The only exception to this occurs in the case where

$$\mathbf{h}[\mathbf{x}(t), t] + \boldsymbol{\lambda}(t)\,\mathbf{g}[\mathbf{x}(t), t] = \mathbf{0} \tag{2.97}$$

Then the Hamiltonian is not a function of $\mathbf{u}(t)$ and can not be minimized with respect to $\mathbf{u}(t)$. When equation (2.97) holds for more than isolated points in time, the optimization problem is said to possess a singular solution, a problem which we will not discuss in detail. A singular solution is possible with respect to a particular control component, $\mathbf{u}(t)$, if the $i$th component of equation (2.97) is 0. For this problem, the first-order necessary conditions may be summarized as

$$\dot{\mathbf{x}}(\mathbf{t}) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(t)} = \mathbf{f}[\mathbf{x}(t), t] + \mathbf{g}[\mathbf{x}(t), t]\,\mathbf{u}(t) \tag{2.98}$$

$$\begin{aligned}-\dot{\boldsymbol{\lambda}}(t) = \frac{\partial \mathcal{H}}{\partial \mathbf{x}(t)} &= \frac{\partial F[\mathbf{x}(t), t]}{\partial \mathbf{x}(t)} + \frac{\partial \mathbf{h}[\mathbf{x}(t), t]}{\partial \mathbf{x}(t)}\mathbf{u}(t) \\ &+ \boldsymbol{\lambda}(t)\frac{\partial \mathbf{f}[\mathbf{x}(t), t]}{\partial \mathbf{x}(t)} + \boldsymbol{\lambda}(t)\frac{\partial \mathbf{g}[\mathbf{x}(t), t]}{\partial \mathbf{x}(t)}\mathbf{u}(t)\end{aligned} \tag{2.99}$$

where $\mathbf{u}(t)$ is determined using the Bang-Bang rule. Since we have not specifically stated the end conditions, we discuss the general problem. When we specify information concerning the desired states at the end time and the initial condition vector, we have a two-point boundary value problem with half of the conditions specified at the initial time and half at the end time.

A possible method for solving the first-order necessary conditions for this formulation consists of reversing time in these equations. Starting at the determined or specified terminal vector $\mathbf{x}(T)$, we integrate back from this point with a constant control (Bang-Bang) until the switching point is obtained using the Bang-Bang control rule.

We now illustrate various solutions to a particular case which results in bang-bang control—the minimum time problem for constant linear systems with a scalar input. In this problem, we desire to transfer an $n$ vector constant differential system

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{2.100}$$

to its origin, $\mathbf{x}(T) = \mathbf{0}$, in minimal time. Thus, we have the objective function

$$J = \int_0^T dt = T \tag{2.101}$$

with the restriction that

$$-1 \le \mathbf{u}(t) \le +1 \tag{2.102}$$

The Hamiltonian for our problem is

$$\mathcal{H}[\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\lambda}(t)] = 1 + \boldsymbol{\lambda}(t)\ \mathbf{A}\ \mathbf{x}(t) + \boldsymbol{\lambda}(t)\ \mathbf{B}\ \mathbf{u}(t) \tag{2.103}$$

We must minimize the Hamiltonian with respect to a choice of $\mathbf{u}(t)$. Thus, we require

$$\mathbf{u}^*(t) = -\text{sign}\ [\boldsymbol{\lambda}(t)\ \mathbf{B}] \tag{2.104}$$

where sign $[\boldsymbol{\lambda}(t)\ \mathbf{B}] = 1$ when $\boldsymbol{\lambda}(t)\ \mathbf{B} > 0$ and sign $[\boldsymbol{\lambda}(t)\ \mathbf{B}] = -1$ when $\boldsymbol{\lambda}(t)\ \mathbf{B} < 0$. Thus, the Hamiltonian with the optimal control $\mathbf{u}^*(t)$ is

$$\mathcal{H}[\mathbf{x}(t), \boldsymbol{\lambda}(t)] = 1 + \boldsymbol{\lambda}(t)\ \mathbf{A}\ \mathbf{x}(t) - \boldsymbol{\lambda}(t)\ \mathbf{B}\ \text{sign}\ [\boldsymbol{\lambda}(t)\ \mathbf{B}] \tag{2.105}$$

Since $\mathcal{H}$ does not depend explicitly on $t$, $d\mathcal{H}/dt = 0$. Furthermore, since the end time $T$ is free, we know from equation (2.92) that $\mathcal{H}[\mathbf{x}(T), \mathbf{u}(T), \boldsymbol{\lambda}(T)] = 0$. Thus, we have

$$\mathcal{H}[\mathbf{x}(t), \boldsymbol{\lambda}(t)] = 0 \qquad \forall\, t \in [0, T] \tag{2.106}$$

On the optimal trajectory, the state equations are

$$\dot{\mathbf{x}}(t) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(t)} = \mathbf{A}\ \mathbf{x}(t) + \mathbf{B}\ \mathbf{u}(t) = \mathbf{A}\ \mathbf{x}(t) - \mathbf{B}\ \text{sign}\ [\boldsymbol{\lambda}(t)\mathbf{B}] \tag{2.107}$$

$$\dot{\boldsymbol{\lambda}}(t) = -\frac{\partial \mathcal{H}}{\partial \mathbf{x}(t)} = -\mathbf{A}\boldsymbol{\lambda}(t) \tag{2.108}$$

To avoid a singular solution, we must ensure that $\boldsymbol{\lambda}(t)\ \mathbf{B}$ can not be zero over a time interval of nonzero length. From equation (2.108), we see that this is almost certainly the case unless $\boldsymbol{\lambda}(0)$ were identically $\mathbf{0}$, which is not possible. The solution to equation (2.108) is

$$\boldsymbol{\lambda}(t) = e^{-\mathbf{A}(t-T)}\boldsymbol{\lambda}(T) \tag{2.109}$$

It is convenient for us to rewrite equation (2.107) using a new time variable

$$\tau = T - t \tag{2.110}$$

and a new state variable

$$\xi(\tau) = x(t) = x(T-\tau) \tag{2.111}$$

Substituting equations (2.109)-(2.111) into equation (2.107), we obtain

$$\frac{d\xi}{d\tau} = \frac{d\xi}{dt} \cdot \frac{dt}{d\tau} = -\frac{d\xi}{dt} = \frac{dx}{dt} = -\mathbf{A}\ \xi(\tau) + \mathbf{B}\ \text{sign}[\boldsymbol{\lambda}(T)e^{\mathbf{A}\tau}\mathbf{B}] \tag{2.112}$$

Since $\xi(0) = x(T) = 0$, the above equation has its solution

$$\xi(\tau) = \int_0^{\tau} e^{-\mathbf{A}(\tau-\omega)}\mathbf{B}\ \text{sign}\ [\boldsymbol{\lambda}(T)e^{\mathbf{A}\omega}\mathbf{B}]\ d\omega \tag{2.113}$$

A state $\mathbf{x}(0) \doteq \mathbf{x}_0$ from which the origin can be reached in a specified minimal time $T$ may now be obtained if we substitute a value of $\boldsymbol{\lambda}(T)$ in equation (2.113) and than calculate $\mathbf{x}(0)$ and $\boldsymbol{\lambda}(T)$. Thus, $\boldsymbol{\lambda}(t)$ is working as a sort of *influence function.*

Since it is the direction and not the magnitude of the $\boldsymbol{\lambda}(T)$ vector which determines sign $[\boldsymbol{\lambda}(T)e^{\mathbf{A}\omega}\mathbf{B}]$, all states which can be reached in a given minimal time may be determined if we allow $\boldsymbol{\lambda}(T)$ to assume all values over a unit sphere. At points where $\boldsymbol{\lambda}(T)e^{\mathbf{A}\omega}\mathbf{B}$ is 0, we have a switching point. It is possible to show that if the eigenvalues of $\mathbf{A}$ are real, there are, at most, $n-1$ switchings or changes of sign of the control. We will now give examples of calculations of Bang-Bang controls.

**Example 2.3.1.** Use the Minimum Principle to discuss possible optimal control laws and optimal trajectories for the system

$$\dot{x}_1(t) = a_{11}x_1(t) + a_{12}x_2(t) + b_1u_1(t) \tag{2.114}$$

$$\dot{x}_2(t) = a_{21}x_1(t) + a_{22}x_2(t) + b_2u_2(t) \tag{2.115}$$

where the objective function is to minimize

$$J = \int_0^T [C_0 + C_1u_1(t) + C_2u_2(t)]\ dt \tag{2.116}$$

Coefficients $a_{ij}$ $(i = 1,2,\ j = 1,2)$, $b_i$ $(i = 1,2)$, $C_k$ $(k = 0,1,2)$ are constants. There is a constraint on control variables

$$0 \le u_i(t) \le 1 \qquad \forall i = 1,2 \tag{2.117}$$

The state variables are unconstrained. The final time $T$ is specified and $x(T)$ is free.

$$x_1(0) = x_{10} \qquad x_2(0) = x_{20} \qquad \text{given} \tag{2.118}$$

The Hamiltonian is formulated as

$$\begin{aligned}\mathcal{H} &= C_0 + C_1u_1(t) + C_2u_2(t) + \lambda_1[a_{11}x_1(t) + a_{12}x_2(t) + b_1u_1(t)] \\ &+ \lambda_2[a_{21}x_1(t) + a_{22}x_2(t) + b_2u_2(t)]\end{aligned}$$

where $\boldsymbol{\lambda}(t)$ are Lagrange multipliers associated with the state equations. Reorganizing the Hamiltonian based on control variables, we have

$$\begin{aligned}\mathcal{H} &= C_0 + \lambda_1[a_{11}x_1 + a_{12}x_2] + \lambda_2[a_{21}x_1 + a_{22}x_2] \\ &+ [C_1 + \lambda_1b_1]u_1 + [C_2 + \lambda_2b_2]u_2\end{aligned}$$

Using the Minimum Principle

$$\mathcal{H}[u^*(t)] \leq \mathcal{H}[u(t)], \tag{2.119}$$

we obtain the optimal control as follows:

$$u_1^*(t) = \begin{cases} 0 & \text{if } C_1 + \lambda_1b_1 \geq 0 \\ 1 & \text{if } C_1 + \lambda_1b_1 < 0 \end{cases}$$

$$u_2^*(t) = \begin{cases} 0 & \text{if } C_2 + \lambda_2b_2 \geq 0 \\ 1 & \text{if } C_2 + \lambda_2b_2 < 0 \end{cases}$$

Substituting the optimal control into the state equation, we obtain the optimal trajectory. There are four possible combinations of optimal controls. We discuss only one here. When

$$u_1^*(t) = 0 \qquad u_2^*(t) = 0,$$

we have state equations as follows

$$\dot{x}_1(t) = a_{11}x_1(t) + a_{12}x_2(t) \tag{2.120}$$

$$\dot{x}_2(t) = a_{21}x_1(t) + a_{22}x_2(t) \tag{2.121}$$

Solving the above two equations, we obtain the optimal trajectories

$$x_1^*(t) = B_1e^{\alpha_1t} + B_2e^{\alpha_2t} \tag{2.122}$$

$$x_2^*(t) = \frac{1}{a_{12}}\left[(\alpha_1 - a_{11})B_1e^{\alpha_1t} + (\alpha_2 - a_{11})B_2e^{\alpha_2t}\right] \tag{2.123}$$

where $B_1$, $B_2$ are integral constants and $\alpha_1$, $\alpha_2$ have values as follows

$$\alpha_1 = \frac{(a_{11} + a_{22}) + \sqrt{(a_{11} - a_{22})^2 + 4a_{12}a_{21}}}{2} \tag{2.124}$$

$$\alpha_2 = \frac{(a_{11} + a_{22}) - \sqrt{(a_{11} - a_{22})^2 + 4a_{12}a_{21}}}{2} \tag{2.125}$$

Substituting boundary conditions $x_1(0) = 1$ and $x_2(0) = 0$ into equations (2.122)-(2.123), we obtain the coefficients

$$B_1 = -\frac{\alpha_2 - a_{11}}{\alpha_1 - \alpha_2} \tag{2.126}$$

$$B_2 = \frac{\alpha_1 - a_{11}}{\alpha_1 - \alpha_2} \tag{2.127}$$

**Example 2.3.2.** Consider a traffic signal control problem at an intersection. Figure 2.2 shows a signal controlled four-leg intersection with one-way streets. To simplify our problem, we assume no turning movements are allowed at this intersection and there are queues at both approaches. Denote arriving flow rates as $q_1(t)$ and $q_2(t)$ in vehicles per hour, and saturation flows as $s_1$ and $s_2$ in vehicles per hour. Define green times for both approaches as $g_1(t)$ and $g_2(t)$ in seconds, respectively, It follows that

$$g_1(t) + g_2(t) = C - L \tag{2.128}$$

where $C$ is the cycle length (in seconds) and $L$ is the lost time (in seconds) due to acceleration and deceleration at the intersection. We assume both $C$ and $L$ are fixed.

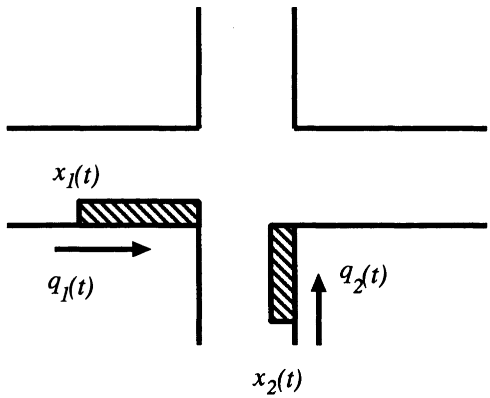



Figure 2.2: Four-Leg Intersection

Denote $u(t)$ as the average number of vehicles per hour which can pass the intersection for approach 1. It follows that

$$u(t) = \frac{s_1\, g_1(t)}{C} \tag{2.129}$$

For approach 2, the average number of vehicles per hour which can pass the intersection is $s_2\, g_2(t)/C$. Thus, we can obtain the relationship between the flows passing the intersection for both approaches as follows

$$\frac{s_2\, g_2(t)}{C} = -\frac{s_2}{s_1}u(t) + s_2(1 - \frac{L}{C}) \tag{2.130}$$

Denote the numbers of queuing vehicles at approaches 1 and 2 as $x_1(t)$ and $x_2(t)$, respectively. Thus, we can write the state equations as follows

$$\dot{x}_1(t) = q_1(t) - u(t) \tag{2.131}$$

$$\dot{x}_2(t) = q_2(t) - s_2(1 - \frac{L}{C}) + \frac{s_2}{s_1}u(t) \tag{2.132}$$

where $u(t)$ is subject to

$$u_{\min} \le u(t) \le u_{\max} \tag{2.133}$$

By substituting the minimum and maximum green times for approach 1 into equation (2.129), we can determine $u_{\min}$ and $u_{\max}$. We also have nonnegativity conditions for state variables as follows

$$x_1(t) \ge 0, \qquad x_2(t) \ge 0 \tag{2.134}$$

In this example, the objective function is to minimize the total number of queuing vehicles at the intersection from time 0 to time $T$ so that $x_1(T) = x_2(T) = 0$ where the end time $T$ is free. In other words, we seek to minimize the cumulative net queue and finally clear queues at both approaches at time $T$. It follows that

$$\text{Min} \qquad J = \int_0^T [x_1(t) + x_2(t)]\, dt \tag{2.135}$$

We assume that $x_1(0)$ and $x_2(0)$ are given in this example. This problem has two fixed boundary points at times 0 and $T$.

The Hamiltonian is constructed as

$$\begin{aligned} \mathcal{H} &= x_1(t) + x_2(t) + \lambda_1(t)[q_1(t) - u(t)] \\ &+ \lambda_2(t)[q_2(t) - s_2(1 - \frac{L}{C}) + \frac{s_2}{s_1}u(t)] \end{aligned} \tag{2.136}$$

The first-order necessary conditions require that

$$\dot{\lambda}_1(t) = -\partial\mathcal{H}/\partial x_1 = -1, \qquad \dot{\lambda}_2(t) = -\partial\mathcal{H}/\partial x_2 = -1 \tag{2.137}$$

Denoting the integral constants as $C_1$ and $C_2$, we have

$$\lambda_1(t) = -t + C_1, \qquad \lambda_2(t) = -t + C_2 \tag{2.138}$$

Define an auxiliary variable $z(t)$ as

$$z(t) = -\lambda_1(t) + \frac{s_2}{s_1}\lambda_2(t) = \frac{s_1 - s_2}{s_1}t + \frac{s_2C_2 - s_1C_1}{s_1} \qquad (2.139)$$

By placing terms associated with control variable $u(t)$ together, the Hamiltonian is rewritten as

$$\begin{aligned} \mathcal{H} &= x_1(t) + x_2(t) + \lambda_1\, q_1(t) \\ &+ \lambda_2[q_2(t) - s_2(1 - \frac{L}{C})] + z(t)\, u(t) \end{aligned} \qquad (2.140)$$

In order to minimize the Hamiltonian, we require that

$$\begin{aligned} z(t) > 0 &\implies u(t) = u_{\mathbf{min}} \\ z(t) < 0 &\implies u(t) = u_{\mathbf{max}} \end{aligned} \qquad (2.141)$$

We assume that approach 1 has higher capacity than approach 2, i.e., $s_1 > s_2$. We note that equation (2.139) is a first-order equation of $t$ and $t$ has a positive coefficient $(s_1 - s_2)/s_1$. If $z(t)$ starts from a negative value, i.e., $z(t) < 0$ for $t < t_c$ (a critical time instant), then $u(t) = u_{\mathbf{max}}$ for $t < t_c$ and $u(t) = u_{\mathbf{min}}$ for $t > t_c$. Otherwise, when $z(t)$ starts from a positive value, i.e., $z(t) > 0$ for $t \geq 0$, $u(t) = u_{\mathbf{min}}$ holds for $t \geq 0$. Thus, from equations (2.139) and (2.141), we conclude that:

1. $u(t)$ can only have values of $u_{\mathbf{max}}$ or $u_{\mathbf{min}}$; it does not take any intermediate value between $u_{\mathbf{max}}$ and $u_{\mathbf{min}}$;

2. the control changes at most only one time;

3. when $u(t)$ changes, the initial control is $u(t) = u_{\mathbf{max}}$ (the higher capacity approach has the priority.)

Denote $Q_i(t)$ as the cumulative number of arrivals at time $t$ along approach $i$ and $G_i(t)$ as the cumulative number of departures at time $t$ along approach $i$. It follows that

$$Q_i(t) = \int_0^t q_i(\tau)\, d\tau \qquad \forall\, i = 1, 2 \qquad (2.142)$$

$$G_1(t) = \int_0^t u(\tau) d\tau \qquad (2.143)$$

$$G_2(t) = \int_0^t s_2 \left[1 - \frac{L}{C} - \frac{1}{s_1}u(\tau)\right] d\tau \qquad (2.144)$$

Thus, the state equations can be reformulated as

$$x_i(t) = x_i(0) + Q_i(t) - G_i(t) \qquad \forall i = 1, 2 \qquad (2.145)$$

In other words, $Q_i(t)$ are the cumulative arrivals and $G_i(t)$ are the cumulative departures. As shown in Figure 2.3, when the cumulative arrivals are given, the optimal cumulative departures are given by the straight lines. The slopes of the lines for approach 1 are $u_{\max}$ and $u_{\min}$. The slopes of the lines for approach 2 are given as

$$\gamma_{\min} = s_2(1 - \frac{L}{C} - \frac{1}{s_1}u_{\max}), \qquad \gamma_{\max} = s_2(1 - \frac{L}{C} - \frac{1}{s_1}u_{\min}) \qquad (2.146)$$

The time $t_s$ and $T$ are determined by the cumulative arrivals. The two curves merge at time $T$.

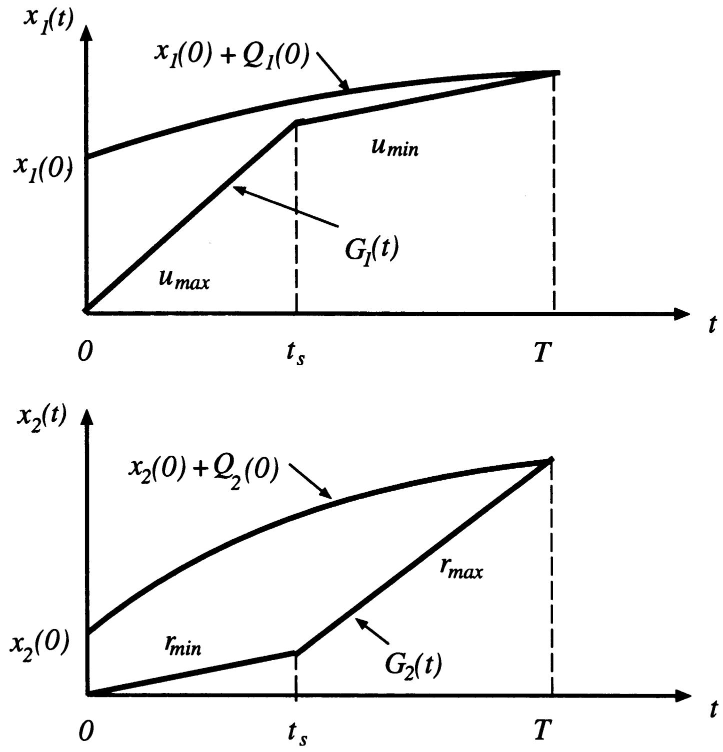



Figure 2.3: Cumulative Arrivals and Departures at the Intersection

## 2.4 Continuous Problems with Equality and Nonnegativity Constraints

We now consider a set of special optimal control problems which will be widely used in the following chapters for problem formulation and analysis for dynamic transportation networks. The state equations are linear

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}\ \mathbf{u}(t) \tag{2.147}$$

where $\mathbf{A}$ is a matrix of constants. We consider only linear equality constraints for control variables

$$\mathbf{G}\ \mathbf{u}(t) = \mathbf{0} \tag{2.148}$$

and nonlinear equality constraints involving only state variables

$$\mathbf{K}[\mathbf{x}(t)] = \mathbf{0} \tag{2.149}$$

where $\mathbf{G}$ is a matrix of constants.

### 2.4.1 Fixed Beginning and Fixed End Times

We now consider an optimal control problem with a fixed end time $T$. The optimal control problem is formulated as follows

$$\min_{\mathbf{x},\mathbf{u}} \quad J = \int_0^T F[\mathbf{x}(t), \mathbf{u}(t)]dt + S[\mathbf{x}(T)] \tag{2.150}$$

| s.t. | | Lagrange Multiplier | |
|---|---|---|---|
| Linear State Equations | $\dot{\mathbf{x}}(t) = \mathbf{A}\ \mathbf{u}(t)$ | $\boldsymbol{\lambda}(t)$ | (2.151) |
| Linear Control Variable Constraints | $\mathbf{G}\ \mathbf{u}(t) = \mathbf{0}$ | $\boldsymbol{\sigma}(t)$ | (2.152) |
| State Variable Constraints | $\mathbf{K}[\mathbf{x}(t)] = \mathbf{0}$ | $\boldsymbol{\eta}(t)$ | (2.153) |

$$\mathbf{x}(t) \geq \mathbf{0}, \qquad \mathbf{u}(t) \geq \mathbf{0}, \tag{2.154}$$

$$\mathbf{x}(0) \quad \text{given.} \tag{2.155}$$

We construct the augmented Hamiltonian $\mathcal{H}$ for the above optimal control problem as

$$\mathcal{H} = F[\mathbf{x}(t), \mathbf{u}(t)] + \boldsymbol{\lambda}(t)\ \mathbf{A}\ \mathbf{u}(t) + \boldsymbol{\sigma}(t)\ \mathbf{G}\ \mathbf{u}(t) + \boldsymbol{\eta}(t)\ \mathbf{K}\ [\mathbf{x}(t)] \tag{2.156}$$

The first order necessary conditions for the optimal control program are

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}(t)} = \frac{\partial F[\mathbf{x}(t), \mathbf{u}(t)]}{\partial \mathbf{u}(t)} + \boldsymbol{\lambda}(t)\ \mathbf{A} + \boldsymbol{\sigma}(t)\ \mathbf{G} \geq \mathbf{0}, \tag{2.157}$$

$$\text{and} \qquad \mathbf{u}(t)\frac{\partial \mathcal{H}}{\partial \mathbf{u}(t)} = 0, \tag{2.158}$$

$$-\dot{\boldsymbol{\lambda}}(t) = \frac{\partial \mathcal{H}}{\partial \mathbf{x}(t)} = \frac{\partial F[\mathbf{x}(t), \mathbf{u}(t)]}{\partial \mathbf{x}(t)} + \boldsymbol{\eta}(t)\frac{\partial \mathbf{K}[\mathbf{x}(t)]}{\partial \mathbf{x}(t)} \tag{2.159}$$

$$\dot{\mathbf{x}}(t) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(t)} = \mathbf{A}\ \mathbf{u}(t) \tag{2.160}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\sigma}(t)} = \mathbf{0} = \mathbf{G}\ \mathbf{u}(t) \tag{2.161}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\eta}(t)} = \mathbf{K}[\mathbf{x}(t)] = \mathbf{0} \tag{2.162}$$

$$\boldsymbol{\lambda}(T) = \frac{\partial \mathcal{S}[\mathbf{x}(T)]}{\partial \mathbf{x}(T)} \tag{2.163}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \tag{2.164}$$

$$\mathbf{x}(t) \geq \mathbf{0}, \qquad \mathbf{u}(t) \geq \mathbf{0}. \tag{2.165}$$

Equations (2.157)-(2.159) are similar to equations (2.17) and (2.18) except for additional terms resulting from equality constraints on control and state variables. The inequality in equation (2.157) is caused by the nonnegativity constraint on the control variable. Equation (2.160) is a restatement of state equation (2.151). Equations (2.161)-(2.162) are restatements of the equality constraints. Equations (2.163)-(2.165) are two-point boundary conditions.

### 2.4.2 Fixed Beginning and Free End Times

We now consider an optimal control problem for a time period $[0, T]$ where end time $T$ is free. This optimal control problem is given as follows:

$$\min_{\mathbf{x},\mathbf{u}} \quad J = \int_0^T F[\mathbf{x}(t), \mathbf{u}(t), t]\, dt + S[\mathbf{x}(T), T] \tag{2.166}$$

| s.t. | | Lagrange Multiplier | |
|---|---|---|---|
| Linear State Equations | $\dot{\mathbf{x}}(t) = \mathbf{A}\ \mathbf{u}(t)$ | $\boldsymbol{\lambda}(t)$ | (2.167) |
| Linear Control Variable Constraints | $\mathbf{G}\ \mathbf{u}(t) = \mathbf{0}$ | $\boldsymbol{\sigma}(t)$ | (2.168) |
| State Variable Constraints | $\mathbf{K}[\mathbf{x}(t)] = \mathbf{0}$ | $\boldsymbol{\eta}(t)$ | (2.169) |

$$\mathbf{x}(t) \geq \mathbf{0}, \qquad \mathbf{u}(t) \geq \mathbf{0}, \tag{2.170}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \qquad \mathbf{N}[\mathbf{x}(T), T] = \mathbf{0} \tag{2.171}$$

$$T \text{ free} \tag{2.172}$$

where $\mathbf{N}[\mathbf{x}(T), T]$ possess continuous partial derivatives with respect to $\mathbf{x}(T)$. We construct the augmented Hamiltonian $\mathcal{H}$ for the above optimal control problem as

$$\mathcal{H} = F[\mathbf{x}(t), \mathbf{u}(t), t] + \boldsymbol{\lambda}(t)\ \mathbf{A}\ \mathbf{u}(t) + \boldsymbol{\sigma}(t)\ \mathbf{G}\ \mathbf{u}(t) + \boldsymbol{\eta}(t)\ \mathbf{K}\ [\mathbf{x}(t)] \tag{2.173}$$

We define another scalar function as

$$\Theta[\mathbf{x}(T), \boldsymbol{\nu}(T), T] = S[\mathbf{x}(T), T] + \boldsymbol{\nu}(T)\ \mathrm{N}[\mathbf{x}(T), T] \tag{2.174}$$

where $\boldsymbol{\nu}(T)$ is a vector of Lagrange multipliers. The first order necessary conditions for the optimal control program are

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}(t)} = \frac{\partial F[\mathbf{x}(t), \mathbf{u}(t), t]}{\partial \mathbf{u}(t)} + \boldsymbol{\lambda}(t)\ \mathbf{A} + \boldsymbol{\sigma}(t)\ \mathbf{G} \geq \mathbf{0}, \tag{2.175}$$

$$\text{and} \qquad \mathbf{u}(t)\frac{\partial \mathcal{H}}{\partial \mathbf{u}(t)} = 0, \tag{2.176}$$

$$-\dot{\boldsymbol{\lambda}}(t) = \frac{\partial \mathcal{H}}{\partial \mathbf{x}(t)} = \frac{\partial F[\mathbf{x}(t), \mathbf{u}(t)]}{\partial \mathbf{x}(t)} + \boldsymbol{\eta}(t)\frac{\partial \mathbf{K}[\mathbf{x}(t)]}{\partial \mathbf{x}(t)} \tag{2.177}$$

$$\dot{\mathbf{x}}(t) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(t)} = \mathbf{A}\ \mathbf{u}(t) \tag{2.178}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\sigma}(t)} = \mathbf{0} = \mathbf{G}\ \mathbf{u}(t) \tag{2.179}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\eta}(t)} = \mathbf{K}[\mathbf{x}(t)] \geq \mathbf{0} \tag{2.180}$$

$$\boldsymbol{\lambda}(T) = \frac{\partial \Theta}{\partial \mathbf{x}(T)} = \frac{\partial \mathrm{S}[\mathbf{x}(T), T]}{\partial \mathbf{x}(T)} + \boldsymbol{\nu}(T)\frac{\partial \mathrm{N}[\mathbf{x}(T), T]}{\partial \mathbf{x}(T)} \tag{2.181}$$

$$\mathrm{N}[\mathbf{x}(T), T] = \mathbf{0} \tag{2.182}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \tag{2.183}$$

$$\mathbf{x}(t) \geq \mathbf{0}, \qquad \mathbf{u}(t) \geq \mathbf{0}. \tag{2.184}$$

Equations (2.175)-(2.177) are $3n$ differential equations for the two-point boundary value problems. Equations (2.178)-(2.180) are simply the restatement of state equation (2.167) and constraints (2.168)-(2.169). Equations (2.181)-(2.183) are boundary or transversality conditions. Equation (2.181) provides $n$ conditions with $q$ Lagrange multipliers to be determined. Equation (2.182) provides $q$ equations to eliminate the Lagrange multipliers, and equation (2.183) provides one additional equation which is used to determine the unspecified end time $T$.

## 2.5 Hierarchical Optimal Control Problems

In this section, we discuss some bilevel optimal control problems. We first discuss the utilization of leader-follower or Stackelberg strategy concepts in the control structuring of interconnected systems. These control methods are appropriate for system problems where there are multiple criteria, multiple decision makers, decentralized information and natural hierarchy of decision making levels. In dynamic transportation network problems, the multi-level formulation approach is very important. For example, travelers' choices, including departure time/mode/destination/route, constitute a natural hierarchy of decision making. Another example is the routing strategies in ATMIS systems. System-optimal and user-optimal strategies are conflicting in general. By imposing dynamic congestion pricing, we can formulate a bilevel model which coordinates the two routing criteria.

The basic leader-follower strategy was originally suggested for static duopoly by von Stackelberg (1952). The generalization of this concept to dynamic nonzero-sum two-person games was given by Cruz. Based on Cruz (1978), we first discuss a static two-person game which will be extended to multi-level optimal control problems.

### 2.5.1 Static Two-Person Games

The basic idea of a leader-follower strategy for a static two-person game is simple. Players 1 and 2 choose static controls $u_1 \in R$ and $u_2 \in R$, respectively. There are two scalar costs $J_1(u_1, u_2)$ and $J_2(u_1, u_2)$ associated with Players 1 and 2, respectively. Designate Player 1 as the leader and Player 2 as the follower. For each control $u_1$ chosen by Player 1, Player 2 chooses $u_2 = f_2(u_1)$ where $f_2$ is determined by $u_1$ and $u_2$ such that for Player 2, his/her cost using controls $[u_1, f_2(u_1)]$ is smaller than or equal to a cost using any controls $(u_1, u_2)$. It follows that

$$J_2[u_1, f_2(u_1)] \leq J_2(u_1, u_2) \qquad \text{for each } u_1 \text{ and for all } u_2 \qquad (2.185)$$

For simplicity, we assume that for each control $u_1$, $f_2(u_1)$ yields a unique $u_2$. The leader chooses an optimal control $u_1^*$ such that his/her cost using optimal controls $[u_1^*, f_2(u_1^*)]$ is smaller than or equal to a cost using any controls $[u_1, f_2(u_1)]$. It follows that

$$J_1[u_1^*, f_2(u_1^*)] \leq J_1[u_1, f_2(u_1)] \qquad \text{for all } u_2 \qquad (2.186)$$

Then, the optimal strategy $u_1^*$ is termed the Stackelberg strategy for Player 1 and $u_2^* = f_2(u_1^*)$ is termed the Stackelberg strategy for Player 2 when the leader is Player 1.

In this problem, we assume that the leader knows the cost mapping of the follower, but the follower doesn't know the cost mapping of the leader. However, the follower knows the control strategy of the leader and he/she takes this into account in computing his/her strategy. This reaction behavior of the follower is known to the leader who optimizes his/her choice of control $u_1$.

Similarly, when Player 1 is the follower and Player 2 is the leader, we have

$$J_1[f_1(u_2), u_2] \leq J_1(u_1, u_2) \qquad \text{for each } u_2 \text{ and for all } u_1 \qquad (2.187)$$

and

$$J_2[f_1(u_2^{**}), u_2^{**}] \leq J_2[f_1(u_2), u_2] \qquad \text{for all } u_2 \tag{2.188}$$

where $f_1$ is determined by $u_2$ and $u_1$. The optimal control $u_2^{**}$ is the leader Stackelberg strategy, and optimal control $u_1^{**} = f_1(u_2^{**})$ is the follower Stackelberg strategy.

Sometimes, we may refer to a Nash strategy or Nash control when we use game theory to tackle multi-level optimization problems. A Nash strategy pair $(\tilde{u}_1, \tilde{u}_2)$ is defined by

$$J_1(\tilde{u}_1, \tilde{u}_2) \leq J_1(u_1, \tilde{u}_2) \qquad \text{for all } u_1 \tag{2.189}$$

and

$$J_2(\tilde{u}_1, \tilde{u}_2) \leq J_2(\tilde{u}_1, u_2) \qquad \text{for all } u_2 \tag{2.190}$$

where the tilde ($\sim$) denotes that the control is a Nash control. We note that the Nash strategy may not be unique. As before, we define $\tilde{u}_2 = f_2(\tilde{u}_1)$ where $f_2$ is determined by $\tilde{u}_1$ and $\tilde{u}_2$. It follows that

$$J_2[\tilde{u}_1, \tilde{u}_2] = J_2[\tilde{u}_1, T_2(\tilde{u}_1)] \tag{2.191}$$

Combining equations (2.186) and (2.191), we have

$$J_1[u_1^*, f_2(u_1^*)] \leq J_1[f_1(\tilde{u}_2), \tilde{u}_2] \tag{2.192}$$

Similarly, we define $\tilde{u}_1 = f_1(\tilde{u}_2)$ where $f_1$ is determined by $\tilde{u}_1$ and $\tilde{u}_2$. It follows that

$$J_1(\tilde{u}_1, \tilde{u}_2) = J_1[T_1(\tilde{u}_2), \tilde{u}_2] \tag{2.193}$$

Combining equations (2.188) and (2.193), we have

$$J_2[f_1(u_2^{**}), u_2^{**}] \leq J_2(\tilde{u}_1, \tilde{u}_2) \tag{2.194}$$

Thus, for the leader, a Stackelberg strategy is at least as good as any Nash strategy. For the follower, the Stackelberg strategy may or may not be preferable compared to a Nash strategy.

### 2.5.2 Dynamic Games

Consider a dynamic system

$$\dot{x}(t) = f(x, u_1, u_2) \tag{2.195}$$

where $x \in R^n$ is the state, $u_1 \in R^{m_1}$ and $u_2 \in R^{m_2}$ are the controls, and $f$ is a piecewise continuous function from $R^n \times R^{m_1} \times R^{m_2}$ to $R^n$. The time interval $[0, T]$ is fixed and the initial state $x(0) = x_0$ is given. In a dynamic system, it is necessary to specify what type of information is available to each player. Suppose no state measurements are available. In this case, we consider open-loop strategies. Associated with each player is a scalar cost function, i.e.,

$$J_i = \int_0^T F_i(x, u_1, u_2)\, dt + S_i[x(T)] \qquad \forall\, i = 1, 2 \tag{2.196}$$

Designate Player 1 as the leader. The dynamic game can be written as a bilevel problem as follows.

**The Leader Problem**

$$\min \quad J_1 = \int_0^T F_1(x, u_1, u_2)\, dt + S_1[x(T)] \tag{2.197}$$

**The Follower Problem**

$$\min \quad J_2 = \int_0^T F_2(x, u_1, u_2)\, dt + S_2[x(T)] \tag{2.198}$$

s.t.

$$\dot{x}(t) = f(x, u_1, u_2) \tag{2.199}$$

$$x(0) = x_0 \tag{2.200}$$

The Hamiltonian function for the follower problem is written as

$$\mathcal{H}_2(x, u_1, u_2, p) = F_2(x, u_1, u_2) + p(t)\ f(x, u_1, u_2) \tag{2.201}$$

where $p(t)$ is the Lagrange multiplier. The necessary conditions for the follower problem are

$$\frac{\partial \mathcal{H}_2}{\partial u_2(t)} = 0 \tag{2.202}$$

$$\dot{p}(t) = -\frac{\partial \mathcal{H}_2}{\partial x(t)} \tag{2.203}$$

$$p(T) = -\frac{\partial S_2[x(T)]}{\partial x(T)} \tag{2.204}$$

In order to obtain the overall necessary conditions for this dynamic game, we need to place the necessary conditions (2.202)-(2.211) and constraints (2.199)-(2.200) of the follower problem as constraints for the leader problem. It follows that

$$\min \quad J_1 = \int_0^T F_1(x, u_1, u_2)\, dt + S_1[x(T)] \tag{2.205}$$

s.t.

$$\dot{x}(t) = f(x, u_1, u_2) \tag{2.206}$$

$$\dot{p}(t) = -\frac{\partial \mathcal{H}_2}{\partial x(t)} \tag{2.207}$$

$$\frac{\partial \mathcal{H}_2}{\partial u_2(t)} = 0 \tag{2.208}$$

$$p(T) + \frac{\partial S_2[x(T)]}{\partial x(T)} = 0 \tag{2.209}$$

$$p(T) + \frac{\partial S_2[x(T)]}{\partial x(T)} = 0 \tag{2.210}$$

$$p(T) = -\frac{\partial S_2[x(T)]}{\partial x(T)} \tag{2.211}$$

$$x(0) = x_0 \tag{2.212}$$

where equations (2.206)-(2.207) are considered as state equations for the redefined one-level problem. This is an optimal control problem with equality constraint and constraint for final states $x(T)$ and $p(T)$. The Hamiltonian function $\mathcal{H}_1$ for the redefined one-level problem is

$$\begin{aligned} \mathcal{H}_1(x, u_1, u_2, p, \lambda_1, \lambda_2, \beta_2) &= F_1(x, u_1, u_2) + \lambda_1(t)\, f(x, u_1, u_2) \\ &+ \lambda_2(t)\, (-\frac{\partial \mathcal{H}_2}{\partial x}) + \beta_2(t)\, \frac{\partial \mathcal{H}_2}{\partial u_2} \end{aligned} \tag{2.213}$$

where $\lambda_1(t), \lambda_2(t), \beta_2(t)$ are Lagrange multipliers. We define another scalar function as

$$\Theta[x(T), p(t), \nu(T), T] = S_1[x(T)] + \nu(T)\, \{p(T) - \frac{\partial S_2[x(T)]}{\partial x(T)}\} \tag{2.214}$$

where $\nu(T)$ is a Lagrange multiplier. The necessary conditions for the redefined one-level problem are

$$\frac{\partial \mathcal{H}_1}{\partial u_i(t)} = 0 \qquad \forall\, i = 1, 2 \tag{2.215}$$

$$\dot{\lambda}_1(t) = -\frac{\partial \mathcal{H}_1}{\partial x(t)} \tag{2.216}$$

$$\lambda_2(t) = -\frac{\partial \mathcal{H}_1}{\partial p(t)} \tag{2.217}$$

$$\lambda_1(T) = \frac{\partial \Theta}{\partial x(T)} = \frac{\partial S_1[(x(T)]}{\partial x(T)} - \nu(T)\,\frac{\partial S_2[(x(T)]}{\partial x(T)} \tag{2.218}$$

$$\lambda_2(T) = \frac{\partial \Theta}{\partial p(T)} = \nu(T) \tag{2.219}$$

Therefore, the necessary conditions for $(u_1, u_2)$ to be an open-loop Stackelberg strategy pair are equations (2.202)-(2.211) and (2.215)-(2.219). Explicit solution in terms of the Riccati matrix equations can be obtained for the linear-quadratic problem. However, necessary conditions for the closed-loop Stackelberg strategy are extremely difficult to characterize. Simplification is possible when the structure of the control law is constrained, e.g., restricting the control law to be linear and the effects of random initial conditions are averaged.

Instead of discussing detailed control strategies, we state some conclusions from the literature (Cruz, 1978). The open-loop strategy for the leader for the entire duration of the game is declared in advance. If the follower minimizes its cost function, it obtains its follower Stackelberg strategy which is the optimal reaction to the declared leader strategy. By declaring his/her strategy in advance, the leader influences the follower to react in a manner which minimizes the follower's cost function, but more importantly, in a manner which is favorable to the leader. This is a direct interpretation of the definition of the leader's strategy. Similarly, for closed-loop strategies where the state is available for measurement, the leader has to declare his control law for the entire duration of the game. In situations where either player might be a leader, both cases should be examined because both players may insist on leader strategies in which case there may be disequilibrium, or both may play follower strategies and a stalemate may occur.

### 2.5.3 Bilevel Optimal Control Problems

We now generalize the above dynamic games into bilevel optimal control problems. Sometimes, we term this kind of optimization problem as hierarchical optimization, which is contrary to the bilevel optimization problems having less coordination or non-hierarchical properties. To simplify our presentation, we only discuss optimal control problems with equality and nonnegativity constraints which are constrained by fixed beginning and end times. Consider two linear dynamic systems

$$\frac{d\mathbf{x}_i(t)}{dt} = \mathbf{A}_i\ \mathbf{u}_i(t) \qquad \forall\, i = 1, 2 \tag{2.220}$$

The two dynamic systems have two decision makers, each with an objective function as follows

$$\min_{\mathbf{x},\mathbf{u}} J_i = \int_0^T F_i[\mathbf{x}_1(t), \mathbf{u}_1(t), \mathbf{x}_2(t), \mathbf{u}_2(t)]dt + S_i[\mathbf{x}_1(T), \mathbf{x}_2(T)] \tag{2.221}$$

$$i = 1, 2$$

where each objective function depends on controls and states of both dynamic systems. To make the problem more general, we assume that control variables of both systems are co-related by the following linear constraints

$$\mathbf{G}_1 \ \mathbf{u}_1(t) + \mathbf{G}_2 \ \mathbf{u}_2(t) = 0 \tag{2.222}$$

where $\mathbf{G}_1, \mathbf{G}_2$ are vectors of constants. We designate Player (decision maker) 1 as the leader. We then formulate the bilevel optimal control problem as follows.

**Upper Level Problem**

$$\min_{\mathbf{x}_1,\mathbf{u}_1} \quad J_1 = \int_0^T F_1[\mathbf{x}_1(t), \mathbf{u}_1(t), \mathbf{x}_2(t), \mathbf{u}_2(t)] \, dt + S_1[\mathbf{x}_1(T), \mathbf{x}_2(T)] \tag{2.223}$$

s.t.

$$\frac{d\mathbf{x}_1(t)}{dt} = \mathbf{A}_1 \ \mathbf{u}_1(t) \tag{2.224}$$

$$\mathbf{G}_1 \ \mathbf{u}_1(t) + \mathbf{G}_2 \ \mathbf{u}_2(t) = \mathbf{0} \tag{2.225}$$

$$\mathbf{K}_1[\mathbf{x}_1(t)] = \mathbf{0} \tag{2.226}$$

$$\mathbf{x}_1(t) \geq \mathbf{0}, \qquad \mathbf{u}_1(t) \geq \mathbf{0} \tag{2.227}$$

$$\mathbf{x}_1(0) \qquad \text{given.} \tag{2.228}$$

where $\mathbf{x}_2(t), \mathbf{u}_2(t)$ are determined by the following lower level problem:

**Lower Level Problem**

$$\min_{\mathbf{x}_2,\mathbf{u}_2} \quad J_2 = \int_0^T F_2[\mathbf{x}_1(t), \mathbf{u}_1(t), \mathbf{x}_2(t), \mathbf{u}_2(t)] \, dt + S_2[\mathbf{x}_1(T), \mathbf{x}_2(T)] \tag{2.229}$$

s.t.

$$\frac{d\mathbf{x}_2(t)}{dt} = \mathbf{A}_2 \ \mathbf{u}_2(t) \tag{2.230}$$

$$\mathbf{x}_2(t) \geq \mathbf{0}, \qquad \mathbf{u}_2(t) \geq \mathbf{0} \tag{2.231}$$

$$\mathbf{x}_2(0) \qquad \text{given.} \tag{2.232}$$

In the lower level problem, $\mathbf{x}_1(t), \mathbf{u}_1(t)$ are determined by the upper level problem. In this bilevel problem, the leader (upper level problem) has the priority to minimize its objective function $J_1$. We now examine the necessary conditions for the bilevel optimal control problem. We first consider the necessary conditions for the lower level problem which is designated as the follower. The overall necessary conditions for the bilevel problem are constructed by placing the necessary conditions and constraints of the lower level problem as constraints for the upper level problem.

We construct the augmented Hamiltonian $\mathcal{H}_2$ for the lower level problem as

$$\mathcal{H}_2 = F_2[\mathbf{x}_1(t), \mathbf{u}_1(t), \mathbf{x}_2(t), \mathbf{u}_2(t)] + \mathbf{p}(t)\ \mathbf{A}_2\ \mathbf{u}_2(t) \tag{2.233}$$

where $\mathbf{p}(t)$ is a vector of Lagrange multipliers. The first order necessary conditions for the lower level optimal control program are

$$\frac{\partial \mathcal{H}_2}{\partial \mathbf{u}_2(t)} = \frac{\partial F_2[\mathbf{x}_1(t), \mathbf{u}_1(t), \mathbf{x}_2(t), \mathbf{u}_2(t)]}{\partial \mathbf{u}_2(t)} + \mathbf{A}_2\ \mathbf{p}(t)\ \geq \mathbf{0} \tag{2.234}$$

$$\mathbf{u}_2(t) \frac{\partial \mathcal{H}_2}{\partial \mathbf{u}_2(t)} = \mathbf{0} \tag{2.235}$$

$$-\dot{\mathbf{p}}(t) = \frac{\partial \mathcal{H}_2}{\partial \mathbf{x}_2(t)} = \frac{\partial F_2[\mathbf{x}_1(t), \mathbf{u}_1(t), \mathbf{x}_2(t), \mathbf{u}_2(t)]}{\partial \mathbf{x}_2(t)} \tag{2.236}$$

$$\frac{d\mathbf{x}_2(t)}{dt} = \frac{\partial \mathcal{H}_2}{\partial \mathbf{p}(t)} = \mathbf{A}_2\ \mathbf{u}_2(t) \tag{2.237}$$

$$\mathbf{p}(T) = \frac{\partial S_2[\mathbf{x}_1(T), \mathbf{x}_2(T)]}{\partial \mathbf{x}_2(T)} \tag{2.238}$$

$$\mathbf{x}_2(t) \geq \mathbf{0}, \qquad \mathbf{u}_2(t) \geq \mathbf{0} \tag{2.239}$$

$$\mathbf{x}_2(0) \qquad \text{given} \tag{2.240}$$

Then, we convert the bilevel optimal control problem into a one-level optimal control problem by placing necessary conditions (2.234)-(2.240) of the lower level problem as constraints for the upper level problem. It follows that

$$\min_{\mathbf{x},\mathbf{u}}\ J_1 = \int_0^T F_1[\mathbf{x}_1(t), \mathbf{u}_1(t), \mathbf{x}_2(t), \mathbf{u}_2(t)]\, dt + S_1[\mathbf{x}_1(T), \mathbf{x}_2(T)] \tag{2.241}$$

s.t.

$$\frac{d\mathbf{x}_1(t)}{dt} = \mathbf{A}_1\ \mathbf{u}_1(t) \tag{2.242}$$

$$\frac{d\mathbf{x}_2(t)}{dt} = \mathbf{A}_2\ \mathbf{u}_2(t) \tag{2.243}$$

$$\dot{\mathbf{p}}(t) = -\frac{\partial \mathcal{H}_2}{\partial \mathbf{x}_2(t)} \tag{2.244}$$

$$\mathbf{G}_1\ \mathbf{u}_1(t) + \mathbf{G}_2\ \mathbf{u}_2(t) = \mathbf{0} \tag{2.245}$$

$$\mathbf{K}_1[\mathbf{x}_1(t)] = \mathbf{0} \tag{2.246}$$

$$\frac{\partial \mathcal{H}_2}{\partial \mathbf{u}_2(t)} \geq \mathbf{0} \tag{2.247}$$

$$\mathbf{u}_2(t) \frac{\partial \mathcal{H}_2}{\partial \mathbf{u}_2(t)} = \mathbf{0} \tag{2.248}$$

$$\mathbf{x}_1(t) \geq \mathbf{0}, \qquad \mathbf{u}_1(t) \geq \mathbf{0}, \qquad \mathbf{x}_2(t) \geq \mathbf{0}, \qquad \mathbf{u}_2(t) \geq \mathbf{0} \tag{2.249}$$

$$\mathbf{x}_1(0), \qquad \mathbf{x}_2(0) \qquad \text{given} \tag{2.250}$$

$$\mathbf{p}(T) = \frac{\partial S_2[\mathbf{x}_1(T), \mathbf{x}_2(T)]}{\partial \mathbf{x}_2(T)} \tag{2.251}$$

where equations (2.242)-(2.244) are considered as state equations. We then construct the augmented Hamiltonian $\mathcal{H}_1$ for the converted one-level optimal control program as

$$\begin{aligned} \mathcal{H}_1 &= F_1 + \boldsymbol{\lambda}_1(t)\ \mathbf{A}_1\ \mathbf{u}_1(t) + \boldsymbol{\lambda}_2(t)\ \mathbf{A}_2\ \mathbf{u}_2(t) + \boldsymbol{\lambda}_3(t)\ [-\frac{\partial \mathcal{H}_2}{\partial \mathbf{x}_2(t)}] \\ &+ \boldsymbol{\sigma}(t)\ [\mathbf{G}_1\ \mathbf{u}_1(t) + \mathbf{G}_2\ \mathbf{u}_2(t)] + \boldsymbol{\eta}(t)\ \mathbf{K}_1[\mathbf{x}_1(t)] \\ &+ \boldsymbol{\beta}_1(t)\ \frac{\partial \mathcal{H}_2}{\partial \mathbf{u}_2(t)} + \boldsymbol{\beta}_2(t)\ [\mathbf{u}_2(t) \frac{\partial \mathcal{H}_2}{\partial \mathbf{u}_2(t)}] \end{aligned} \tag{2.252}$$

where $\boldsymbol{\lambda}_1(t), \boldsymbol{\lambda}_2(t), \boldsymbol{\lambda}_3(t), \boldsymbol{\sigma}(t), \boldsymbol{\eta}(t), \boldsymbol{\beta}_1(t), \boldsymbol{\beta}_2(t)$ are Lagrange multipliers and $\boldsymbol{\beta}_1(t) \leq 0$. We define another scalar function as

$$\Theta[\mathbf{x}(T), \mathbf{p}(t), \boldsymbol{\nu}(T), T] = S_1[\mathbf{x}(T)] + \boldsymbol{\nu}(T)\ \{\mathbf{p}(T) - \frac{\partial S_2[\mathbf{x}(T)]}{\partial \mathbf{x}_2(T)}\} \tag{2.253}$$

where $\boldsymbol{\nu}(T)$ is a vector of Lagrange multipliers. The first order necessary conditions for the converted one-level optimal control program are

$$\frac{\partial \mathcal{H}_1}{\partial \mathbf{u}_i(t)} \geq \mathbf{0} \qquad \forall\, i = 1, 2 \tag{2.254}$$

$$\mathbf{u}_i(t) \frac{\partial \mathcal{H}_1}{\partial \mathbf{u}_i(t)} = 0 \qquad \forall\, i = 1, 2 \tag{2.255}$$

$$\dot{\boldsymbol{\lambda}}_i(t) = -\frac{\partial \mathcal{H}_1}{\partial \mathbf{x}_i(t)} \qquad \forall\, i = 1, 2 \tag{2.256}$$

$$\dot{\boldsymbol{\lambda}}_3(t) = -\frac{\partial \mathcal{H}_1}{\partial \mathbf{p}(t)} \tag{2.257}$$

$$\dot{\mathbf{x}}_i(t) = \frac{\partial \mathcal{H}_1}{\partial \boldsymbol{\lambda}_i(t)} = \mathbf{A}_i\ \mathbf{u}_i(t) \qquad \forall\, i = 1, 2 \tag{2.258}$$

$$\dot{\mathbf{p}}(t) = \frac{\partial \mathcal{H}_1}{\partial \boldsymbol{\lambda}_3(t)} = -\frac{\partial \mathcal{H}_2}{\partial \mathbf{x}_2(t)} \tag{2.259}$$

$$\frac{\partial \mathcal{H}_1}{\partial \boldsymbol{\sigma}(t)} = \mathbf{0} = \mathbf{G}_1\, \mathbf{u}_1(t) + \mathbf{G}_2\, \mathbf{u}_2(t) \tag{2.260}$$

$$\frac{\partial \mathcal{H}_1}{\partial \boldsymbol{\eta}(t)} = \mathbf{K}_1[\mathbf{x}_1(t)] = \mathbf{0} \tag{2.261}$$

$$\frac{\partial \mathcal{H}_1}{\partial \boldsymbol{\beta}_1(t)} = \frac{\partial \mathcal{H}_2}{\partial \mathbf{u}_2(t)} \tag{2.262}$$

$$\frac{\partial \mathcal{H}_1}{\partial \boldsymbol{\beta}_2(t)} = \mathbf{u}_2(t)\frac{\partial \mathcal{H}_2}{\partial \mathbf{u}_2(t)} \tag{2.263}$$

$$\boldsymbol{\lambda}_i(T) = \frac{\partial \Theta}{\partial \mathbf{x}_i(T)} \qquad \forall\, i = 1,2 \tag{2.264}$$

$$\boldsymbol{\lambda}_3(T) = \frac{\partial \Theta}{\partial \mathbf{p}(T)} = \boldsymbol{\nu}(T) \tag{2.265}$$

$$\mathbf{x}_i(0) = \mathbf{x}_{i0}, \qquad \forall\, i = 1,2 \tag{2.266}$$

$$\mathbf{x}_i(t) \geq \mathbf{0}, \qquad \mathbf{u}_i(t) \geq \mathbf{0} \qquad \forall\, i = 1,2 \tag{2.267}$$

Therefore, the necessary conditions for the bilevel program are equations (2.234)-(2.240) and (2.254)-(2.267). However, the detailed analysis of the necessary conditions is complicated, especially when the two level problems have many constraints. In general, the analytical solution for the above bilevel program is not possible even for simple cost functions.

On the other hand, if we assume that the above problem has no hierarchical relationship, we can solve the two level problems separately and transmit some shared variables back and forth. In this way, no leader-follower relationship is assumed between the two level problems. Thus, this situation can be termed a coordination problem where the shared variables function as coordinators. This approach can also be extended to a multi-level optimal control problem where the hierarchical relationship is hard to identify.

## 2.6 Notes

In classical control system design, the ultimate objective is to obtain a controller that will allow a system to perform in a desirable manner. The objective in control problem formulation for transportation networks is also to obtain a set of control variables which can adjust traffic flows in the network to behave in desired ways.

Many complex transportation problems can be formulated and solved using optimal control theory. However, at the present time, optimal control theory does not constitute a generally applicable procedure for the design of a simple controller. The optimal control law, if it can be obtained, usually requires a digital computer for implementation (an important exception is the linear regulator problem discussed in section 2.3), and all of the states must be available for feedback to the controller. These limitations may preclude implementation of the optimal control law; however, the theory of optimal control is still useful, because knowing the optimal control law may provide insight helpful in designing a suboptimal, but easily implemented controller. The optimal control law provides a standard for evaluating proposed suboptimal designs. In other words, by knowing the optimal control law we have a quantitative measure of performance degradation caused by using a suboptimal controller.

The basic knowledge of the optimal control problem can be found in any standard optimal control text. At the elementary level, readers may consult the text by Kirk (1970). Bryson and Ho (1975) provide many examples with which readers may find it is easier to understand the theory. Other closely related texts are Kamien and Schwartz (1981) and Sage and White (1977). A proof of optimality conditions for optimal control problems with general constraints for state and control variables is provided by Russak (1970). The applications of game theory in dynamic systems and optimal control problems are summarized by Cruz (1978). For an advanced text on multi-level optimal control problems, readers may consult the text by Singh and Titli (1978).

# Chapter 3

# Discrete Optimal Control, Mathematical Programming and Variational Inequality Problems

In this chapter, we present more mathematical background which is necessary for modeling and solution of dynamic transportation network problems. This chapter will cover discrete optimal control, mathematical programming and variational inequality problems. First, we introduce the discrete optimal control problem (OCP). To simplify our presentation, we consider discrete optimal control problems with fixed end times as examples in Section 3.1. The discussion is focused on the analysis of optimality conditions. Then, some mathematical programming (MP) problems are presented in Section 3.2. Specifically, nonlinear programming (NLP) problems with equality and nonnegativity constraints are presented for comparison. Similarities between discrete optimal control problems and mathematical programming are emphasized.

Beyond optimal control problems and nonlinear programming problems, we also provide some basic concepts of variational inequality problems which are capable of formulating and analyzing more general problems than the constrained optimization approach. Variational inequality (VI) problems are presented in Section 3.3 and are suggested for advanced readers with knowledge of optimization problems. We first define variational inequality problems for both static and dynamic problems. We then introduce some fundamental definitions, along with qualitative results for variational inequality problems, such as conditions for existence and uniqueness of solutions.

In Section 3.4, we present algorithms for solving NLP and VI, including one-dimensional search, the Frank-Wolfe algorithm and a relaxation method. Unlike traditional algorithms for solving optimal control problems, we suggest using these algorithms to solve discrete optimal control problems.

## 3.1 Discrete Optimal Control Problems with Fixed Beginning and End Times

In this section, we consider several discrete optimal control problems with fixed end times. These problems are widely used in our formulations and solutions in the following chapters. The analysis of optimal control problems with free end times follows very easily. We first discuss discrete optimal control problems with no constraints. Then, discrete optimal control problems with equality and inequality constraints are analyzed. Finally, we investigate discrete optimal control problems with equality and nonnegativity constraints.

In discrete optimal control problems, we discretize the time period $[0, T]$ into $K + 1$ small time intervals or increments, i.e., $k = 1, 2, \cdots, K + 1$. For simplicity, each interval is assumed to have equal length $\Delta$. In general, we use the first difference approximation to replace $\dot{\mathbf{x}}(t)$. It follows that

$$\dot{\mathbf{x}}(t) \approx \frac{\mathbf{x}(k+1) - \mathbf{x}(k)}{\Delta} \tag{3.1}$$

In most analysis, it is convenient to assume $\Delta$ is a unit value. Thus, the above difference approximation can be rewritten as

$$\dot{\mathbf{x}}(t) \approx \mathbf{x}(k+1) - \mathbf{x}(k) \tag{3.2}$$

Thus, the discrete state equation can be written as

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), k] \tag{3.3}$$

In the same way, the discretization of the Lagrange multipliers

$$\dot{\boldsymbol{\lambda}}(t) \approx \frac{\boldsymbol{\lambda}(k+1) - \boldsymbol{\lambda}(k)}{\Delta} \tag{3.4}$$

can be rewritten as

$$\dot{\boldsymbol{\lambda}}(t) \approx \boldsymbol{\lambda}(k+1) - \boldsymbol{\lambda}(k) \tag{3.5}$$

In continuous optimal control problems, we minimize cost functions which are integrals of scalar cost functions. We now consider a dynamic system having a cost function $F[\mathbf{x}(k), \mathbf{u}(k), k]$ for each time interval $k$. Thus, we are interested in minimization of cost functions which are summations of scalar functions. It follows that

$$\min_{\mathbf{x},\mathbf{u}} \quad J = \sum_{k=1}^{K} F[\mathbf{x}(k), \mathbf{u}(k), k] \; + \; S[\mathbf{x}(K+1), K+1] \tag{3.6}$$

### 3.1.1 Fixed End Times: No Constraints

The discrete optimal control problem is to seek an admissible control function $\mathbf{u}$ in order to minimize the objective function

$$\min_{\mathbf{x},\mathbf{u}} \quad J = \sum_{k=1}^{K} F[\mathbf{x}(k), \mathbf{u}(k), k] \; + \; S[\mathbf{x}(K+1), K+1] \tag{3.7}$$

s.t.

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), k] \tag{3.8}$$

$$K \quad \text{and} \quad \mathbf{x}(1) = \mathbf{x}_1 \text{ fixed}; \quad \mathbf{x}(K+1) \text{ free} \tag{3.9}$$

where $F[\mathbf{x}(k), \mathbf{u}(k), k]$ possesses continuous partial derivatives with respect to $\mathbf{x}(k)$ and $\mathbf{u}(k)$, and $S[\mathbf{x}(K+1), K+1]$ has continuous partial derivatives with respect to $\mathbf{x}(K+1)$. $S[\mathbf{x}(K+1), K+1]$ is associated with the end time $(K+1)$ only, and is termed the salvage cost in many economics problems.

We define the Hamiltonian as

$$\mathcal{H} = F[\mathbf{x}(k), \mathbf{u}(k), k] + \boldsymbol{\lambda}(k)\, \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), k] \tag{3.10}$$

where $\boldsymbol{\lambda}(k)$ is the vector of Lagrange multipliers associated with the dynamic state equations. Then, the first-order necessary conditions for the optimal control problem are summarized as follows:

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}(k)} = \frac{\partial F[\mathbf{x}(k), \mathbf{u}(k), k]}{\partial \mathbf{u}(k)} + \boldsymbol{\lambda}(k) \frac{\partial \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), k]}{\partial \mathbf{u}(k)} = \mathbf{0} \tag{3.11}$$

$$\begin{aligned} \boldsymbol{\lambda}(k+1) - \boldsymbol{\lambda}(k) &= \frac{\partial \mathcal{H}}{\partial \mathbf{x}(k)} \\ &= \frac{\partial F[\mathbf{x}(k), \mathbf{u}(k), k]}{\partial \mathbf{x}(k)} + \boldsymbol{\lambda}(k) \frac{\partial \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), k]}{\partial \mathbf{x}(k)} \end{aligned} \tag{3.12}$$

$$\mathbf{x}(k+1) - \mathbf{x}(k) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(k)} = \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), k] \tag{3.13}$$

$$\mathbf{x}(1) = \mathbf{x}_1 \tag{3.14}$$

$$\boldsymbol{\lambda}(K+1) = \frac{\partial S[\mathbf{x}(K+1), K+1]}{\partial \mathbf{x}(K+1)} \tag{3.15}$$

Equations (3.13) and (3.14) are simply the restatement of state equation (3.8) and boundary condition (3.9). Equations (3.14) and (3.15) are termed transversality conditions which constitute two-point boundary conditions for the set of differential equations. The above first-order necessary conditions are derived using the discrete Minimum (or Maximum) Principle or the Hamilton-Jacobi equation. The discrete Minimum Principle can simply be stated as follows. The inequality

$$\mathcal{H}[\mathbf{u}^*(k)] \leq \mathcal{H}[\mathbf{u}(k)] \tag{3.16}$$

is valid for all admissible $\mathbf{u}(k)$, where $*$ represents that the solution is optimal. Sometimes, we directly use this principle to perform solution analyze of complicated discrete optimal control problems.

### 3.1.2 Fixed End Times: Equality and Inequality Constraints

In this section, we consider discrete optimal control problems with nonlinear equality and inequality constraints on state and control variables. Those constraints can represent most practical constraints in realistic applications. The set of equality constraints for control and state variables are denoted as $\mathbf{G}[\mathbf{u}(k), \mathbf{x}(k), k] = 0$ and the set of inequality constraints are denoted as $\mathbf{K}[\mathbf{u}(k), \mathbf{x}(k), k] \leq 0$. We assume both $\mathbf{G}$ and $\mathbf{K}$ are continuous and differentiable with respect to $\mathbf{u}$, $\mathbf{x}$ and $k$. The optimal control problem is formulated as follows

$$\min_{\mathbf{x},\mathbf{u}} \quad J = \sum_{k=1}^{K} F[\mathbf{x}(k), \mathbf{u}(k), k] \; + \; S[\mathbf{x}(K+1), K+1] \tag{3.17}$$

s.t.

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), k] \tag{3.18}$$

$$\mathbf{G}[\mathbf{u}(k), \mathbf{x}(k), k] = \mathbf{0} \tag{3.19}$$

$$\mathbf{K}[\mathbf{u}(k), \mathbf{x}(k), k] \leq \mathbf{0} \tag{3.20}$$

$$\mathbf{x}(1) = \mathbf{x}_1, \qquad K \;\; \text{fixed.} \tag{3.21}$$

Denote the Lagrangian multipliers associated with equations (3.18)-(3.20) as $\boldsymbol{\lambda}(k)$, $\boldsymbol{\sigma}(k)$ and $\boldsymbol{\eta}(k)$, respectively. We construct the augmented Hamiltonian $\mathcal{H}$ for the above optimal control problem as

$$\begin{aligned} \mathcal{H} \;\; &= \;\; F[\mathbf{x}(k), \mathbf{u}(k), k] + \boldsymbol{\lambda}(k)\; \mathbf{f}[\mathbf{u}(k), \mathbf{x}(k), k] \\ &+ \;\; \boldsymbol{\sigma}(k)\; \mathbf{G}[\mathbf{u}(k), \mathbf{x}(k), k] + \boldsymbol{\eta}(k)\; \mathbf{K}[\mathbf{u}(k), \mathbf{x}(k), k] \end{aligned} \tag{3.22}$$

where

$$\boldsymbol{\eta}(k) \begin{cases} > \mathbf{0} & \text{if } \mathbf{K} = \mathbf{0} \\ = \mathbf{0} & \text{if } \mathbf{K} < \mathbf{0} \end{cases}$$

The first order necessary conditions for the optimal control program are

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial \mathbf{u}(k)} = \mathbf{0} \;\; &= \;\; \frac{\partial F[\mathbf{x}(k), \mathbf{u}(k), k]}{\partial \mathbf{u}(k)} + \boldsymbol{\lambda}(k) \frac{\partial \mathbf{f}[\mathbf{u}(k), \mathbf{x}(k), k]}{\partial \mathbf{u}(k)} \\ &+ \;\; \boldsymbol{\sigma}(k) \frac{\partial \mathbf{G}[\mathbf{u}(k), \mathbf{x}(k), k]}{\partial \mathbf{u}(k)} + \boldsymbol{\eta}(k) \frac{\partial \mathbf{K}[\mathbf{u}(k), \mathbf{x}(k), k]}{\partial \mathbf{u}(k)} \end{aligned} \tag{3.23}$$

$$\begin{aligned} -\boldsymbol{\lambda}(k) + \boldsymbol{\lambda}(k+1) &= \frac{\partial \mathcal{H}}{\partial \mathbf{x}(k)} \\ &= \;\; \frac{\partial F[\mathbf{x}(k), \mathbf{u}(k), k]}{\partial \mathbf{x}(k)} + \boldsymbol{\lambda}(k) \frac{\partial \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), k]}{\partial \mathbf{x}(k)} \\ &+ \;\; \boldsymbol{\sigma}(k) \frac{\partial \mathbf{G}[\mathbf{u}(k), \mathbf{x}(k), k]}{\partial \mathbf{x}(k)} + \boldsymbol{\eta}(k) \frac{\partial \mathbf{K}[\mathbf{u}(k), \mathbf{x}(k), k]}{\partial \mathbf{x}(k)} \end{aligned} \tag{3.24}$$

$$\mathbf{x}(k+1) - \mathbf{x}(k) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(k)} = \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), k] \tag{3.25}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\sigma}(k)} = \mathbf{0} = \mathbf{G}[\mathbf{u}(k), \mathbf{x}(k), k] \tag{3.26}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\eta}(k)} = \mathbf{K}[\mathbf{u}(k), \mathbf{x}(k), k] \leq \mathbf{0} \tag{3.27}$$

$$\mathbf{x}(1) = \mathbf{x}_1 \tag{3.28}$$

$$\boldsymbol{\lambda}(K+1) = \frac{\partial S[\mathbf{x}(K+1), K+1]}{\partial \mathbf{x}(K+1)} \tag{3.29}$$

Equations (3.23)-(3.24) are similar to equations (3.11) and (3.12) except for additional terms resulting from equality and inequality constraints on the control and state variables. Equation (3.25) is a restatement of state equation (3.18). Equations (3.26)-(3.27) are restatements of the equality and inequality constraints. Equations (3.28)-(3.29) are two-point boundary conditions.

### 3.1.3 Fixed End Times: Equality and Nonnegativity Constraints

We now consider a set of special discrete optimal control problems which will be widely used in the following chapters for formulation and analysis of dynamic transportation network models. The state equations are linear

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{A}\ \mathbf{u}(k) \tag{3.30}$$

where $\mathbf{A}$ is a matrix of constants. We consider only linear equality constraints for control variables

$$\mathbf{G}\ \mathbf{u}(k) = \mathbf{0} \tag{3.31}$$

and nonlinear equality constraints involving only state variables

$$\mathbf{K}[\mathbf{x}(k)] = \mathbf{0} \tag{3.32}$$

where $\mathbf{G}$ is a matrix of constants. The discrete optimal control problem is formulated as follows

$$\min_{\mathbf{x},\mathbf{u}} \quad J = \sum_{k=1}^{K} F[\mathbf{x}(k), \mathbf{u}(k), k] \ + \ S[\mathbf{x}(K+1), K+1] \tag{3.33}$$

| s.t. | | Lagrange Multiplier | |
|---|---|---|---|
| Linear State Equations | $\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{A}\ \mathbf{u}(k)$ | $\boldsymbol{\lambda}(k)$ | (3.34) |
| Linear Control Variable Constraints | $\mathbf{G}\ \mathbf{u}(k) = \mathbf{0}$ | $\boldsymbol{\sigma}(k)$ | (3.35) |
| State Variable Constraints | $\mathbf{K}[\mathbf{x}(k)] = \mathbf{0}$ | $\boldsymbol{\eta}(k)$ | (3.36) |

$$\mathbf{x}(k) \geq \mathbf{0}, \qquad \mathbf{u}(k) \geq \mathbf{0}, \tag{3.37}$$

$$\mathbf{x}(1) \text{ given.} \tag{3.38}$$

We construct the augmented Hamiltonian $\mathcal{H}$ for the above discrete optimal control problem as

$$\mathcal{H} = F[\mathbf{x}(k), \mathbf{u}(k), k] + \boldsymbol{\lambda}(k)\ \mathbf{A}\ \mathbf{u}(k) + \boldsymbol{\sigma}(k)\ \mathbf{G}\ \mathbf{u}(k) + \boldsymbol{\eta}(k)\ \mathbf{K}\ [\mathbf{x}(k)] \tag{3.39}$$

The first order necessary conditions for the discrete optimal control program are

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}(k)} = \frac{\partial F[\mathbf{x}(k), \mathbf{u}(k), k]}{\partial \mathbf{u}(k)} + \boldsymbol{\lambda}(k)\ \mathbf{A} + \boldsymbol{\sigma}(k)\ \mathbf{G} \geq \mathbf{0}, \tag{3.40}$$

$$\text{and} \qquad \mathbf{u}(k)\frac{\partial \mathcal{H}}{\partial \mathbf{u}(k)} = 0, \tag{3.41}$$

$$\begin{aligned} -\boldsymbol{\lambda}(k) + \boldsymbol{\lambda}(k+1) &= \frac{\partial \mathcal{H}}{\partial \mathbf{x}(k)} \\ &= \frac{\partial F[\mathbf{x}(k), \mathbf{u}(k), k]}{\partial \mathbf{x}(k)} + \boldsymbol{\eta}(k)\frac{\partial \mathbf{K}[\mathbf{x}(k)]}{\partial \mathbf{x}(k)} \end{aligned} \tag{3.42}$$

$$\mathbf{x}(k+1) - \mathbf{x}(k) = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}(k)} = \mathbf{A}\ \mathbf{u}(k) \tag{3.43}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\sigma}(k)} = \mathbf{0} = \mathbf{G}\ \mathbf{u}(k) \tag{3.44}$$

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\eta}(k)} = \mathbf{K}[\mathbf{x}(k)] = \mathbf{0} \tag{3.45}$$

$$\boldsymbol{\lambda}(K+1) = \frac{\partial S[\mathbf{x}(K+1)]}{\partial \mathbf{x}(K+1)} \tag{3.46}$$

$$\mathbf{x}(1) = \mathbf{x}_1 \tag{3.47}$$

$$\mathbf{x}(k) \geq \mathbf{0}, \qquad \mathbf{u}(k) \geq \mathbf{0}. \tag{3.48}$$

Equations (3.40)-(3.42) are similar to equations (3.11) and (3.12) except for additional terms resulting from equality constraints on control and state variables. The inequality sign in equation (3.40) is caused by the nonnegativity constraint for the control variable. Equation (3.43) is a restatement of the state equation (3.34). Equations (3.44)-(3.45) are restatements of the equality constraints. Equations (3.46)-(3.48) are two-point boundary conditions.

# 3.2 Mathematical Programming Problems

This section reviews some concepts related to the formulation and solution of mathematical optimization programs. The focus of the discussion is on the conditions that characterize the solution of such programs and the constraint conditions which are similar to discrete optimal control problems. To have a systematic description of mathematical programming problems, we first discuss unconstrained minimization problems. Then, we explore minimization problems with general constraints and the Karush-Kuhn-Tucker conditions. The relationship between the discrete optimal control problem and mathematical programming problem is then investigated. In comparison with bilevel optimal control problems, bilevel mathematical programming problems are also presented.

## 3.2.1 Unconstrained Minimization

We first discuss minimization problems without constraints. Our problem is to find a set of variables $x_1, x_2, \cdots, x_n$ that minimize an objective function $Z(x_1, x_2, \cdots, x_n)$. Let $\mathbf{x}$ denote the vector of decision variables, i.e. $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. Using vector notation, the minimization program without constraints can be stated as

$$\min_{\mathbf{x}} \quad Z(\mathbf{x}) \tag{3.49}$$

Since the function to be minimized is unconstrained, the first-order necessary condition for a minimum at $\mathbf{x} = \mathbf{x}^*$ is that the gradient of $Z(\mathbf{x})$ vanish at $\mathbf{x}^*$. The gradient of $Z(\mathbf{x})$ with respect to $\mathbf{x}^*$, $\nabla_{\mathbf{x}} Z(\mathbf{x})$, is the vector of partial derivatives, that is,

$$\nabla_{\mathbf{x}}\, Z(\mathbf{x}) = \left[\frac{\partial\, Z(\mathbf{x})}{\partial x_1}, \frac{\partial\, Z(\mathbf{x})}{\partial x_2}, \cdots, \frac{\partial\, Z(\mathbf{x})}{\partial x_n}\right] \tag{3.50}$$

At every point $\mathbf{x}$, the gradient points in the direction of the steepest increase in $Z(\mathbf{x})$. The first-order necessary conditions for a minimum are

$$\nabla\, Z(\mathbf{x}^*) = \mathbf{0} \tag{3.51}$$

In other words, each element of the gradient has to equal to zero. Equivalently,

$$\frac{\partial Z(\mathbf{x}^*)}{\partial x_i} = 0 \qquad \forall\, i = 1, 2, \cdots, n \tag{3.52}$$

The sufficient conditions that $x^*$ is the local minimum of $Z(\mathbf{x})$ depend on establishing that $Z(\mathbf{x})$ is locally convex in the vicinity of $x^*$. In other words, any line sequent joining two points $x_1$ and $x_2$ lies above the surface $Z(\mathbf{x})$. The strict convexity of $Z(\mathbf{x})$ can be established if $Z(\mathbf{x})$ is positive definite, as now defined. In general, suppose we have an $n \times n$ matrix $F(\mathbf{x})$

$$\begin{bmatrix} F_{11}(\mathbf{x}) & \cdots & F_{1n}(\mathbf{x}) \\ \vdots & & \vdots \\ F_{n1}(\mathbf{x}) & \cdots & F_{nn}(\mathbf{x}) \end{bmatrix}$$

and an arbitrary vector $\mathbf{v} = [v_1, \cdots, v_n]$. The matrix $F(\mathbf{x})$ is *positive semidefinite* if

$$\mathbf{v}^T\ F(\mathbf{x})\ \mathbf{v} \geq 0, \qquad \forall\ \mathbf{v} \in R^n$$

where $T$ denotes the transpose. $F(x)$ is *positive definite* if

$$\mathbf{v}^T\ F(\mathbf{x})\ \mathbf{v} > 0, \qquad \forall\ \mathbf{v} \neq 0, \mathbf{v} \in R^n$$

$F(x)$ is strongly positive definite if

$$\mathbf{v}^T\ F(\mathbf{x})\ \mathbf{v} > \alpha||v||^2, \qquad \text{for some } \alpha > 0, \qquad \forall\ \mathbf{v} \in R^n$$

If $\gamma(\mathbf{x})$ is the smallest eigenvalue, which is necessarily real, of the symmetric part of $F(\mathbf{x})$, that is, $\frac{1}{2}[F(\mathbf{x}) + F^T(\mathbf{x})]$, then it follows that:

1. $F(\mathbf{x})$ is positive semidefinite if and only if $\gamma(\mathbf{x}) \geq 0$;
2. $F(\mathbf{x})$ is positive definite if and only if $\gamma(\mathbf{x}) > 0$.

We define the Hessian matrix of the objective function $Z(\mathbf{x})$ as

$$\nabla^2 Z(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 Z(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 Z(\mathbf{x})}{\partial x_1\, \partial x_2} & \cdots & \frac{\partial^2 Z(\mathbf{x})}{\partial x_1\, \partial x_n} \\ \frac{\partial^2 Z(\mathbf{x})}{\partial x_2\, \partial x_1} & \frac{\partial^2 Z(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 Z(\mathbf{x})}{\partial x_2\, \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 Z(\mathbf{x})}{\partial x_n\, \partial x_1} & \frac{\partial^2 Z(\mathbf{x})}{\partial x_n\, \partial x_2} & \cdots & \frac{\partial^2 Z(\mathbf{x})}{\partial x_n^2} \end{bmatrix}$$

To show that a stationary point, $\mathbf{x}^*$, is a local minimum, it is sufficient to demonstrate that $Z(\mathbf{x})$ is positive definite in the vicinity of $\mathbf{x} = \mathbf{x}^*$.

### 3.2.2 Nonlinear Programs with General Constraints

A typical mathematical programming problem is to choose the values of a set of variables, $x_1, x_2, \cdots, x_n$, which minimize an objective function $Z(x_1, x_2, \cdots, x_n)$, subject to certain constraints. Each constraint can be expressed as an inequality of a function $g(x_1, x_2, \cdots, x_n)$ of the variables. The set of possible values of $x_1, x_2, \cdots, x_n$ that comply with all constraints is termed the *feasible region.* A general minimization problem with $M$ constraints can be written as

$$\min_{\mathbf{x}} \quad Z(x_1, x_2, \cdots, x_n) \tag{3.53}$$

s.t.

$$g_1(x_1, x_2, \cdots, x_n) \geq b_1 \tag{3.54}$$

$$g_2(x_1, x_2, \cdots, x_n) \geq b_2 \tag{3.55}$$

$$\vdots$$

$$g_m(x_1, x_2, \cdots, x_n) \geq b_m \tag{3.56}$$

where $g_i(x_1, x_2, \cdots, x_n) \geq b_i$ denotes the $i$th constraint on the variables. This problem can include any type of constraints. Using vector notation, the above problem can be written in a standard form as follows

$$\min_{\mathbf{x}} \quad Z(\mathbf{x}) \tag{3.57}$$

s.t.

$$\mathbf{g}(\mathbf{x}) \geq \mathbf{b} \tag{3.58}$$

where $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), g_2(\mathbf{x}), \cdots, g_m(\mathbf{x})]$ and $\mathbf{b} = (b_1, b_2, \cdots, b_m)$. If $Z(\mathbf{x})$ is a nonlinear function or the constraints are nonlinear, the above mathematical programming problem is termed a nonlinear programming (NLP) problem. On the other hand, when $Z(\mathbf{x})$ is a linear function and the constraints are linear, the above mathematical programming problem is termed a linear programming (LP) problem.

A generalization of the Lagrangian method can be used to derive the first-order necessary conditions for general mathematical programs. The Lagrangian for this program is given by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = Z(\mathbf{x}) + \sum_j \mu_j \, [b_j - g_j(\mathbf{x})] \tag{3.59}$$

where $\mu_j$ is the Lagrange multiplier for constraint $j$ and $\mu_j \geq 0$.

The stationary point of the Lagrangian of a convex function is not at a minimum or a maximum of $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$, but rather at a saddle point of the Lagrangian. In fact, $\mathcal{L}(\mathbf{x}^*, \mathbf{u}^*)$ minimizes $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$ with respect to $\mathbf{x}$ and maximizes it with respect to $\boldsymbol{\mu}$. This condition can be stated as

$$\mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}) \leq \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*) \leq \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}^*) \tag{3.60}$$

In order to write the first-order necessary conditions of Lagrangian, note that its minimization is unconstrained with respect to $\mathbf{x}$. Maximization with respect to $\boldsymbol{\mu}$, however, is subject to the nonnegativity constraints. Therefore, the saddle point of $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$ satisfies the following set of first-order necessary conditions:

$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)}{\partial x_i} = 0 \qquad \forall\, i \tag{3.61}$$

$$\mu_j \frac{\partial \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)}{\partial \mu_j} = 0 \qquad \text{and} \qquad \frac{\partial \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)}{\partial \mu_j} \leq 0 \qquad \forall\, j \tag{3.62}$$

In addition, it is required that $\mu_j \geq 0$, $\forall\, j$. Condition (3.61) simply states that the gradient vanishes at the stationary point. Condition (3.62) describes the condition for a maximum of a function subject to nonnegativity constraints. Since $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$ has to be maximized with respect to $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_m)$, the maximum of $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$ with respect to $\mu_j$ can occur either at a point where $\partial \mathcal{L}(\mathbf{x}, \boldsymbol{\mu})/\partial \mu_j = 0$ or at a point where $\mu_j = 0$. In the later case, it must be true that $\partial\mathcal{L}(\mathbf{x},\boldsymbol{\mu})/\partial\mu_j \leq 0$. This observation gives rise to conditions (3.62). Thus, conditions (3.61)-(3.62) can be written explicitly as

$$\frac{\partial Z(\mathbf{x}^*)}{\partial x_i} - \sum_j \mu_j^* \frac{\partial g_j(\mathbf{x}^*)}{\partial x_i} = 0 \qquad \forall\, j \tag{3.63}$$

$$b_j - g_j(\mathbf{x}^*) \leq 0 \qquad \forall\, j \tag{3.64}$$

$$\mu_j\, [b_j - g_j(\mathbf{x}^*)] = 0 \qquad \forall\, j \tag{3.65}$$

$$\mu_j^* \geq 0 \qquad \forall\, j \tag{3.66}$$

These necessary conditions (3.63)-(3.66) are called the Karush-Kuhn-Tucker conditions. They are widely used in the analysis of optimality conditions for mathematical programming problems.

The Lagrangian approach implies that constrained minimization problems can be solved as unconstrained problems by finding the saddle point of the Lagrangian. This point can be found by minimizing the Lagrangian with respect to $\mathbf{x}$ given $\boldsymbol{\mu}$, and then maximizing over all values of $\boldsymbol{\mu}$. The Lagrangian is widely used as an aid in the formulation of first-order necessary conditions.

Note that the functional form of the Lagrangian demonstrates why the Lagrange multipliers, or dual variables, can be interpreted as a measure of the sensitivity of the optimal solution to a constraint relaxation. At the solution point,

$$\mathcal{L}(\mathbf{x}^*,\boldsymbol{\mu}^*) = Z(\mathbf{x}^*) + \sum_j \mu_j\, [b_j - g_j(\mathbf{x}^*)] \tag{3.67}$$

At this point $\mathcal{L}(\mathbf{x}^*,\boldsymbol{\mu}^*) = Z(\mathbf{x}^*)$. If constraint $k$ is relaxed by a small amount, $\Delta b_k$, and $b_k$ in (3.67) is replaced by $b_k - \Delta b_k$, the new minimum value of $\mathcal{L}(\mathbf{x}^*,\boldsymbol{\mu}^*)$ will approximately equal the old value (before the relaxation) minus $\mu_k \Delta b_k$. Thus a relaxation of constraint $k$ by $\Delta b_k$ improves the optimal value of the objective function by approximately $\mu_k \Delta b_k$.

A special case of mathematical programming is linear programming (LP). In a linear minimization problem, both the objective function and the constraints are linear functions of the decision variables. A linear program can be written as

$$\min_{\mathbf{x}} \quad Z(\mathbf{x}) = \sum_{i=1}^{I} c_i\, x_i \tag{3.68}$$

s.t.

$$\sum_{i=1}^{I} a_{ij}\, x_i \geq b_j \qquad \forall\, j \tag{3.69}$$

where $c_i$ and $a_{ij}$ are constants. In some cases, multiple minima may exist because the strict convexity conditions of nonlinear programming do not apply to linear programs. Some of these minima, however, will always be at the intersection of several constraints or at the corners of the feasible region. Therefore, the minimum value of $Z(\mathbf{x})$ can still be determined if only the corners of the feasible region are considered.

### 3.2.3 Discrete Optimal Control and Nonlinear Programs

Section 3.1 described the relationship between discrete and continuous optimal control problems. In this section, we explore the relationship between discrete optimal control and nonlinear programming problems. We recognize both as being multivariate extremization problems subject to various equality and inequality constraints. There are several approaches to reduce discrete optimal control problems to nonlinear programming problems. In the following, we only present a simple transformation to show the analytical relationship. As an example, we discuss how we can transform the following discrete optimal control problem into an NLP. The discrete optimal control problem is as follows:

$$\min_{\mathbf{x},\mathbf{u}} \quad J = \sum_{k=1}^{K} F[\mathbf{x}(k), \mathbf{u}(k), k] \; + \; S[\mathbf{x}(K+1), K+1] \tag{3.70}$$

s.t.

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{f}[\mathbf{x}(k), \mathbf{u}(k), k] \tag{3.71}$$

$$K \text{ and } \; \mathbf{x}(1) = \mathbf{x}_1 \text{ fixed}; \quad \mathbf{x}(K+1) \text{ free} \tag{3.72}$$

We define a vector $\mathbf{y}$ as follows:

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{u}(1) \\ &\vdots \\ \mathbf{y}_K &= \mathbf{u}(K) \\ \mathbf{y}_{K+1} &= \mathbf{x}(1) \\ &\vdots \\ \mathbf{y}_{2K+1} &= \mathbf{x}(K+1) \end{aligned}$$

Note that $\mathbf{y}_i$, $i = 1, \cdots, 2K+1$, must satisfy all corresponding constraints and boundary conditions of the discrete optimal control problem. The objective function can be rewritten as

$$\min_{\mathbf{y}} \quad Z(\mathbf{y}) = \sum_{i=1}^{K} F[\mathbf{y}_{K+i}, \mathbf{y}(i), i] \; + \; S[\mathbf{y}(2K+1), 2K+1] \tag{3.73}$$

The state equation and boundary conditions can be rewritten as

$$\mathbf{y}_{K+i+1} = \mathbf{y}_{K+i} + \mathbf{f}[\mathbf{y}_{K+i}, \mathbf{y}_i, i] \qquad \forall\, i = 1, 2, \cdots, K \tag{3.74}$$

$$\mathbf{y}_{K+1} \text{ given} \tag{3.75}$$

Thus, the discrete optimal control problem can be written in the form of the nonlinear programming problem:

$$\min_{\mathbf{y}} \quad Z(\mathbf{y}) \tag{3.76}$$

s.t.

$$\mathbf{g}(\mathbf{y}) \leq \mathbf{0} \tag{3.77}$$

If the discrete optimal control problem was generated from a continuous optimal control problem, clearly the solution to the associated nonlinear programming problem will only approximate the solution to the original continuous control problem. From the discussion of the discrete optimal control problem and the nonlinear programming problem, we can see that if control variables $\mathbf{u}(k)$ and state variables $\mathbf{x}(k)$ are considered as individual variables and the state equations are transformed into corresponding NLP constraints, then the discrete optimal control problem is in fact a kind of nonlinear programming problem. Thus, solution algorithms for nonlinear programming problems can be used for solving discrete optimal control problems.

### 3.2.4 Nonlinear Programs with Linear Equality and Nonnegativity Constraints

Constrained minimization problems with nonnegativity and equality constraints are of special interest in the study of static network equilibrium problems. They also have an important role in dynamic network equilibrium problems, especially when continuous dynamic problems are transformed into discrete forms. The general form of these problems is

$$\min_{\mathbf{x}} \; Z(\mathbf{x}) \tag{3.78}$$

s.t.

$$\sum_i a_{ij}\, x_i = b_j \qquad j = 1, 2, \cdots, J \tag{3.79}$$

$$x_i \geq 0 \qquad i = 1, 2, \cdots, I \tag{3.80}$$

To find the first-order necessary conditions for a minimum for such a problem, write the Lagrangian with respect to the equality constraints:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = Z(\mathbf{x}) + \sum_j \mu_j \,[b_j - \sum_i a_{ij}\, x_i] \tag{3.81}$$

Then the stationary point of this Lagrangian has to be determined, subject to the constraint

$$x_i \geq 0 \qquad i = 1, 2, \cdots, I \tag{3.82}$$

Unlike the case discussed previously, this problem includes nonnegativity constraints. Consequently, the stationary point of the NLP program has to be determined by the following conditions:

$$x_i^* \frac{\partial \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)}{\partial x_i} = 0 \qquad \text{and} \qquad \frac{\partial \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)}{\partial x_i} \geq 0 \qquad \forall\, i \tag{3.83}$$

$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \boldsymbol{\mu}^*)}{\partial \mu_j} = 0 \qquad \forall\, i \tag{3.84}$$

$$x_i \geq 0 \qquad \forall\, i \tag{3.85}$$

Equation (3.84) requires, simply, that the derivatives of $\mathcal{L}(\mathbf{x}, \boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$ vanish at the minimum. No other condition is necessary since the values of the $\boldsymbol{\mu}$ are not constrained to be nonnegative. This condition, then, is identical to the original constraints. The first-order conditions for the NLP programs with linear equality and nonnegativity constraints can be written explicitly as follows:

$$x_i^* [\frac{\partial Z(\mathbf{x}^*)}{\partial x_i} - \sum_j \mu_j^* a_{ij}\,] = 0 \qquad \forall\, i \tag{3.86}$$

$$\frac{\partial Z(\mathbf{x}^*)}{\partial x_i} - \sum_j u_j^* a_{ij} \geq 0 \qquad \forall\, i \tag{3.87}$$

$$\sum_i a_{ij}\, x_i^* = b_j \qquad \forall\, j \tag{3.88}$$

$$x_i^* \geq 0 \qquad \forall\, i \tag{3.89}$$

The same conditions can be derived also by applying the Karush-Kuhn-Tucker conditions (3.63)-(3.66) directly.

### 3.2.5 Bilevel Mathematical Programs

In this section, we focus on bilevel mathematical programming problems, which have a correspondence to bilevel optimal control problems. We consider two decision-makers or competitive players who must find vectors $\mathbf{x}$ and $\mathbf{y}$, respectively, to minimize their individual objective functions $Z_1(\mathbf{x}, \mathbf{y})$ and $Z_2(\mathbf{x}, \mathbf{y})$.

It is assumed that player 1 has the first choice and selects $\mathbf{x}$, followed by player 2 who selects $\mathbf{y}$. In the most general situation, the value of each objective function depends upon the decisions made by both players. In addition, the choice made by player 1 may affect the set of feasible strategies open to player 2, implying the existence of jointly dependent constraints.

The bilevel nonlinear programming problem (BNP) can be stated as:

$$\min_{\mathbf{x}} Z_1(\mathbf{x}, \mathbf{y}) \tag{3.90}$$

where $\mathbf{y}$ solves

$$\min_{\mathbf{y}} Z_2(\mathbf{x}, \mathbf{y}) \tag{3.91}$$

s.t.

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \tag{3.92}$$

Alternately, this problem can be viewed as a two-person, nonzero-sum game with perfect information where the order of play is specified at the outset and the players' strategy sets are no longer assumed to be disjoint. As a consequence, the moves available to player 2 (the follower) depend on the actions of player 1 (the leader). This interpretation is in accord with the definition of the Stackelberg game which is discussed in Section 2.5. The leader problem is termed the upper level problem and the follower is termed the lower level problem.

This problem can be considered a generalization of a mathematical program where the constraint region is determined implicitly by the lower level optimization problem. An alternate representation of the BNP may be derived by converting the bilevel program into a standard one-level mathematical program. This can be achieved by appending the follower's Karush-Kuhn-Tucker conditions to the leader's constraint set.

The Karush-Kuhn-Tucker conditions for the lower level problem are:

$$\nabla_{\mathbf{y}} Z_2(\mathbf{x}, \mathbf{y}) + \boldsymbol{\lambda}\, \nabla_{\mathbf{y}} \mathbf{g}(\mathbf{x}, \mathbf{y}) = \mathbf{0} \tag{3.93}$$

$$\boldsymbol{\lambda}\, \mathbf{g}(\mathbf{x}, \mathbf{y}) = \mathbf{0} \tag{3.94}$$

$$\boldsymbol{\lambda} \geq \mathbf{0} \tag{3.95}$$

where $\boldsymbol{\lambda}$ is the $m$-dimensional vector of Lagrange multipliers associated with the lower level problem and $\nabla$ is the gradient operator. The above conditions have to be satisfied by the overall bilevel problem.

The reformulated one-level problem is stated as:

$$\min_{\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}} Z_1(\mathbf{x}, \mathbf{y}) \tag{3.96}$$

s.t.

$$\nabla_{\mathbf{y}} Z_2(\mathbf{x},\mathbf{y}) + \boldsymbol{\lambda}\ \nabla_{\mathbf{y}} \mathbf{g}(\mathbf{x},\mathbf{y}) = \mathbf{0} \tag{3.97}$$

$$\boldsymbol{\lambda}\ \mathbf{g}(\mathbf{x},\mathbf{y}) = \mathbf{0} \tag{3.98}$$

$$\mathbf{g}(\mathbf{x},\mathbf{y}) \leq \mathbf{0} \tag{3.99}$$

$$\boldsymbol{\lambda} \geq \mathbf{0} \tag{3.100}$$

By imposing certain regularity conditions, it can be shown that the solution to the above one-level problem is also the solution to the BNP when the follower's problem is convex.

Now we derive the first-order necessary conditions for the one-level program. The Lagrangian for the equivalent one-level program is

$$\begin{aligned}\mathcal{L}(\mathbf{x},\mathbf{y},\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\omega}) &= Z_1(\mathbf{x},\mathbf{y}) + \boldsymbol{\mu}\left[\nabla_{\mathbf{y}} Z_2(\mathbf{x},\mathbf{y}) + \boldsymbol{\lambda}\ \nabla_{\mathbf{y}} \mathbf{g}(\mathbf{x},\mathbf{y})\right] \\ &+ \boldsymbol{\nu}\left[\boldsymbol{\lambda}\ \mathbf{g}(\mathbf{x},\mathbf{y})\right] + \boldsymbol{\omega}\ \mathbf{g}(\mathbf{x},\mathbf{y})\end{aligned} \tag{3.101}$$

where $\boldsymbol{\mu}$, $\boldsymbol{\nu}$, $\boldsymbol{\omega}$ are Lagrange multipliers, and $\boldsymbol{\omega} \geq \mathbf{0}$. Let $Z_1$, $Z_2$, and $\mathbf{g}$ be once continuously differentiable. Then the necessary conditions for the bilevel program are

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \mathbf{0} &= \nabla_{\mathbf{x}} Z_1(\mathbf{x},\mathbf{y}) + \boldsymbol{\mu}\left[\nabla_{\mathbf{xy}} Z_2(\mathbf{x},\mathbf{y}) + \boldsymbol{\lambda}\ \nabla_{\mathbf{xy}} \mathbf{g}(\mathbf{x},\mathbf{y})\right] \\ &+ \left[\boldsymbol{\nu}\ \boldsymbol{\lambda} + \boldsymbol{\omega}\right] \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x},\mathbf{y})\end{aligned} \tag{3.102}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{y}} = \mathbf{0} &= \nabla_{\mathbf{y}} Z_1(\mathbf{x},\mathbf{y}) + \boldsymbol{\mu}\left[\nabla^2_{\mathbf{y}} Z_2(\mathbf{x},\mathbf{y}) + \boldsymbol{\lambda}\ \nabla^2_{\mathbf{y}} \mathbf{g}(\mathbf{x},\mathbf{y})\right] \\ &+ \left[\boldsymbol{\nu}\ \boldsymbol{\lambda} + \boldsymbol{\omega}\right] \nabla_{\mathbf{y}} \mathbf{g}(\mathbf{x},\mathbf{y})\end{aligned} \tag{3.103}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}} = \mathbf{0} = \nabla_{\mathbf{y}} Z_2(\mathbf{x},\mathbf{y}) + \boldsymbol{\lambda}\ \nabla_{\mathbf{y}} \mathbf{g}(\mathbf{x},\mathbf{y}) \tag{3.104}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\nu}} = \mathbf{0} = \boldsymbol{\lambda}\ \mathbf{g}(\mathbf{x},\mathbf{y}) \tag{3.105}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\omega}} = \mathbf{g}(\mathbf{x},\mathbf{y}) \leq \mathbf{0} \tag{3.106}$$

$$\boldsymbol{\omega} \geq \mathbf{0} \tag{3.107}$$

$$\boldsymbol{\lambda} \geq \mathbf{0} \tag{3.108}$$

In general, the analysis and interpretation of the above necessary conditions are very difficult for most dynamic transportation network problems. Conceptually, the bilevel programming framework can be extended to more than two levels, which is also termed the hierarchical or multilevel programming problem.

## 3.3 Variational Inequality Problems

The variational inequality problem is a general problem formulation that encompasses a set of mathematical problems, including nonlinear equations, optimization problems, complementarity problems and fixed point problems. Variational inequalities were originally developed as a tool for the study of certain classes of partial differential equations such as those that arise in mechanics. The focus of this section is on variational inequality problems suitable for the analysis of dynamic network equilibrium models.

### 3.3.1 Definitions for Variational Inequality Problems

In this section, we present several types of variational inequality problems. First, we discuss the variational inequality for static problems. Here, we are dealing with a vector of decision variables $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ and a vector of cost functions $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_n(\mathbf{x})]$. Define $G$ as a given closed convex set of the decision variables $\mathbf{x}$; $\mathbf{f}$ is a vector of given continuous functions defined on $R^n$. Then, we define the static case as follows.

> **Definition 3.3.1.** *The finite-dimensional variational inequality problem is to determine a vector* $\mathbf{x}^* \in G \subset R^n$*, such that*
>
> $$\mathbf{f}[\mathbf{x}^*] \cdot [\mathbf{x} - \mathbf{x}^*] \geq 0, \qquad \forall\, \mathbf{x} \in G \tag{3.109}$$

In geometric terms, variational inequality (3.109) states that $\mathbf{f}(\mathbf{x}^*)$ is orthogonal to the feasible set $G$ at the point $\mathbf{x}^*$.

Now, we discuss the variational inequality for dynamic problems. Unlike the static problem, we are concerned with a vector of control variables $\mathbf{u}(t) = [u_1(t), u_2(t), \cdots, u_m(t)]$ and their dynamic processes

$$\dot{\mathbf{x}}(t) = \mathbf{h}[\mathbf{x}(t), \mathbf{u}(t)]$$

where the state variables $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_n(t)]$ and state equations $\mathbf{h} = [h_1(t), h_2(t), \cdots, h_n(t)]$. Associated with the dynamic processes, there is a vector of cost functions $\mathbf{F}(t) = [F_1(t), F_2(t), \cdots, F_m(t)]$. Each element of the cost function vector is a function of state and control variables, i.e.,

$$F_i(t) = F_i[\mathbf{x}(t), \mathbf{u}(t)] \qquad i = 1, 2, \cdots, m$$

Since the state variables $\mathbf{x}(t)$ can be determined by the state equations when the control variables $\mathbf{u}(t)$ are given, the vector of cost functions can be further simplified as $\mathbf{F}(t) = \mathbf{F}[\mathbf{u}(t)]$. Define $G(t)$ as a given closed convex set of the control variables $\mathbf{u}(t)$. We assume $\mathbf{F}(t)$ is a set of given continuous functions from $G(t)$ to $R^n(t)$. Then, we give the following definition of the dynamic variational inequality problem.

> **Definition 3.3.2.** *The finite-dimensional variational inequality problem is to determine a control vector* $\mathbf{u}^*(t) \in G(t) \subset R^n(t)$*, such that*

$$\mathbf{F}[\mathbf{u}^*(t)] \cdot [\mathbf{u}(t) - \mathbf{u}^*(t)] \geq 0, \qquad \forall\, \mathbf{u}(t) \in G(t) \tag{3.110}$$

This definition facilitates the formulation of dynamic network equilibrium problems as variational inequality problems. However, the following definition is also useful where continuous time problems need to be transformed to discrete time problems and comparisons need to be made with static problems.

**Definition 3.3.3.** *The finite-dimensional variational inequality problem is to determine a control vector* $\mathbf{u}^*(t) \in G(t) \subset R^n(t)$*, such that*

$$\int_0^T \mathbf{F}^T[\mathbf{u}^*(t)] \cdot [\mathbf{u}(t) - \mathbf{u}^*(t)]\, dt \geq 0, \qquad \forall\, \mathbf{u}(t) \in G(t) \tag{3.111}$$

Many dynamic transportation network equilibrium problems can be formulated as systems of equations. The systems of equations can be written as

$$\mathbf{F}[\mathbf{u}^*(t), \mathbf{x}^*(t)] = 0 \tag{3.112}$$

This problem can be regarded as a special case of a variational inequality.

We next discuss several problems related to the variational inequality. These problems include optimization problems and complementarity problems. We first discuss the relationship between an optimization problem and a variational inequality problem. In this discussion, we mainly consider the static case; the analysis can be readily extended to their dynamic counterparts.

### Optimization Problems

A general optimization problem is to maximize or minimize an objective function, and in the case of a constrained problem, subject to a given set of constraints. Both unconstrained and constrained mathematical programming problems can be formulated as variational inequality problems. The following two theorems describe the relationship between an optimization problem and a variational inequality problem.

**Theorem 3.3.1.** *Let* $\mathbf{x}^*$ *be a solution to the minimization problem:*

$$\min_{\mathbf{x}} \quad Z(\mathbf{x}) \tag{3.113}$$

$$\text{s.t.} \quad \mathbf{x} \in G,$$

*where* $Z$ *is continuously differentiable and* $G$ *is closed and convex. Then,* $\mathbf{x}^*$ *is a solution to the variational inequality problem:*

$$\nabla Z^T(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*) \geq 0, \qquad \forall\, \mathbf{x} \in G. \tag{3.114}$$

**Proof:** Denote an auxiliary function $Y(\alpha) = Z[\mathbf{x}^* + \alpha(\mathbf{x} - \mathbf{x}^*)]$, where $\alpha \in [0,1]$ is the decision variable. Note that $Y(\alpha)$ achieves its minimum at $\alpha = 0$, since

$$[\mathbf{x}^* + \alpha(\mathbf{x} - \mathbf{x}^*)]|_{\alpha=0} = \mathbf{x}^*$$

Thus, the derivative $dY/d\alpha$ must be nonnegative within the interval $\alpha \in [0,1]$; i.e.,

$$\frac{dY}{d\alpha} = \nabla Z^T(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*) \geq 0$$

Therefore, we obtain variational inequality (3.114) and $\mathbf{x}^*$ is a solution to (3.114).

**Theorem 3.3.2.** *If $Z(\mathbf{x})$ is a convex function and $\mathbf{x}^*$ is a solution to variational inequality (3.114), then $\mathbf{x}^*$ is a solution to minimization problem (3.113).*

**Proof:** Since $Z(\mathbf{x})$ is convex, we have

$$Z(\mathbf{x}) \geq Z(\mathbf{x}^*) + \nabla Z^T(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*) \geq 0, \qquad \forall\, \mathbf{x} \in G. \tag{3.115}$$

Note that ($\nabla Z^T(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*) \geq 0$) is true because $\mathbf{x}^*$ is a solution to variational inequality (3.114). Therefore, from inequality (3.115) we have

$$Z(\mathbf{x}) \geq Z(\mathbf{x}^*), \forall\, \mathbf{x} \in G \tag{3.116}$$

In other words, $\mathbf{x}^*$ is a minimum of mathematical program (3.113).

Note that the above two theorems apply to both constrained and unconstrained optimization problems, because the feasible set $G$ may or may not be constrained. In addition, the variational inequality problem can be formulated as an optimization problem when a certain symmetry condition holds. More specifically, if the variational inequality formulation of the optimality conditions is characterized by a function with a symmetric Jacobian, then the solution of the optimality conditions and the solution of a particular optimization problem are the same. We have the following theorem to depict this relationship.

**Theorem 3.3.3.** *If $\mathbf{F}(\mathbf{x})$ is a set of continuously differentiable functions on $G$ and the Jacobian matrix*

$$\nabla F(\mathbf{x}) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial F_n}{\partial x_1} & \frac{\partial F_n}{\partial x_2} & \cdots & \frac{\partial F_n}{\partial x_n} \end{bmatrix}$$

*is symmetric and positive semidefinite. Then, there is a real-valued convex function $Z(\mathbf{x})$ satisfying*

$$\nabla Z(\mathbf{x}) = \mathbf{F}(\mathbf{x})$$

*where $\nabla Z(\mathbf{x})$ is the gradient vector of function $Z(\mathbf{x})$. The solution $\mathbf{x}^*$ to variational inequality (3.114) is also the solution to the optimization problem:*

$$\min_{\mathbf{x}} \quad Z(\mathbf{x}) \tag{3.117}$$

$$\text{s.t.} \quad \mathbf{x} \in G.$$

**Proof:** Since $\nabla\mathbf{F}(\mathbf{x})$ is symmetric and positive semidefinite, it follows from Green's Theorem that

$$Z(\mathbf{x}) = \int \mathbf{F}(\mathbf{x})dx$$

where $\int$ is a line integral. Since $\mathbf{x}^*$ is the solution to variational inequality (3.114), using Theorem 3.3.2, we conclude that $\mathbf{x}^*$ is also the solution to optimization problem (3.117).

By Theorem 3.3.3, a variational inequality problem can be reformulated as a convex optimization problem only when the cost function $\mathbf{F}(\mathbf{x})$ is symmetric and positive semidefinite. Thus, the variational inequality problem encompasses the optimization problem. Therefore, the variational inequality is the more general problem in that it can also accomodate a function $\mathbf{F}(\mathbf{x})$ with an asymmetric Jacobian matrix. Historically, some static transportation network equilibrium problems which cannot be formulated as optimization problems were formulated successfully and solved as variational inequality problems. A similar observation can be made for dynamic transportation network equilibrium problems. Only certain types of dynamic problems can be formulated as optimal control problems. Many dynamic problems do not satisfy the symmetry condition and have to be formulated as variational inequalities directly.

We note that although the above theorems are only proven for static problems, they apply to dynamic problems as well. To avoid repetition, we omit the proofs. Note that in dynamic problems, the corresponding optimization problems have to be replaced by optimal control problems.

### Complementarity Problems

Complementarity problems are defined on the nonnegative orthant. The nonlinear complementarity problem is a system of equations and inequalities stated as follows:

Find $\mathbf{x}^* \geq 0$ such that

$$\mathbf{f}(\mathbf{x}^*) \geq \mathbf{0} \quad \text{and} \quad \mathbf{f}^T(\mathbf{x}^*) \cdot \mathbf{x}^* = 0 \tag{3.118}$$

When $\mathbf{f}(\mathbf{x})$ is a set of linear functions, that is, $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{B}$, where $\mathbf{A}$ is an $n \times n$ matrix and $\mathbf{B}$ an $n \times 1$ vector, problem (3.118) is a linear complementarity problem. In general, the complementarity problem is a special case of the variational inequality problem. The relationship between the complementarity problem and the variational inequality problem is as follows.

**Theorem 3.3.4.** *The variational inequality (3.109) and the complementarity problem (3.118) have precisely the same solutions, if any solutions exist.*

**Proof:** First, we need to prove that if $\mathbf{x}^*$ satisfies variational inequality (3.109), then it also satisfies complementarity problem (3.118). Denote $\mathbf{e}_i$ as an $n$-dimensional vector with 1 in the $i$th location and 0 elsewhere, i.e.,

$$\mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Substituting $\mathbf{x} = \mathbf{x}^* + \mathbf{e}_i$ into variational inequality (3.109), we have $f_i(\mathbf{x}^*) \geq 0$. We can choose any $\mathbf{e}_i$ with 1 at any $i$-th location so that each component of $\mathbf{f}_i(\mathbf{x}^*)$ is nonnegative. Thus, $\mathbf{f}(\mathbf{x}^*) \geq \mathbf{0}$.

Now substituting $\mathbf{x} = 2\mathbf{x}^*$ into variational inequality (3.109), we obtain

$$\mathbf{f}(\mathbf{x}^*) \cdot (\mathbf{x}^*) \geq 0. \tag{3.119}$$

Then, substituting $\mathbf{x} = \mathbf{0}$ into variational inequality (3.109), we obtain

$$\mathbf{f}(\mathbf{x}^*) \cdot (-\mathbf{x}^*) \geq 0. \tag{3.120}$$

Equations (3.119) and (3.120) together imply that $\mathbf{f}(\mathbf{x}^*) \cdot \mathbf{x}^* = 0$. Thus, we obtain complementarity problem (3.118).

Second, if $\mathbf{x}^*$ satisfies complementarity problem (3.118), then

$$\mathbf{f}(\mathbf{x}^*) \cdot \mathbf{x}^* = 0 \tag{3.121}$$

Since we can find any feasible $\mathbf{x} \geq \mathbf{0}$ and $\mathbf{f}(\mathbf{x}^*) \geq \mathbf{0}$, we obtain

$$\mathbf{f}(\mathbf{x}^*) \cdot \mathbf{x} \geq 0 \tag{3.122}$$

Subtracting equation (3.121) from equation (3.122), we obtain the variational inequality

$$\mathbf{f}(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*) \geq 0 \tag{3.123}$$

### 3.3.2 Existence and Uniqueness Conditions

Next we discuss the existence and uniqueness of the variational inequality problem for static problems. The conclusions will also apply to variational inequalities for the dynamic problems. Existence of a solution to a variational inequality problem follows from continuity of the function $\mathbf{f}$ entering the variational inequality, provided that the feasible set $G$ is compact. In general, we have the following existence theorem.

**Theorem 3.3.5.** *If $G$ is a compact convex set and $\mathbf{f}(\mathbf{x})$ is continuous on $G$, then the variational inequality problem has at least one solution $\mathbf{x}^*$.*

The proof of this theorem requires the use of Brouwer's Fixed Point Theorem and is not given here (Nagurney, 1993). Qualitative properties of existence and uniqueness are easily obtained under certain monotonicity conditions. First, we present the following definitions.

**Definition 3.3.4.** *A vector of functions $\mathbf{f}(\mathbf{x})$ is monotone on $G$ if*

$$[\mathbf{f}(\mathbf{x}^1) - \mathbf{f}(\mathbf{x}^2)] \cdot (\mathbf{x}^1 - \mathbf{x}^2) \geq 0 \qquad \forall\, \mathbf{x}^1, \mathbf{x}^2 \in G \tag{3.124}$$

*where $\mathbf{x}^1$ and $\mathbf{x}^2$ are any two points on $G$.*

**Definition 3.3.5.** *A vector of functions $\mathbf{f}(\mathbf{x})$ is* strictly *monotone on $G$ if*

$$[\mathbf{f}(\mathbf{x}^1) - \mathbf{f}(\mathbf{x}^2)] \cdot (\mathbf{x}^1 - \mathbf{x}^2) > 0 \quad \forall\, \mathbf{x}^1, \mathbf{x}^2 \in G, \quad \mathbf{x}^1 \neq \mathbf{x}^2 \tag{3.125}$$

**Definition 3.3.6.** *A vector of functions $\mathbf{f}(\mathbf{x})$ is* strongly *monotone on $G$ if for some $\alpha > 0$*

$$[\mathbf{f}(\mathbf{x}^1) - \mathbf{f}(\mathbf{x}^2)]^T \cdot (\mathbf{x}^1 - \mathbf{x}^2) \geq \alpha\, ||\mathbf{x}^1 - \mathbf{x}^2||^2, \quad \forall\, \mathbf{x}^1, \mathbf{x}^2 \in G \tag{3.126}$$

Assume that $\mathbf{f}(\mathbf{x})$ is continuously differentiable on $G$ and $\nabla\mathbf{f}(\mathbf{x})$ is strongly positive definite. Then $\mathbf{f}(\mathbf{x})$ is strongly monotone. Then, we have the following theorem for uniqueness.

**Theorem 3.3.6.** *Suppose that $\mathbf{f}(\mathbf{x})$ is strictly monotone on $G$. Then, the solution is unique, if one exists.*

In the following, we present some easier methods for checking the monotonicity of functions.

**Theorem 3.3.7.** *Suppose that $\mathbf{f}(\mathbf{x})$ is continuously differentiable on $G$ and the Jacobian matrix*

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

*is positive semidefinite (or positive definite), then $\mathbf{f}(\mathbf{x})$ is monotone (or strictly monotone).*

**Theorem 3.3.8.** *Assume that* $\mathbf{f}$ *is continuously differentiable at some* $\bar{\mathbf{x}}$. *Then* $\mathbf{f}(x)$ *is locally strictly (or strongly) monotone at* $\bar{\mathbf{x}}$ *if* $\nabla\mathbf{f}(\bar{\mathbf{x}})$ *is positive definite (or strongly positive definite), that is,*

$$\mathbf{v}^T\mathbf{f}(\bar{\mathbf{x}})\mathbf{v} > 0, \qquad \forall\, \mathbf{v} \in R^n, \mathbf{v} \neq 0 \tag{3.127}$$

$$\mathbf{v}^T\nabla\mathbf{f}(\bar{\mathbf{x}})\mathbf{v} \geq \alpha||\mathbf{v}||^2, \qquad \text{for some} \quad \alpha > 0, \quad \forall\, \mathbf{v} \in R^n \tag{3.128}$$

*where* $\mathbf{v}$ *is an arbitrary vector with components of real values.*

Given the above two theorems for monotonicity, we have the following theorem for uniqueness.

**Theorem 3.3.9.** *Assume that* $\mathbf{f}(\mathbf{x})$ *is continuously differentiable on* $G$ *and that* $\nabla\mathbf{f}(\mathbf{x})$ *is strongly positive definite, then* $\mathbf{f}(\mathbf{x})$ *is strongly monotone.*

The following theorem provides a condition under which both existence and uniqueness of the solution to the variational inequality problem are guaranteed. No assumption on the compactness of the feasible set $G$ is made. This is important for very complicated dynamic problems when convexity of the feasible set is difficult to prove.

**Theorem 3.3.10.** *If* $\mathbf{f}(\mathbf{x})$ *is strongly monotone, then there exists precisely one solution* $\mathbf{x}^*$ *to the variational inequality (3.109).*

The proof of existence follows from the fact that strong monotonicity implies coercivity, whereas uniqueness follows from the fact that strong monotonicity implies strict monotonicity. In conclusion, in the case of an unbounded feasible set $G$, strong monotonicity of the function $\mathbf{f}$ guarantees both existence and uniqueness. If $G$ is compact, then existence is guaranteed if $\mathbf{f}$ is continuous, and only the strict monotonicity condition is needed for uniqueness to be guaranteed. The first conclusion is important for some complicated dynamic problems.

## 3.4 Solution Algorithms for Mathematical Programs and Variational Inequalities

In this section, we present several solution algorithms for mathematical programming problems. These algorithms are also suitable for solving corresponding discrete optimal control problems. To simplify our presentation, we concentrate on the most widely used algorithms for dynamic transportation network equilibrium problems. For other algorithms, we refer readers to mathematical programming texts. We first present the interval reduction algorithm for a one-dimensional minimization problem. Then, the Frank-Wolfe algorithm is discussed. Finally, we present the relaxation or diagonalization algorithm for variational inequality problems. Although the algorithms are presented for static mathematical programming problems, they are extended to dynamic problems in subsequent chapters.

### 3.4.1 One Dimentional Minimization

In this section, we consider the minimization of a nonlinear function $Z(x)$ of a single variable $x$. It is well known from elementary calculus that the necessary condition for a differentiable function in one variable, $Z(x)$, to have a minimum at $x = x^*$ is that the derivative of $Z(x)$ evaluated at $x^*$ equals zero. In other words,

$$\frac{d\,Z(x^*)}{dx} = 0 \tag{3.129}$$

This is a first-order condition for a minimum. If there is a minimum at $x^*$, this condition must hold. To prove that this stationary point is a minimum, we need to prove that it is a global minimum. In other words, the value of $Z(x)$ is lower than $Z(x)$ at any other $x$.

A sufficient condition for a stationary point to be a global minimum is that the function is strictly convex . The strict convexity condition is equivalent to requiring that the second derivative of $Z(x)$ be positive, that is

$$\frac{d^2 Z(x^*)}{dx^2} > 0 \tag{3.130}$$

We now discuss the methods for determining $x^*$. It is assumed that $x$ lies within some finite interval $[a, b]$ and $Z(x)$ is continuous and uniquely defined everywhere in that interval. These requirements ensure the existence of a finite minimum of $Z(x)$ for some $x$ in the interval. For simplicity, we assume $Z(x)$ is ditonic (has one extreme point) over the interval $[a, b]$, implying that it has only a single, unique minimum in that interval.

The study of one-dimensional optimization methods is important mainly because such an optimization or line search is often a part of an algorithm designed to find a minimum of multivariate functions. We discuss two basic methods, the bisection and golden section methods. Both methods use the interval reduction approach. The interval reduction approach involves iterative procedures in which each iteration is focused on a current interval. The current interval is a portion of $[a, b]$, denoted as $[a^{(n)}, b^{(n)}]$. This interval must be determined to include the minimal point $x^*$. At each iteration, this interval is examined and divided into two parts: the part in which the minimum cannot lie and the current interval for the next iteration. The part in which the minimum cannot lie is discarded and the procedure is repeated for the new current interval. The procedure starts by designating $[a, b]$ as the first current interval, i.e., $a^{(0)} = a$ and $b^{(0)} = b$. The interval is then reduced at each successive iteration until a small enough current interval (smaller than a prespecified value) for $x^*$ is obtained. Basically we have two interval reduction methods which differ from each other only in the rules used to examine the current interval and to decide which portion of it can be discarded.

### Bisection Method

The bisection method exploits the fact that the function is monotonic on each side of the minimum. In other words, the derivative of the function, $dZ(x)/dx$, is negative for $x < x^*$ and positive for $x > x^*$. The algorithm computes the derivative of $Z(x)$ at the midpoint of the current interval, $[a^{(n)}, b^{(n)}]$, at iteration $n$. Denote this point as $x^{(n)}$. If $dZ(x^{(n)})/dx < 0$, then $x^* > x^{(n)}$. Thus, the interval $[a^{(n)}, x^{(n)}]$ can be discarded. The next current interval will be $[x^{(n)}, b^{(n)}]$. If $dZ(x^{(n)})/dx > 0$, then $x^* < x^{(n)}$. Thus, the interval $[x^{(n)}, b^{(n)}]$ can be discarded. The next current interval will be $[a^{(n)}, x^{(n)}]$. A prespecified convergence criterion $|a^{(n)} - b^{(n)}| \leq \epsilon$ ($\epsilon$ is a very small value) can be used to terminate the procedure and the middle point of the remaining interval is taken as the estimate of $x^*$.

### Golden Section Method

The golden section method is based on a comparison of the values of $Z(x)$ at two points, $x_1^{(n)}$ and $x_2^{(n)}$, where $x_1^{(n)} < x_2^{(n)}$. The two points are within the current interval, $[a^{(n)}, b^{(n)}]$, at iteration $n$. The two interior points are determined by using a reduction ratio of 0.618 or precisely $(\sqrt{5} - 1)/2$. The interior points are selected so that $x_1^{(n)}$ is 0.618 of the current interval length to the left of $b^{(n)}$ and $x_2^{(n)}$ is 0.618 of the current interval length to the right of $a^{(n)}$. Since the value of 0.618 is known as the golden section, this one-dimensional search method is termed the golden section method.

At iteration $n$, if $Z(x_1^{(n)}) > Z(x_2^{(n)})$, the optimum must lie to the right of $Z(x_1^{(n)})$ because the function is ditonic. Thus, the interval $[a^{(n)}, x_1^{(n)}]$ can be discarded. The new interval for iteration $(n+1)$ is $[a^{(n+1)}, b^{(n+1)}]$, where $a^{(n+1)} = x_1^{(n)}$ and $b^{(n+1)} = b^{(n)}$. If $Z(x_1^{(n)}) < Z(x_2^{(n)})$, the optimum must lie to the left of $Z(x_2^{(n)})$ because the function is ditonic. Thus, the interval $[x_2^{(n)}, b^{(n)}]$ can be discarded. The new interval for iteration $(n+1)$ is $[a^{(n+1)}, b^{(n+1)}]$, where $a^{(n+1)} = a^{(n)}$ and $b^{(n+1)} = x_2^{(n)}$.

The interval reduction process continues with two new interior points $x_1^{(n+1)}$ and $x_2^{(n+1)}$ for iteration $(n + 1)$. Thus, at each iteration, the golden section procedure makes use of one of the interior points from the last interval; only one new point needs to be calculated at each iteration.

The bisection method has a faster convergence speed than the golden section method. However, it requires computing the derivatives of $Z(x)$ at each iteration. On the other hand, the golden section method involves computing the function $Z(x)$ itself. Thus, the bisection algorithm can be preferable to the golden section method if calculating the derivatives is easier than calculating the function itself.

### 3.4.2 Frank-Wolfe Algorithm

The Frank-Wolfe (F-W) algorithm was originally suggested by Frank and Wolfe (1956) as a procedure for solving quadratic programming problems with linear constraints. It is also known as the convex combination algorithm. This method is extensively used in determining equilibrium flows in static transportation network problems. In this book, it is extended to solve the dynamic transportation network equilibrium problems.

We consider a convex minimization program with linear constraints:

$$\min_{\mathbf{x}} \quad Z(\mathbf{x}) \tag{3.131}$$

s.t.

$$\sum_i a_{ij}\, x_i \geq b_j \qquad \forall\, j \tag{3.132}$$

where $a_{ij}$ and $b_j$ are constant coefficients ($i = 1, \cdots, I; j = 1, \cdots, J$). The algorithm is basically a feasible descent direction method. At iteration ($n$+1), it generates a point $\mathbf{x}^{(n+1)} = (x_1^{(n+1)}, \cdots, x_I^{(n+1)})$ from $\mathbf{x}^{(n)} = (x_1^{(n)}, \cdots, x_I^{(n)})$ so that $Z(\mathbf{x}^{(n+1)}) < Z(\mathbf{x}^{(n)})$. Thus, the essence of this algorithm is the calculation of $\mathbf{x}^{(n+1)}$ from $\mathbf{x}^{(n)}$. The algorithmic step can be written in a standard form as

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \alpha^{(n)}\, \mathbf{d}^{(n)} \tag{3.133}$$

where $\mathbf{d}^{(n)} = (d_1^{(n+1)}, \cdots, d_I^{(n+1)})$ is a descent direction vector and $\alpha^{(n)}$ is a nonnegative scalar known as the step size or move size. This equation means that at each point $\mathbf{x}^{(n)}$, a direction $\mathbf{d}^{(n)}$ is identified along which the function is decreasing. Then, the step size $\alpha^{(n)}$ determines how far the next point $\mathbf{x}^{(n+1)}$ will be along the direction $\mathbf{d}^{(n)}$.

The F-W method selects the feasible descent direction not only based on how steep each candidate direction is in the vicinity of $\mathbf{x}^{(n)}$, but also according to how far it is possible to move along this direction. It chooses a direction based on the product of the rate of descent in the vicinity of $\mathbf{x}^{(n)}$ in a given direction and the length of the feasible region in that direction. This product is the drop or the possible reduction in the objective function value which can be achieved by moving in this direction. Thus, the algorithm uses the direction that maximizes the drop.

To find a descent direction, the algorithm checks the entire feasible region for an auxiliary feasible solution, $\mathbf{y}^{(n)} = [y_1^{(n)}, \cdots, y_1^{(n)}]$, such that the direction from $\mathbf{x}^{(n)}$ to $\mathbf{y}^{(n)}$ provides the maximum drop. In seeking the feasible direction, the bounding of the move size does not require a separate step of the algorithm. The bounding is accomplished as an integral part of choosing the decent direction. The direction from $\mathbf{x}^{(n)}$ to any feasible solution, $\mathbf{y}$, is the vector $(\mathbf{y} - \mathbf{x}^{(n)})$ (or the unit vector $(\mathbf{y} - \mathbf{x}^{(n)})/||\mathbf{y} - \mathbf{x}^{(n)}||$). The slope of $Z(\mathbf{x}^{(n)})$ in the direction of $(\mathbf{y} - \mathbf{x}^{(n)})$ is given by the projection of the opposite gradient $[-\nabla Z(\mathbf{x}^{(n)})]$ in this direction, i.e.,

$$-\nabla Z^T(\mathbf{x}^{(n)}) \cdot \frac{(\mathbf{y} - \mathbf{x}^{(n)})}{||\mathbf{y} - \mathbf{x}^{(n)}||} \tag{3.134}$$

The drop in the objective function in the direction $(\mathbf{y} - \mathbf{x}^{(n)})$ is obtained by multiplying this slope by the distance from $\mathbf{x}^{(n)}$ to $\mathbf{y}$, $||\mathbf{y} - \mathbf{x}^{(n)}||$, i.e.,

$$-\nabla Z^T(\mathbf{x}^{(n)}) \cdot (\mathbf{y} - \mathbf{x}^{(n)}) \tag{3.135}$$

This term has to be maximized (in $\mathbf{y}$) subject to the feasibility of $\mathbf{y}$. Alternatively, the term can be multiplied by (-1) and minimized. It follows that

$$\min \ \nabla Z^T(\mathbf{x}^{(n)}) \cdot (\mathbf{y} - \mathbf{x}^{(n)}) = \sum_i \frac{\partial Z(\mathbf{x}^{(n)})}{\partial x_i}(y_i - x_i^{(n)}) \tag{3.136}$$

s.t.

$$\sum_i a_{ij}\, y_i \geq b_j \qquad \forall\, j \tag{3.137}$$

where constraint set (3.137) is equivalent to the original constraint set (3.132) by replacing $\mathbf{x}$ with $\mathbf{y}$. Thus, finding the descent direction amounts to solving a linear program, in which $y_i$ is the decision variable. Note that $\nabla Z(\mathbf{x}^{(n)})$ is constant at $\mathbf{x}^{(n)}$ and the term $\nabla Z^T(\mathbf{x}^{(n)}) \cdot (\mathbf{x}^{(n)})$ can be discarded from the objective function. Thus, the linearized problem can be simplified as

$$\min \ \ F^{(n)}(\mathbf{y}) = \nabla Z^T(\mathbf{x}^{(n)}) \cdot \mathbf{y} = \sum_i \frac{\partial Z(\mathbf{x}^{(n)})}{\partial x_i}\, y_i \tag{3.138}$$

s.t.

$$\sum_i a_{ij}\, y_i \geq b_j \qquad \forall\, j \tag{3.139}$$

The objective function coefficients are $\partial Z(\mathbf{x}^{(n)})/\partial x_1$, $\partial Z(\mathbf{x}^{(n)})/\partial x_2$, $\cdots$, $\partial Z(\mathbf{x}^{(n)})/\partial x_I$. These coefficients are the derivatives of the original objective function at $\mathbf{x}^{(n)}$, which are known at iteration $n$. The decision variables of program (3.136)-(3.137) are $\mathbf{y}^{(n)} = (y_1^{(n)}, y_2^{(n)}, \cdots, y_I^{(n)})$ and the decent direction is the vector pointing from $\mathbf{x}^{(n)}$ to $\mathbf{y}^{(n)}$, i.e., $\mathbf{d}^{(n)} = (\mathbf{y}^{(n)} - \mathbf{x}^{(n)})$, or in an expanded form, $d_i^{(n)} = y_i^{(n)} - x_i^{(n)}$, $\forall\, i$. Once the decent direction is known, other algorithmic steps involve the determination of the move size and a convergence test.

As in many other descent methods, the move size in the direction of $d^{(n)}$ equals the distance to the point along $\mathbf{d}^{(n)}$ which minimizes $Z(\mathbf{x})$. The F-W method does not require a special step to bracket the search for an optimal move size in order to maintain feasibility. The new solution, $\mathbf{x}^{(n+1)}$, must lie between $\mathbf{x}^{(n)}$ and $\mathbf{y}^{(n)}$. Because $\mathbf{y}^{(n)}$ is a solution of the linearized problem, it naturally lies at the boundary of the feasible region. In other words, the search for a descent direction automatically generates a bound for a line search by accounting for all constraints when the descent direction is determined. Since the search interval is bracketed, the bisection or golden section method can be used to find the step size $\alpha$ by solving the minimization of $Z(\mathbf{x})$ along $\mathbf{d}^{(n)} = (\mathbf{y}^{(n)} - \mathbf{x}^{(n)})$. It follows that

$$\min_{0\leq\alpha\leq 1} Z[\mathbf{x}^{(n)} + \alpha(\mathbf{y}^{(n)} - \mathbf{x}^{(n)})] \tag{3.140}$$

Once the optimal solution of this line search, $\alpha^{(n)}$, is found, the next point can be generated using the following formula

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \alpha^{(n)}(\mathbf{y}^{(n)} - \mathbf{x}^{(n)}) \tag{3.141}$$

Note that equation (3.141) can be written as

$$\mathbf{x}^{(n)} = (1 - \alpha^{(n)})\mathbf{x}^{(n)} + \alpha^{(n)}\mathbf{y}^{(n)}$$

The new solution is thus a convex combination (or a weighted average) of $\mathbf{x}^{(n)}$ and $\mathbf{y}^{(n)}$. An appropriate convergence criterion is to check the lower bound of the objective function at each iteration. By convexity,

$$Z(\mathbf{x}^*) \geq Z(\mathbf{x}^{(n)}) + \nabla Z(\mathbf{x}^{(n)}) \cdot (\mathbf{x}^* - \mathbf{x}^{(n)}) \tag{3.142}$$

Thus, the value of the linearized objective function yields a lower bound at $Z(\mathbf{x}^{(n)})$.

$$LB(\mathbf{x}^{(n)}) = Z(\mathbf{x}^{(n)}) + \nabla Z(\mathbf{x}^{(n)}) \cdot (\mathbf{y} - \mathbf{x}^{(n)}) \tag{3.143}$$

An appropriate convergence criterion is

$$\nabla Z(\mathbf{x}^{(n)}) \cdot (\mathbf{y} - \mathbf{x}^{(n)})/LB(x^{(n)}) < \epsilon \tag{3.144}$$

The numerator of equation (3.144) is sometimes called the *gap*.

The F-W algorithm can be summarized as follows:

**Step 0: Initialization.**
Find a feasible solution $\mathbf{x}^{(0)}$. Set iteration counter $n := 0$.

**Step 1: Direction Finding.**
Find $\mathbf{y}^{(n)}$ that solves the linear program (3.138)-(3.139).

**Step 2: Step Size Determination.**
Find $\alpha^{(n)}$ that solves

$$\min_{0\leq\alpha\leq 1} Z[\mathbf{x}^{(n)} + \alpha(\mathbf{y}^{(n)} - \mathbf{x}^{(n)})] \tag{3.145}$$

**Step 3: Move.**
Set $\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \alpha^{(n)}(\mathbf{y}^{(n)} - \mathbf{x}^{(n)})$.

**Step 4: Convergence Test.**
If $\nabla Z(\mathbf{x}^{(n)}) \cdot (\mathbf{y} - \mathbf{x}^{(n)})/LB(x^{(n)}) < \epsilon$, stop. Otherwise, let $n := n + 1$ and go to Step 1.

The algorithm converges in a finite number of iterations. Since the F-W algorithm involves a minimization of a linear program as part of the direction-finding step, it is useful only in cases in which this linear program can be solved relatively easily. It is also useful when algorithms which are generally more efficient than the F-W method can not be utilized due to the size of the problem. Minimization problems for dynamic transportation networks possess both properties: they include a large number of variables and the linear program associated with the direction finding step can be efficiently solved.

### 3.4.3 Relaxation Algorithm

In this section, we present an iterative method for the solution of the variational inequality problem. For a static problem, the variational inequality problem is to determine a vector $\mathbf{x}^* \in G \subset R^n$, such that

$$\mathbf{f}^T(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*) \geq 0, \qquad \forall\, \mathbf{x} \in G \tag{3.146}$$

where $\mathbf{f}$ is a vector of continuous functions. Assume that there exists a vector of auxiliary smooth functions $\mathbf{g}(\mathbf{x}, \mathbf{y})$ on $G \times G$ and $\mathbf{g} \in R^n$. The function $\mathbf{g}(\mathbf{x}, \mathbf{y})$ has the following properties:

1. $\mathbf{g}(\mathbf{x}, \mathbf{x}) = \mathbf{f}(\mathbf{x})$ for all $\mathbf{x}$ on $G$;
2. for every fixed $\mathbf{x}$, $\mathbf{y} \in G$, the Jacobian matrix $\nabla\mathbf{g}(\mathbf{x}, \mathbf{y})$ is symmetric and positive definite.

In other words, the decision variables $\mathbf{x}$ for the functions $\mathbf{f}$ are partitioned into two groups $\mathbf{x}$ and $\mathbf{y}$. Since the Jacobian matrix $\nabla\mathbf{g}(\mathbf{x}, \mathbf{y})$ is symmetric and positive definite, the line integral $\int \mathbf{g}(\mathbf{x}, \mathbf{y})\, dx$ creates a new function $Z(\mathbf{x}, \mathbf{y})$ on $G \times G$ and $Z \in R$. For any fixed $\mathbf{y} \in G$, function $Z(\mathbf{x}, \mathbf{y})$ is strictly convex and

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) = \nabla_{\mathbf{x}} Z(\mathbf{x}, \mathbf{y}) \tag{3.147}$$

At each iteration $n$, we solve the following variational inequality subproblem:

$$\mathbf{g}(\mathbf{x}^{(n)}, \mathbf{x}^{(n-1)}) \cdot (\mathbf{x} - \mathbf{x}^{(n)}) \geq 0, \qquad \forall\, \mathbf{x} \in G \tag{3.148}$$

or an equivalent mathematical programming problem:

$$\min_{\mathbf{x} \in G} \quad Z(\mathbf{x}, \mathbf{x}^{(n-1)}) \tag{3.149}$$

for which a unique solution $\mathbf{x}^{(n)}$ exists. The solution to program (3.149) may be computed using any appropriate mathematical programming algorithm. For transportation equilibrium network problems, the F-W algorithm is most widely used for such purpose. We note that a variational inequality subproblem can be constructed in various ways, some easier to solve than others.

If the sequence of solution $\mathbf{x}^{(n)}$ is convergent, i.e., $\mathbf{x}^{(n)} \rightarrow \mathbf{x}^*$ as $n \rightarrow \infty$, variational inequality subproblem (3.148) yields

$$\mathbf{f}^T(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*) = \mathbf{g}^T(\mathbf{x}^*, \mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*) \geq 0, \qquad \forall\, \mathbf{x} \in G \tag{3.150}$$

Thus, $\mathbf{x}^*$ is a solution to variational inequality problem (3.146). The relaxation method is stated as follows:

**Step 0: Initialization.**
Find a set of feasible decision variables $\mathbf{x}^{(n)}$. Set $n := 0$.

**Step 1: Relaxation.**
Solve the mathematical programming subproblem:

$$\min_{\mathbf{x} \in G} \quad Z(\mathbf{x}^{(n)}, \mathbf{x}^{(n-1)}) \tag{3.151}$$

obtaining solution $\mathbf{x}^{(n)}$.

**Step 2: Convergence Test.**
If $|\mathbf{x}^{(n)} - \mathbf{x}^{(n-1)}| \leq \epsilon$, for a prespecified small value $\epsilon$, then stop. Otherwise, set $n := n + 1$, and go to Step 1.

We note that in the relaxation method, each component $g_i(\mathbf{x}, \mathbf{y})$ of function $\mathbf{g}(\mathbf{x}, \mathbf{y})$ should correspond to the relaxation of variable $x_i$

$$g_i(\mathbf{x}, \mathbf{y}) = f_i(y_1, \cdots, y_{i-1}, x_i, y_{i+1}, \cdots, y_n) \qquad i = 1, \cdots, n \tag{3.152}$$

This relaxation method is sometimes termed the *diagonalization method* because the Hessian matrix of $Z$ is diagonal, since all cross-link effects have been fixed. The mathematical programming subproblem is also known as the *diagonalized* problem. Therefore, the relaxation method resolves variational inequality (3.146) into a sequence of variational inequality subproblems (3.148) or mathematical programming subproblems (3.149).

## 3.5 Notes

Discrete optimal control problems have been discussed in various optimal control texts. At the elementary level, readers may consult the text by Sage and White (1977) and Bryson and Ho (1975). The necessary conditions for general nonlinear programming problems were provided by Kuhn and Tucker (1951).

The material reviewed for mathematical programming can be found in any standard mathematical programming text, such as Luenberger (1984). Further discussion of reducing discrete optimal control problems to mathematical programming problems can be found in texts by Tabak and Kuo (1971) and Canon et al (1970). For multi-level mathematical programming problems, readers may consult the paper by Bard (1984).

A comprehensive summary of variational inequality problems is provided by Nagurney (1993). The text by Kinderlehrer and Stampacchia (1980) provides an introduction to some variational inequality problems. For the rigorous proofs of the existence and uniqueness for dynamic problems, especially optimal control problems, please refer to Cesari (1983).

The computational algorithms for mathematical programming problems are summarized by Bazaraa et al (1993). The F-W algorithm and its applications in static transportation equilibrium problems are described in detail by Sheffi (1985). The relaxation method is described in more detail in Nagurney (1993).

# Chapter 4

# Network Flow Constraints and Definitions of Travel Times

In this chapter, the constraints for dynamic traffic necessary for a urban transportation network are presented. These constraints include flow conservation for links and nodes, flow propagation, first-in-first-out (FIFO) and oversaturation. Associated with these constraints and different needs for dynamic travel time information, two definitions of travel time are considered.

## 4.1 Flow Conservation Constraints

The multiple origin-destination network flow problem is considered. A traffic network is represented by a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{A}$ is the set of directed links. A node can represent an origin and a destination, as well as an intersection. In the following, the index $r$ will denote an origin node and the index $s$ will denote a destination node.

Consider the fixed time period [0,T], which is long enough to allow all travelers departing during the peak period to complete their trips. Let

| | | |
|---|---|---|
| $x_a(t)$ | = | number of vehicles traveling on link $a$ at time $t$; |
| $x_{ap}^{rs}(t)$ | = | number of vehicles traveling on link $a$ over route $p$ with origin $r$ and destination $s$ at time $t$. |

In the following, all variables with subscript $p$ and superscripts $rs$ denote the variables with route $p$, origin $r$ and destination $s$. It follows that

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t) \qquad \forall a. \tag{4.1}$$

Let $u_a(t)$ denote the inflow rate (vehicles/hour or vehicles/minute) into link $a$ at time $t$ and $v_a(t)$ denote the exit flow rate from link $a$ at time $t$. The inflow $u_a(t)$ and exit flow $v_a(t)$ on link $a$ are both regarded as control variables.

The number of vehicles $x_a(t)$ on link $a$ is defined as the state variable for link $a$. The state equation for link $a$ can then be written as

$$\frac{dx_{ap}^{rs}(t)}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s. \tag{4.2}$$

Figure 4.1 illustrates the flow variables for link $a$.

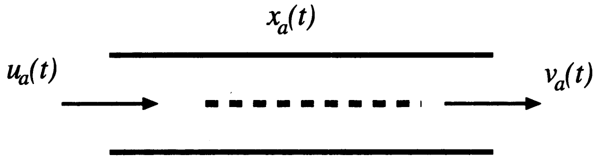



Figure 4.1: Flow Variables for Link $a$

The number of vehicles on link $a$ at an initial time $t = 0$ is assumed to equal 0:

$$x_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s. \tag{4.3}$$

Therefore, the number of vehicles on link $a$ at any time $t$ is given by

$$x_{ap}^{rs}(t) = \int_0^t [u_{ap}^{rs}(\omega) - v_{ap}^{rs}(\omega)]d\omega \qquad \forall a, p, r, s. \tag{4.4}$$

If the number of vehicles on link $a$ at an initial time $t = 0$ is not equal to 0,

$$x_{ap}^{rs}(0) > 0, \qquad \forall a, p, r, s, \tag{4.5}$$

then, the number of vehicles on link $a$ at any time $t$ is given by

$$x_{ap}^{rs}(t) = x_{ap}^{rs}(0) + \int_0^t [u_{ap}^{rs}(\omega) - v_{ap}^{rs}(\omega)]d\omega \qquad \forall a, p, r, s. \tag{4.6}$$

In most models in this book, we consider the case when $x_{ap}^{rs}(0) = 0$. For the case when $x_{ap}^{rs}(0) > 0$, the models and corresponding solution algorithms require modification at time 0 accordingly. Finally, all variables must be nonnegative at all times:

$$x_{ap}^{rs}(t) \geq 0, \qquad u_{ap}^{rs}(t) \geq 0, \qquad v_{ap}^{rs}(t) \geq 0, \qquad \forall a, p, r, s. \tag{4.7}$$

Denote the required instantaneous flows from origin node $r$ to destination node $s$ at time $t$ as $f^{rs}(t)$, which is a given function of time in any dynamic route choice problem. Flow conservation at node $j$ ($j \neq r, s$) for route $p$ between O-D pair $rs$ requires that the flow exiting from the link pointing into node $j$ at time $t$ equals the flow entering the link which leave node $j$ at time $t$. Thus, the flow conservation equations can be expressed as

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j \neq r, s; p, r, s \tag{4.8}$$

where $A(j)$ is the set of links whose tail node is $j$ (after $j$), and $B(j)$ is the set of links whose head node is $j$ (before $j$). Figure 4.1 illustrates flow conservation at node $j$.

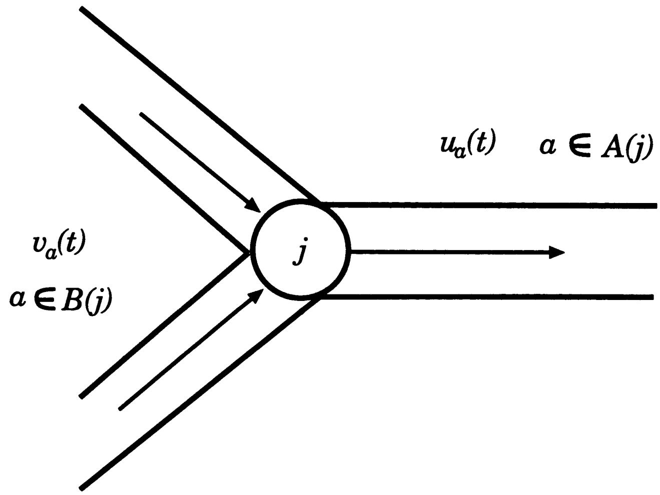



Figure 4.2: Flow Conservation at Node $j$

Conservation of flow at origin node $r$ ($r \neq s$) requires the flow originating at origin $r$ at time $t$ to equal the flow entering the links which leave origin $r$ at time $t$. Thus, the flow conservation equations for the origin nodes can be expressed as

$$\sum_{a \in A(r)} \sum_{p} u_{ap}^{rs}(t) = f^{rs}(t) \qquad \forall r \neq s; s. \tag{4.9}$$

Denote the instantaneous flows arriving at destination node $s$ from origin node $r$ at time $t$ as the control variable $e^{rs}(t)$, and let $e_p^{rs}(t)$ denote these flows over route $p$ at time $t$. Conservation of flow at destination node $s$ ($s \neq r$) requires the flow exiting at destination $s$ at time $t$ to equal the flow exiting the links which lead to destination $s$ at time $t$. Thus, the flow conservation equations for the destination nodes can be expressed as

$$\sum_{a \in B(s)} \sum_{p} v_{ap}^{rs}(t) = e^{rs}(t) \qquad \forall r; s \neq r. \tag{4.10}$$

Denote the cumulative number of vehicles arriving at destination $s$ from origin $r$ over route $p$ by time $t$ as $E_p^{rs}(t)$. It follows that

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r, s \neq r. \tag{4.11}$$

At the initial time $t = 0$,

$$E_p^{rs}(0) = 0, \qquad \forall p, r, s. \tag{4.12}$$

Finally, the variables must be nonnegative at all times:

$$e_p^{rs}(t) \geq 0, \qquad E_p^{rs}(t) \geq 0, \qquad \forall p, r, s. \tag{4.13}$$

## 4.2 Definitions of Dynamic Travel Times

The *instantaneous* link travel time at any time $t$ is defined as the travel time that would be experienced by vehicles traversing a link when prevailing traffic conditions remain unchanged. The instantaneous route travel time at any time $t$ is the sum of the instantaneous link travel times over all links in this route at time $t$. Thus, the instantaneous route travel time would be experienced by a vehicle if prevailing traffic conditions do not vary until the vehicle reaches its destination.

The instantaneous travel time $c_a[x_a(t), u_a(t), v_a(t)]$, or simply $c_a(t)$, over link $a$ is assumed to be dependent on the number of vehicles $x_a(t)$, the inflow $u_a(t)$ and the exit flow $v_a(t)$ on link $a$ at time $t$. We assume the instantaneous travel time $c_a(t)$ on link $a$ is the sum of two components: 1) an instantaneous flow-dependent cruise time $g_{1a}[x_a(t), u_a(t)]$ over link $a$; 2) an instantaneous queuing delay $g_{2a}[x_a(t), v_a(t)]$. It follows that

$$c_a(t) = g_{1a}[x_a(t), u_a(t)] + g_{2a}[x_a(t), v_a(t)]. \tag{4.14}$$

The two components $g_{1a}[x_a(t), u_a(t)]$ and $g_{2a}[x_a(t), v_a(t)]$ are assumed to be nonnegative and differentiable with respect to $x_a(t), u_a(t)$ and $x_a(t), v_a(t)$, respectively.

Consider the flow which originates at node $r$ at time $t$ and is destined for node $s$. There is a set of routes $\{p\}$ between O-D pair $(r, s)$. Define the instantaneous travel time function $\psi_p^{rs}(t)$ for each route $p$ between $(r, s)$ as

$$\psi_p^{rs}(t) = \sum_{a \in rsp} c_a[x_a(t), u_a(t), v_a(t)] \qquad \forall r, s, p; \tag{4.15}$$

the summation is over all links $a$ in route $p$ from origin $r$ to destination $s$.

Define the minimal instantaneous route travel time $\sigma^{rs}(t)$ as the minimal travel time that would be experienced by a vehicle departing from origin $r$ to destination $s$ at time $t$, if prevailing traffic conditions do not vary until the vehicle reaches its destination. If the instantaneous link travel time $c_a(t)$ is determined, the minimal instantaneous O-D travel time $\sigma^{rs}(t)$ can be computed as $\sigma^{rs}(t) = \min_p \psi_p^{rs}(t)$, where $\sigma^{rs}(t)$ is a functional of all link flow variables at time $t$, or $\pi^{rs}(t) = \pi^{rs}[u(t), v(t), x(t), t]$.

The future link travel time or actual link travel time is the travel time over a link actually experienced by vehicles. This time can also be called the projected time. Define $\tau_a(t)$ as the actual travel time over link $a$ for vehicles entering link $a$ at time $t$. $\tau_a(t)$ is assumed to be dependent on the number of vehicles $x_a(t)$, the inflow $u_a(t)$ and the exit flow $v_a(t)$ on link $a$ at time $t$. It follows that

$$\tau_a(t) = \sum_{a \in rsp} \tau_a[x_a(t), u_a(t), v_a(t)] \qquad \forall a \tag{4.16}$$

Similarly, the future route travel time or actual route travel time is the time actually experienced over a route by vehicles. Define $\eta_p^{rs}(t)$ as the travel time *actually* experienced over route $p$ by vehicles departing origin $r$ toward destination $s$ at time $t$. We use a recursive formula to compute the route travel time $\eta_p^{rs}(t)$ for all allowable routes. Assume route $p$ consists of nodes $(r, 1, 2, \cdots, i, \cdots, s)$. Denote $\eta_p^{ri}(t)$ as the travel time *actually* experienced over route $p$ from origin $r$ to node $i$ by vehicles departing origin $r$ at time $t$. Then, a recursive formula for route travel time $\eta_p^{rs}(t)$ is:

$$\eta_p^{ri}(t) = \eta_p^{r(i-1)}(t) + \tau_a[t + \eta_p^{r(i-1)}(t)] \qquad \forall p, r, i; i = 1, 2, \cdots, s;$$

where link $a = (i-1, i)$.

Define $\pi^{rs}(t)$ as the minimal travel time *actually* experienced by motorists departing from origin $r$ to destination $s$ at time $t$. If the actual link travel time $\tau_a(t)$ is determined, the minimal actual O-D travel time $\tau^{rs}(t)$ can be computed as $\pi^{rs}(t) = \min_p \eta_p^{rs}(t)$. $\pi^{rs}(t)$ is a functional of all link flow variables at time $\omega \geq t$, or $\pi^{rs}(t) = \pi^{rs}[u(\omega), v(\omega), x(\omega)|\omega \geq t]$. This functional is neither a state variable nor a control variable, and it is not fixed. This functional is not available in closed form. Nevertheless, it can be evaluated when $u(\omega)$, $v(\omega)$ and $x(\omega)$ are temporarily fixed.

Various dynamic route choice models are formulated in the following chapters based on instantaneous and actual travel times. More discussion on various definitions of travel times and their applications is found in Chapter 16.

## 4.3 Flow Propagation Constraints

It is necessary to ensure that the entering and exiting flows, as well as the vehicles remaining on links, are consistent with the link travel times. We write these constraints using actual link travel times. In static network models, these flow propagation constraints are not necessary because a flow is assumed to propagate instantaneously over its entire journey from its origin to its destination. Vehicles don't remain on a link for a duration of time in order that a queue can form. Thus, queuing phenomena cannot be captured correctly in any static network model.

Flow propagation constraints can be written based on links or nodes. In this book, we suggest using link flow propagation constraints, which can be formulated in two different ways as now described.

### 4.3.1 Type I

Let $U_{ap}^{rs}(t)$ denote the cumulative number of vehicles entering link $a$ on route $p$ with O-D pair $rs$ by time $t$, and $V_{ap}^{rs}(t)$ denote the cumulative number of vehicles leaving link $a$ on route $p$ with O-D pair $rs$ by time $t$. $U_{ap}^{rs}(t)$ and $V_{ap}^{rs}(t)$ are state variables for link $a$. The state equation for link $a$ can then be written as

$$\frac{dU_{ap}^{rs}(t)}{dt} = u_{ap}^{rs}(t) \qquad \forall a, p, r, s. \tag{4.17}$$

$$\frac{dV_{ap}^{rs}(t)}{dt} = v_{ap}^{rs}(t) \qquad \forall a, p, r, s. \tag{4.18}$$

The cumulative numbers of vehicles entering and exiting link $a$ at an initial time $t = 0$ are assumed to equal 0:

$$U_{ap}^{rs}(0) = 0, \qquad V_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s. \tag{4.19}$$

When the number of vehicles on link $a$ at an initial time $t = 0$ equals 0, the number of vehicles at any time $t$ is given by

$$x_{ap}^{rs}(t) = U_{ap}^{rs}(t) - V_{ap}^{rs}(t) \qquad \forall a, p, r, s. \tag{4.20}$$

Figure 4.3 illustrates the relationship of cumulative entering and exiting flows on link $a$. Finally, all variables must be nonnegative at all times:

$$x_{ap}^{rs}(t) \geq 0, \qquad U_{ap}^{rs}(t) \geq 0, \qquad V_{ap}^{rs}(t) \geq 0, \qquad \forall a, p, r, s. \tag{4.21}$$

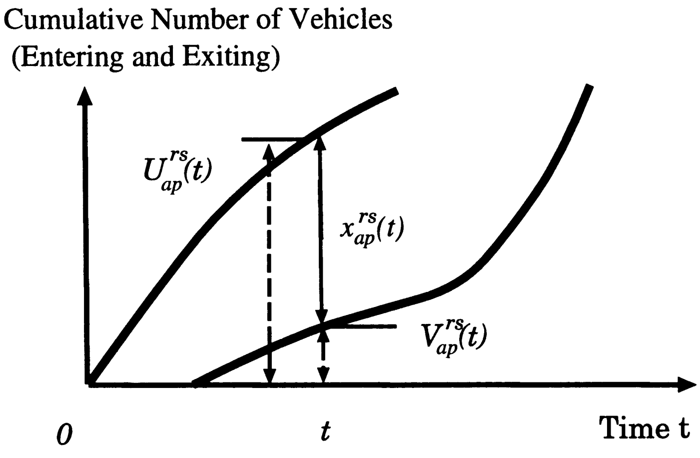



Figure 4.3: Cumulative Entering and Exiting Flows on Link $a$

The node flow conservation equations defined in the last section do not change. For link $a$ on route $p$ between O-D pair $rs$, the total number of vehicles entering link $a$ by time $t$ must have exited link $a$ by time $[t + \tau_a(t)]$. It follows that

$$U_{ap}^{rs}(t) = V_{ap}^{rs}[t + \tau_a(t)] \qquad \forall a, p, r, s. \tag{4.22}$$

Figure 4.4 illustrates the flow propagation on link $a$. The expression of this

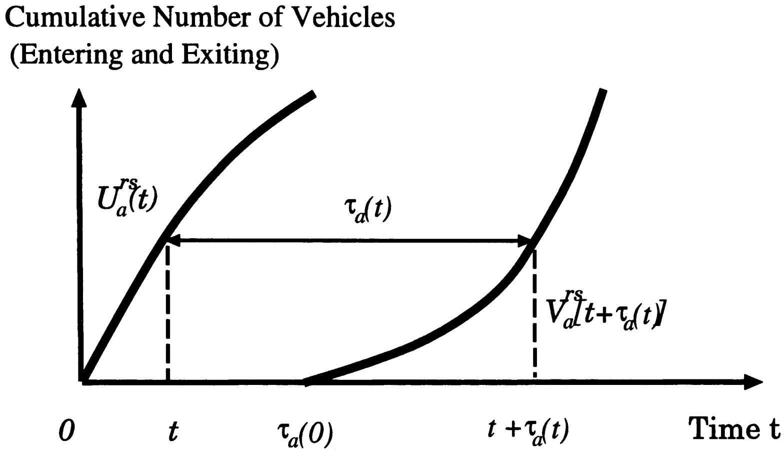



Figure 4.4: Flow Propagation on Link $a$

flow propagation constraint is very simple and its meaning is also intuitive. However, the following flow propagation constraints are easier to handle in optimal control problems and are used in most of our models.

### 4.3.2 Type II

We formulate the second type of link flow propagation constraints as follows. Let $x_{ap}^{rs}(t)$ denote the number of vehicles on link $a$ using route $p$ between O-D pair $rs$ at time $t$. By definition,

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t) \qquad \forall a. \tag{4.23}$$

For any intermediate node $j \neq r$ on route $p$, denote a subroute $\tilde{p}$ as the section of route $p$ from node $j$ to destination $s$. For any link $a \in B(j)$, vehicles on link $a$ using route $p$ at any time $t$ must result in either:

1. extra vehicles on downstream links on subroute $\tilde{p}$ at time $t + \tau_a(t)$, or
2. increased exiting vehicles at the destination at time $t + \tau_a(t)$.

It follows that

$$x_{ap}^{rs}(t) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[t + \tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall a, j, p, r, s; j \neq r; a \in B(j). \qquad (4.24)$$

A detailed discussion of this constraint is given in Chapter 5 following the formulation of an instantaneous DUO route choice model. We note that the flow conservation constraints discussed in the previous section can be used directly when the above link flow propagation constraint is used in the formulation.

In some of our dynamic network models, especially in optimization formulations, we need to write these constraints using *estimates* of actual link travel times. These link travel time estimates must be updated in an iterative procedure known in the transportation science literature as the *relaxation* or *diagonalization technique* (Sheffi, 1985). In this procedure, the travel times over each link $a$, $\bar{\tau}_a(t)$, are estimated for flows entering the link at each time $t$. These functions $\bar{\tau}_a(t)$ are held fixed, and the model is solved. Then, the link travel times corresponding to the solution $x_a(t)$, $u_a(t)$ and $v_a(t)$ obtained are compared to the functions $\bar{\tau}_a(t)$. If the link travel times corresponding to the solution are different from $\bar{\tau}_a(t)$, the values of $\bar{\tau}_a(t)$ are reset to these travel times and the process is repeated. Given the robust nature of the relaxation (diagonalization) technique, we expect that the solution will converge to the DUO solution. This procedure is discussed in detail in Chapter 6 and is justified in Chapter 13 as a standard approach for solving a general variational inequality model. The revised flow propagation constraints are as follows:

$$x_{ap}^{rs}(t) = \sum_{b \in \bar{p}} \{x_{bp}^{rs}[t + \bar{\tau}_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \bar{\tau}_a(t)] - E_p^{rs}(t)\}$$

$$\forall a, j, p, r, s; j \neq r; a \in B(j). \qquad (4.25)$$

These constraints are associated with state variables only. This property is fully exploited in the analysis of the optimality conditions for optimal control models in Chapter 5.

In the later chapters, the above two sets of flow propagation constraints are used in appropriate models so to simplify the formulation and to improve the effectiveness of the corresponding computational algorithms. As we noted above, the flow propagation constraints can be formulated using node-based constraints instead of link-based constraints. Thus, other flow propagation constraints can be formulated. However, these constraints should be decided in conjunction with the corresponding formulation of the dynamic model.

## 4.4 First-In-First-Out Constraints

First-In-First-Out (FIFO) conditions may or may not occur in actuality. However, FIFO should be strictly guaranteed when there is only one lane and no extra space for turning movements at intersections. When there are both left-turn and right-turn lanes on a street link, FIFO may still be violated and the extent of violation will depend on the channelization of lanes at the intersection.

In a continuous time model, the flow propagation constraints state that any inflow $u_a^{rs}(t)$ into link $a$ has to remain on the link for travel time $\tau_a(t)$, regardless of the origin-destination source of this inflow. Thus, the flow propagation constraints imply FIFO constraints in a continuous time model if a rigorous travel time function is used in the flow propagation constraints. In the following, we use a discrete time example to illustrate how the flow propagation constraints should imply FIFO constraints. The example network is shown in Figure 4.5. It is 3-link, 4-node network (links 1-2, 2-3, 3-4) with O-D trips $f^{13}(k) = (0,1,0,1,0,1,0,1,0,1)$ and $f^{14}(k) = (1,0,1,0,1,0,1,0,1,0)$ for time interval $k = 1, \cdots, 10$. Therefore, one vehicle enters link 1-2 in each time interval $k = 1, \cdots, 10$.


Figure 4.5: Example Network

The FIFO condition requires that one vehicle exit link 1-2 and go to node 4 and then another go to node 3, etc. We assume that link 2-3 is blocked due to some incident at $k = 1$. When the first vehicle arrives at node 2 and enters link 2-3, it cannot exit node 3 and enter link 2-4 since the travel time is extremely high and it has to remain link 2-3. Thus, the link travel time on link 2-3 will increase for subsequent vehicles. Our physical flow propagation constraint for link flow and travel time states that the second vehicle entering link 1-2 during period 2 ($k = 2$) cannot arrive at node 3, since link 2-3 is highly congested. This will increase the travel time for vehicles entering link 1-2 during period 3. This process continues and there will be no vehicles exiting node 3. This framework implicitly defines the FIFO constraints.

In the following, we discuss FIFO or its violation, overtaking, in more detail. Overtaking denotes that a late entering vehicle flow propagates faster than an earlier entering vehicle flow and exits earlier than the earlier entering vehicle flow. Overtaking violates the FIFO rule for traffic propagation on links, although it might happen on two-lane links. We consider traffic propagation on link $a$ for two time instants $t$ and $t + \Delta t$ in continuous time problems. Denote the link travel time for flows entering link $a$ at time $t$ as $\tau_a(t)$. The travel time for flows entering link $a$ at time $t + \Delta t$ is $\tau_a(t + \Delta t)$. When the summation of link travel time at time $t + \Delta t$ plus $\Delta t$ is smaller than the link travel time at time $t$, i.e., $\tau_a(t) > \tau_a(t + \Delta t) + \Delta t$, overtaking will occur so that FIFO is violated. Because flows entering link $a$ during time $t + \Delta t$ will exit the link after clock time $t + \Delta t + \tau_a(t + \Delta t)$ which is earlier than the clock time $t + \tau_a(t)$, the exiting time for flows entering link $a$ at time $t$.

Thus, if we require that overtaking should not occur, we must allow the clock time $t + \tau_a(t)$, when flows entering at time $t$ must exit link $a$, to be smaller than the clock time $t + \Delta t + \tau_a(t + \Delta t)$, the exiting time for flows entering link

$a$ at time $t + \Delta t$. It follows that

$$t + \tau_a(t) < t + \Delta t + \tau_a(t + \Delta t) \tag{4.26}$$

Dividing the above equation by $\Delta t$, we obtain

$$1 + \frac{\tau_a(t + \Delta t) - \tau_a(t)}{\Delta t} > 0 \tag{4.27}$$

Taking the limit of the above equation ($\Delta t \to 0$),

$$1 + \dot{\tau}_a(t) > 0 \tag{4.28}$$

or

$$\dot{\tau}_a(t) > -1 \tag{4.29}$$

The above condition must be met to avoid overtaking in any dynamic route choice model using link travel time functions in the flow propagation constraints. If the decreasing rate of travel time on any link $a$ exceeds 1, overtaking will occur.

Even with the flow propagation constraints in a general discrete time model, the FIFO constraints may still be violated when the time interval is quite large. For example, assume link travel times at instants $t$ and $t + \Delta t$ are 2 minutes and 40 seconds, respectively. If $\Delta t = 1$ minute, then

$$\frac{\tau_a(t + \Delta t) - \tau_a(t)}{\Delta t} = \frac{40 - 120}{60} = -1.33 < -1 \tag{4.30}$$

Thus, overtaking does occur in this example. However, condition (4.27) can be satisfied by choosing appropriate link length and time interval length in a discrete time environment.

Some nonconvex conditions can be introduced to guarantee FIFO condition (4.27). However, those constraints greatly increase the complexity of the model and its solution. In the continuous time environment, FIFO condition (4.27) is generally satisfied by realistic travel time functions such as those proposed in Chapter 16. As shown there, our proposed link travel time function for a signal-controlled arterial has three parts:

1. an uncongested cruise time over the first part of the link;

2. a queuing delay at the exit part of the link;

3. a uniform delay due to signal setting.

Since the major delay is the queuing delay which assumes FIFO, the condition, $\dot{\tau}_a(t) > -1$, will not occur if the link travel time function is validated using real traffic flow data.

Our major concern is the overtaking problem in discrete models. In particular, overtaking may occur in the afternoon peak period when traffic flow is declining rapidly together with travel time. Recall that by equation (4.30), overtaking does not occur when

$$\frac{\tau_a(t+\Delta t)-\tau_a(t)}{\Delta t} > -1 \tag{4.31}$$

Thus, the length of time interval $\Delta t$ and the link travel time $\tau_a(t)$, or the link length, determine whether overtaking occurs or not. Note that the link length determines the free flow travel time, which is a major factor affecting equation (4.31) because the queuing delay assumes FIFO on the link. If the time interval $\Delta t$ increases, link travel time $\tau_a(t)$ must be smaller or the link must be shorter so that equation (4.31) holds. If time interval $\Delta t$ decreases, link travel time $\tau_a(t)$ must be higher or the link must be longer in order for equation (4.31) to hold.

The detailed values of time interval $\Delta t$ and link length associated with any specific link travel time function should be determined using numerical experiments in any practical application. In general, overtaking will not occur for most definitions of link lengths (over 100 feet) and time intervals (shorter than 2 minutes) if the free flow speed is assumed to be 50 miles/hour in the experiment. However, as we note in discrete models, no matter how accurate the link traffic dynamics model is, overtaking or a "jump" may still occur when the time interval is too large. On the other hand, for most problems, we are only interested in the aggregate behavior of flows and FIFO is not so important in those situations. Furthermore, we should note that the FIFO assumption itself is also an approximation of reality.

## 4.5 Link Capacity and Oversaturation

There are two basic constraints for link capacity. The first constraint is the maximal number of vehicles on the link. The second constraint is the maximal exit flow rate from the link. More detailed analyses of their impacts on dynamic network models are discussed in Chapter 16.

### 4.5.1 Maximal Number of Vehicles on a Link

Let $l_a$ denote the length of link $a$ and $e_{am}$ denote the maximal traffic density (vehicles/mile). The maximal number of vehicles that link $a$ can accommodate is $l_a e_{am}$. The number of vehicles on link $a$ must be less than or equal to the maximal number of vehicles on the link. It follows that

$$x_a(t) \leq l_a e_{am} \qquad \forall a \tag{4.32}$$

This constraint applies for any dynamic network model. Since this constraint involves only state variables, it could be added to the formulation without incurring any analysis problem in our desired optimality conditions for optimal control models. However, the computational algorithm would need to be revised. When solving a route choice problem including the above constraint, we could add a penalty term to the objective function and solve it as an ordinary dynamic route choice problem.

### 4.5.2 Maximal Exit Flow from a Link

Another constraint concerns the exit flow capacity $v_{am}$ at the exit of a link. It follows that

$$v_a(t) \leq v_{am} \qquad \forall a \tag{4.33}$$

In a network, the exit capacity constraint for an upstream link is also an inflow capacity constraint for downstream links. This constraint can be added directly in the formulation or combined in the link travel time functions. If it is directly added in the optimal control formulation, more analysis of the optimality conditions of dynamic network models is necessary when the exit flow capacity is reached. Moreover, the computational algorithm needs to be revised. A method of combining this constraint with link travel time functions is discussed in Chapter 16.

In most of our dynamic traffic network models, we consider this exit capacity constraint in the travel time functions. Thus, it is not necessary to define an explicit constraint in these network models.

### 4.5.3 Constraints for Spillback

Oversaturation may occur anywhere and during any time interval when traffic demand exceeds capacity. When queues at critical intersections develop upstream, then they cause the so-called spillback problem.

In an oversaturated situation, continuing excess demand relative to supply could transform local oversaturation to regional oversaturation. Thus, in dynamic network models, corresponding constraints should be formulated to reflect this phenomenon. In an advanced control/assignment framework, those constraints should be consistent with each type of traffic control strategy. The two main types of traffic control strategies are: 1) minimize delay and stops; 2) keep traffic moving or maximize productivity (Lieberman, 1993). We leave the control policy for further study in the context of combined dynamic travel choice/signal control models. Here, we only formulate constraints that describe physical spillback queues.

An extreme case occurs if the queue on link $a$ approaches the maximum, i.e. $x_a(t) = l_a e_{am}$. In this situation, the queue may extend into upstream links, which causes a spillback problem. However, if the above constraint as well as upper bound constraints for cumulative O-D departures and associated access delays at origin nodes are added in the assignment problem, the spillback problem can still be handled in our dynamic traffic network models.

When a spillback queue develops upstream towards origins, such as parking lots, queuing delays at origins will occur. This is an additional constraint for dynamic network models. We construct a dummy link $b$ associated with each origin node $r$; see Figure 4.6. The state equation for any origin $r$ is as follows:

$$\frac{dx_{bp}^{s}(t)}{dt} = f_p^{rs}(t) - v_{bp}^{s}(t) \qquad \forall b, p, r, s; b \in rs. \tag{4.34}$$

$$\sum_p f_p^{rs}(t) = f^{rs}(t) \qquad \forall r, s. \tag{4.35}$$

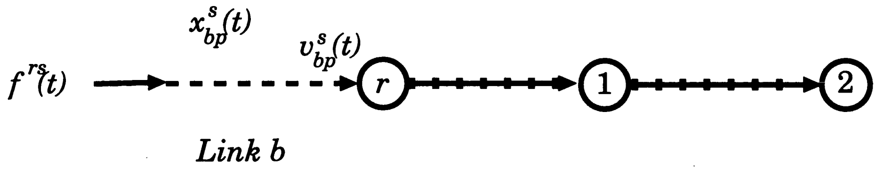



Figure 4.6: Example of Spillback

We assume the number of spillback vehicles at time 0 is 0. It follows that

$$x_{bp}^{s}(0) = 0 \qquad \forall b, p, r, s; b \in rs. \tag{4.36}$$

Thus, the flow conservation equation for origin $r$ should be revised as

$$\sum_{a \in A(r)} u_{ap}^{rs}(t) = v_{bp}^{s}(t) \qquad \forall b, p, r, s; b \in rs. \tag{4.37}$$

Of course, we also need the flow propagation constraint for dummy link $b$. We assume there is no upper bound for the queue length $x_b^s(t)$ since an origin always has enough capacity to accommodate vehicles. The queuing delay at the origin is as follows

$$\tau_b(t) = \tau_b[x_b(t), v_b(t)] \qquad \forall b \in rs \tag{4.38}$$

where

$$x_b(t) = \sum_s \sum_p x_{bp}^{s}(t) \qquad v_b(t) = \sum_s \sum_p v_{bp}^{s}(t)$$

The exact form of this formula should be associated with specific queuing patterns at origin nodes. Figure 4.7 illustrates the queuing delay at origin $r$.

## 4.6 Summary of Notation

| | | |
|---|---|---|
| $x_a(t)$ | = | number of vehicles on link $a$ at time $t$ (main problem variable) * |
| $u_a(t)$ | = | inflow rate into link $a$ at time $t$ (main problem variable) ** |
| $v_a(t)$ | = | exit flow rate from link $a$ at time $t$ (main problem variable) ** |
| $y_a(k)$ | = | number of vehicles on link $a$ at the beginning of time interval $k$ (subproblem variable) |
| $p_a(k)$ | = | inflow into link $a$ during interval $k$ (subproblem variable) |
| $q_a(k)$ | = | exit flow from link $a$ during interval $k$ (subproblem variable) |
| $f^{rs}(t)$ | = | departure flow rate from origin $r$ toward destination $s$ at time $t$ (given) |
| $F^{rs}(t)$ | = | cumulative number of departing vehicles from origin $r$ to destination $s$ by time $t$ (given) |
| $e^{rs}(t)$ | = | arrival flow rate at destination $s$ from origin $r$ at time $t$ ** |
| $E^{rs}(t)$ | = | cumulative number of vehicles arriving at destination $s$ from origin $r$ by time $t$ (main problem variable) * |
| $\bar{E}^{rs}(k)$ | = | cumulative number of vehicles arriving at destination $s$ from origin $r$ by time $t$ (subproblem variable) |
| $c_a(t)$ | = | instantaneous travel time for link $a$ at time $t$ |
| $\psi_p^{rs}(t)$ | = | instantaneous route travel time for route $p$ between $(r, s)$ at time $t$ |
| $\sigma^{rs}(t)$ | = | minimal instantaneous route travel time between $(r, s)$ at time $t$ |
| $A(j)$ | = | set of links whose tail node is $j$ (after $j$) |
| $B(j)$ | = | set of links whose head node is $j$ (before $j$) |
| $\tau_a(t)$ | = | actual travel time over link $a$ for flows entering link $a$ at time $t$ |
| $\bar{\tau}_a(t)$ | = | estimated actual travel time over link $a$ for flows entering link $a$ at time $t$ |
| $\eta_p^{rs}(t)$ | = | actual travel time for route $p$ between $(r, s)$ for flows departing origin $r$ at time $t$ |
| $\pi^{rs}(t)$ | = | minimal actual route travel time between $(r, s)$ for flows departing origin $r$ at time $t$ |

* state variable
** control variable

## 4.7 Notes

Flow conservation constraints are intuitive and essential for any dynamic network model. In contrast with static traffic network models, there are two kinds of travel times in a dynamic traffic network model: instantaneous travel time and actual travel time. These two kinds of travel times reflect the dynamic nature and complexity of dynamic traffic problems.

Flow propagation constraints can also be represented by some alternative constraints which imply flow propagation in flow conservation constraints.

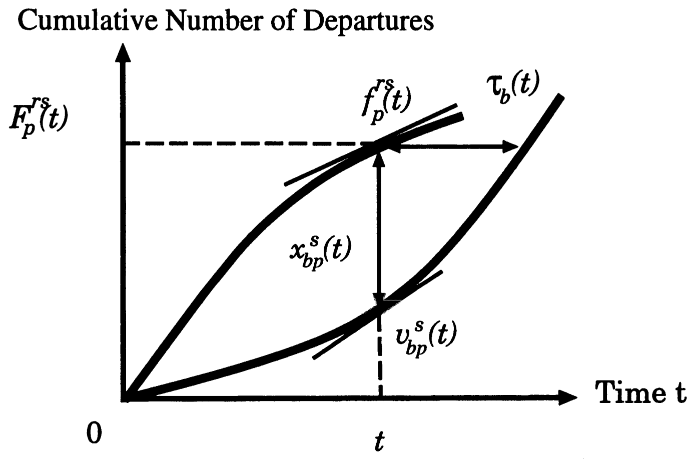



Figure 4.7: Queuing Delay at Origin $r$

Those models include traffic simulation models and the hydrodynamics method. One recent example is the cell transmission model proposed by Daganzo (1993) which assumes traffic flow propagation by cell transmission (flow transfers from cell to cell, where a cell is a small link segment). The advantages of those models are that link flow propagation is directly represented and link travel times are directly provided, while link travel time function is not presented in the model. For more discussion on FIFO, readers may refer to Carey (1992). Detailed discussion on oversaturation and spillback can be found in Lieberman (1993).

# Chapter 5

# Instantaneous Dynamic User-Optimal Route Choice Models

In this chapter, we discuss optimal control models for instantaneous dynamic user-optimal route choice problems. Using a network with two parallel routes, we first present an example to illustrate the instantaneous dynamic user-optimal concept in Section 5.1. The general definition of instantaneous dynamic user-optimal state is given in Section 5.2. Then, we present three instantaneous dynamic user-optimal route choice models. Model 1 is described in Section 5.3. In Section 5.4, the equivalence of Model 1 with DUO route choice is demonstrated by proving the equivalence of the first order necessary conditions of the model with the instantaneous DUO route choice conditions. In Section 5.5, the second DUO model employing a different link travel time function assumption is formulated, and the equivalence of Model 2 with the instantaneous DUO route choice conditions is also demonstrated. In Section 5.6, the third instantaneous DUO model employing a simpler link travel time function assumption is formulated, and its equivalence with the instantaneous DUO conditions is also demonstrated. Finally, we present a discrete-time numerical example indicating that this class of models yields realistic results.

## 5.1 An Example with Two Parallel Routes

Consider a network with one O-D pair and two parallel routes (see Figure 5.1). Assume that there exists only one bottleneck on each route. For simplicity, we assume that the bottleneck is close to the entry point on each route. Then the route travel times are assumed to depend on the inflow rate and have the following simple form:

$$c_1(t) = 10 + 5u_1(t) \tag{5.1}$$

$$c_2(t) = 15 + 3u_2(t) \tag{5.2}$$

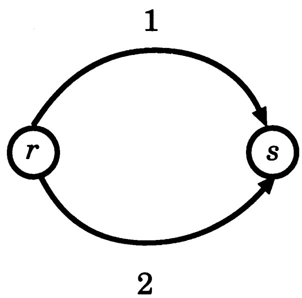



Figure 5.1: Example Network

Suppose there is a departure flow $f(t) = 0.5t$ from origin $r$ to destination $s$ at time $t \in [0, 4]$. Flow conservation at origin $r$ requires that

$$u_1(t) + u_2(t) = 0.5t \tag{5.3}$$

In addition, the inflows must be nonnegative, i.e.,

$$u_1(t) \geq 0, \qquad u_2(t) \geq 0 \tag{5.4}$$

The dynamic user-optimal route choice criterion requires that the departing flow use the minimal travel time route. It follows that

$$c_1(t) = c_2(t) \qquad \text{if } u_1(t) > 0 \text{ and } u_2(t) > 0 \tag{5.5}$$

Solving equations (5.3)-(5.5), we obtain the optimal inflows as follows:

$$u_1^*(t) = \begin{cases} 0.5t & \text{if } t < 2 \\ 0.1875t + 0.625 & \text{if } t \geq 2 \end{cases}$$

$$u_2^*(t) = \begin{cases} 0. & \text{if } t < 2 \\ 0.3125t - 0.625 & \text{if } t \geq 2 \end{cases}$$

The corresponding optimal route travel times are

$$c_1^*(t) = \begin{cases} 10. + 2.5t & \text{if } t < 2 \\ 13.125 + 0.9375t & \text{if } t \geq 2 \end{cases}$$

$$c_2^*(t) = \begin{cases} 15. & \text{if } t < 2 \\ 13.125 + 0.9375t & \text{if } t \geq 2 \end{cases}$$

The minimal route travel time $\sigma^{rs}(t)$ is

$$\sigma^{rs}(t) = \begin{cases} c_1(t) = 10. + 2.5t & \text{if } t < 2 \\ c_1^*(t) = c_2^*(t) = 13.125 + 0.9375t & \text{if } t \geq 2 \end{cases}$$

The optimal inflows and optimal route travel times are illustrated in Figure 5.2. Thus, the dynamic user-optimal route choice conditions for this example can be summarized as follows:

$$c_a(t) \geq \sigma^{rs}(t) \qquad a = 1, 2; \tag{5.6}$$

$$c_a(t) = \sigma^{rs}(t) \qquad \text{if } u_a > 0 \qquad a = 1, 2; \tag{5.7}$$

$$u_1(t) \geq 0, \qquad u_2(t) \geq 0 \tag{5.8}$$

## 5.2 Definition of Instantaneous DUO State

We now consider a general transportation network. We define a decision node for each route $p$ of each O-D pair as any node on the route including the origin. The instantaneous route travel time between a decision node and the destination node is calculated using the currently prevailing link travel times. Significantly, with many current traveler information systems, such as radio broadcasts of traffic conditions, the information provided to travelers on freeways is the estimated *instantaneous* route travel time. Thus, at present many travelers do choose routes based on current or instantaneous travel times.

Consider the flow originating at node $r$ at time $t$ and destined for node $s$. There is a set of routes $\{p\}$ between O-D pair $(r, s)$. In general, for any link $a$ and any O-D pair $rs$, link $a$ is defined as being used at time $t$ if $u_a^{rs}(t) > 0$. Furthermore, a route $p$ between $r$ and $s$ is defined as being used at time $t$ if $u_{ap}^{rs}(t) > 0$, where link $a$ is the first link on route $p$ from $r$ to $s$. The above general definition will be used in Chapter 12 for general variational inequality models for instantaneous DUO route choice problems.

In this chapter, we formulate three alternative optimal control models, each of which equivalent to the instantaneous DUO route choice conditions. Our objective here is to explore the complexity of the problem. As shown in Chapter 12, these optimal control models are specific versions of a more general variational inequality model. Thus, in this chapter we use more restricted definitions of used links and routes as follows. For any link $a$ on any route from origin $r$ to destination $s$, link $a$ is defined as being used at time $t$ if $u_a^{rs}(t) > 0$ and $v_a^{rs}(t) > 0$. Furthermore, a route $p$ between $r$ and $s$ is defined as being used at time $t$ if $u_{ap}^{rs}(t) > 0$ and $v_{ap}^{rs}(t) > 0$ for all links $a$ on route $p$ from $r$ to $s$.

We assume that the time-dependent origin-destination trip pattern is known *a priori*. In other words, the departure times of travelers are given. The instantaneous dynamic user-optimal (DUO) route choice problem is to determine vehicle flows at each instant of time on each link resulting from drivers using minimal-time routes under the currently prevailing travel times. In this chapter, we consider the following dynamic generalization of the conventional static user-optimal state.

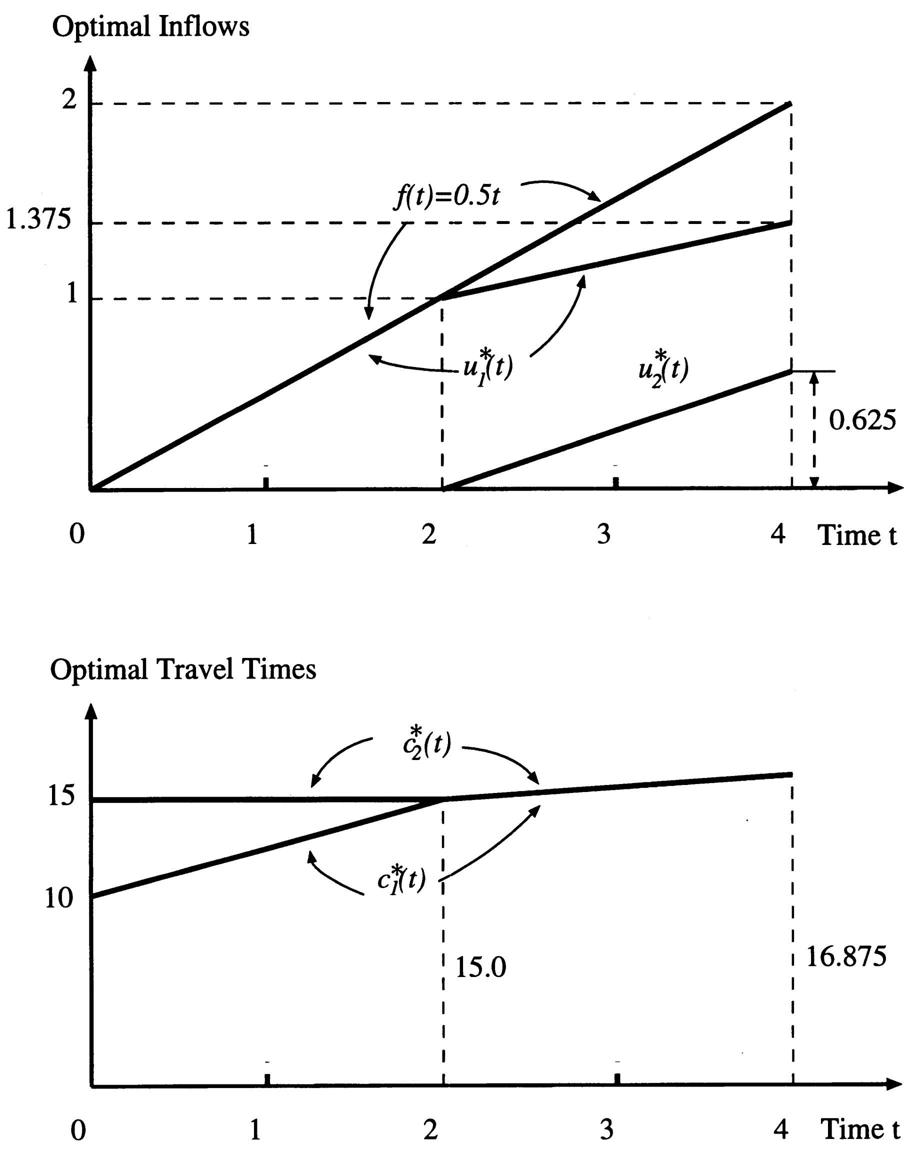



Figure 5.2: Dynamic User-Optimal Inflows and Travel Times

**Link-Time-Based Instantaneous DUO State:** *If, for each O-D pair at each decision node at each instant of time, the instantaneous travel times for all routes that are being used equal the minimal instantaneous route travel time, the dynamic traffic flow over the network is in a link-time-based instantaneous dynamic user-optimal state.*

In Chapter 12, we generalize the above definition to the situation of travel disutility and multi-group travelers instead of travel times for a single group of travelers only.

Although the instantaneous user-optimal travel times for all routes that are being used are equal at each decision node at each instant of time, route flows with the same departure time and the same origin-destination may actually experience somewhat different route travel times. This is because the route time may subsequently change due to changing network traffic conditions, even though at each decision node the flows select the route that is currently the best.

In optimal control theory terminology, the dynamic user-optimal route choice problem is to find the dynamic trajectories of link states and inflow and exit flow control variables, given the time-dependent O-D flow requirements, the network and the link travel time functions. The formulation of the problem in this chapter is based on the underlying choice criterion that each traveler uses the route that minimizes his/her instantaneous travel time when departing from the origin or any intermediate node to his/her destination. This route choice rule implies that in a travel-time-based instantaneous dynamic user-optimal state, the trajectories of the link flow states and inflows and exit flows are such that the instantaneous travel times at each decision node of all used routes connecting any given O-D pair will be identical and not greater than the instantaneous travel times of routes which are not being used.

# 5.3 Instantaneous Route Choice Model 1

## 5.3.1 Model Formulation

Recall from Chapter 4 that the instantaneous travel time $c_a[x_a(t), u_a(t), v_a(t)]$, or simply $c_a(t)$, over link $a$ is assumed to be dependent on the number of vehicles $x_a(t)$, the inflow $u_a(t)$ and the exit flow $v_a(t)$ on link $a$ at time $t$. This instantaneous link time is the travel time that would be incurred if traffic conditions on the link remain unchanged while traversing the link. In Model 1, we assume the instantaneous travel time $c_a(t)$ on link $a$ is the sum of two components: 1) an instantaneous flow-dependent running time $g_{1a}[x_a(t), u_a(t)]$ over link $a$; 2) an instantaneous queuing delay $g_{2a}[x_a(t), v_a(t)]$. It follows that

$$c_a(t) = g_{1a}[x_a(t), u_a(t)] + g_{2a}[x_a(t), v_a(t)]. \tag{5.9}$$

The two components $g_{1a}[x_a(t), u_a(t)]$ and $g_{2a}[x_a(t), v_a(t)]$ of the time-dependent link travel time function $c_a[x_a(t), u_a(t), v_a(t)]$ are assumed to be nonnegative and differentiable with respect to $x_a(t), u_a(t)$ and $x_a(t), v_a(t)$, respectively. The instantaneous travel time function $\psi_p^{rs}(t)$ for each route $p$ between O-D pair $(r, s)$ is

$$\psi_p^{rs}(t) = \sum_{a \in rsp} c_a[x_a(t), u_a(t), v_a(t)] \qquad \forall r, s, p; \tag{5.10}$$

the summation is over all links $a$ in route $p$ from origin $r$ to destination $s$.

Using optimal control theory, the equivalent optimization model of the instantaneous dynamic user-optimal route choice problem (**Model 1**) is formulated as follows.

$$\min_{u,v,x,e,E} \int_0^T \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t), \omega] d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t), \omega] d\omega \right\} dt \tag{5.11}$$

s.t.

Relationship between state and control variables:

$$\frac{dx_{ap}^{rs}(t)}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \tag{5.12}$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r; s \neq r; \tag{5.13}$$

Flow conservation constraints:

$$f^{rs}(t) = \sum_{a \in A(r)} \sum_p u_{ap}^{rs}(t) \qquad \forall r, s; \tag{5.14}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \tag{5.15}$$

$$\sum_{a \in B(s)} \sum_p v_{ap}^{rs}(t) = e^{rs}(t) \qquad \forall r, s; s \neq r; \tag{5.16}$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b \in \bar{p}} \{x_{bp}^{rs}[t + \bar{\tau}_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \bar{\tau}_a(t)] - E_p^{rs}(t)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; \tag{5.17}$$

Definitional constraints:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \tag{5.18}$$

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \qquad \forall a; \tag{5.19}$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \qquad u_{ap}^{rs}(t) \geq 0, \qquad v_{ap}^{rs}(t) \geq 0 \qquad \forall a, p, r, s; \tag{5.20}$$

$$e_p^{rs}(t) \geq 0, \qquad E_p^{rs}(t) \geq 0, \qquad \forall p, r, s; \tag{5.21}$$

Boundary conditions:

$$E_p^{rs}(0) = 0, \qquad \forall p, r, s; \tag{5.22}$$

$$x_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s. \tag{5.23}$$

The objective function terms are similar to the objective function of the well-known static user-optimal (UO) model. The first two constraints (5.12)-(5.13) are state equations for the flow on each link $a$ and for cumulative arrivals at each destination. Equations (5.14)-(5.16) are flow conservation constraints at each node including origins and destinations. The other constraints include flow propagation constraints, definitional constraints, nonnegativity, and boundary conditions. In addition, we need several definitional constraints as follows:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t),$$

$$\sum_{p} x_{ap}^{rs}(t) = x_a^{rs}(t), \qquad \sum_{rs} x_a^{rs}(t) = x_a(t),$$

$$\sum_{p} e_p^{rs}(t) = e^{rs}(t).$$

In summary, the control variables are $u_{ap}^{rs}(t)$, $v_{ap}^{rs}(t)$, and $e_p^{rs}(t)$; the state variables are $x_{ap}^{rs}(t)$ and $E_p^{rs}(t)$. Model 1 can be solved by discretizing time so that it becomes an ordinary nonlinear program and by using a fixed estimate of each link travel time $\bar{\tau}_a(t)$, which is updated in an iterative diagonalization (or relaxation) fashion.

We illustrate the constraints of Model 1 using the example network with three links in Figure 5.3. Assume that O-D flow $f^{14}(t) = 10$ for a short time period $t \in [0, \epsilon]$ ($\epsilon \ll \bar{\tau}_1(t)$) and $f^{14}(t) = 0$ at time $t > \epsilon$. Conservation of flow constraints (5.14) require that $u_1^{14}(t) = 10$ at time $t \in [0, \epsilon]$ (flow enters link 1). Thus, $x_1^{14}(t)$ becomes positive by constraints (5.12) (vehicles are on link 1). When $x_1^{14}(t)$ decreases, constraint (5.12) requires $v_1^{14}(t)$ to become positive (flow exits link 1), and constraint (5.15) for node 2 requires $u_2^{14}(t)$ to equal $v_1^{14}(t)$ (flow enters link 2). Analogously, constraint (5.12) requires $x_2^{14}(t)$ to increase, since $u_2^{14}(t)$ has become positive. Thus, the conservation of flow constraints (together with the flow propagation constraints) requires that flow moves from the origin to the destination, successively entering a link, staying on the link, and then exiting the link.

The flow propagation constraints (5.17) for this network are now stated; since there is only one route, we suppress the route subscript $p$ and begin with link 3.

$$x_3^{14}(t) = E^{14}[t + \bar{\tau}_3(t)] - E^{14}(t) \tag{5.24}$$

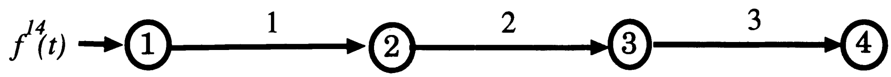



Figure 5.3: Example Network with Inflow

$$x_2^{14}(t) = \{x_3^{14}[t + \bar{\tau}_2(t)] - x_3^{14}(t)\} + \{E^{14}[t + \bar{\tau}_2(t)] - E^{14}(t)\} \qquad (5.25)$$

$$\begin{aligned} x_1^{14}(t) &= \{x_2^{14}[t + \bar{\tau}_1(t)] - x_2^{14}(t)\} + \{x_3^{14}[t + \bar{\tau}_1(t)] - x_3^{14}(t)\} \\ &+ \{E^{14}[t + \bar{\tau}_1(t)] - E^{14}(t)\} \end{aligned} \qquad (5.26)$$

These constraints ensure that flows stay on each link for an amount of time consistent with the link's travel time. For example, (5.22) and (5.24) require that $E^{14}[t+\bar{\tau}_3(t)] = 0$ until $x_3^{14}(t)$ becomes positive. When $x_3^{14}(t)$ does become positive, $E[t + \bar{\tau}_3(t)]$ must also become positive. Only then can $e[t + \bar{\tau}_3(t)]$ become positive (see (5.13) and note that this is route $p = 1$), and thus by (5.16), only then can $v_3^{14}[t + \bar{\tau}_3(t)]$ become positive. This ensures that $v_3^{14}(t)$ cannot become "prematurely" positive; i.e., flows must stay on link 3 for an amount of time consistent with $\bar{\tau}_3(t)$.

Constraints (5.25) require that flow on link 2 at time $t$ must result in either:

1. *added* flow on link 3 at time $t + \bar{\tau}_2(t)$ (in case any existing flow on link 3 at time $t$ has not yet cleared by time $t + \bar{\tau}_2(t)$); or

2. added arrivals at the destination at time $t + \bar{\tau}_2(t)$ (if link 3 is very short, vehicle flows at the *end of link 2* may have traversed link 3 and exited destination node 4 at time $t + \bar{\tau}_2(t)$, while vehicle flows at the *beginning of link 2* may still be on link 3 at time $t + \bar{\tau}_2(t)$).

To illustrate constraint (5.25) further, suppose that vehicle flows first appear on link 2 at time $t_o$: $x_2^{14}(t_o) > 0$, but $x_2^{14}(t) = 0$ for $t < t_o$. Suppose also that there are no vehicle flows on link 3 and none have exited at time $t_o$. Thus, $0 = x_2^{14}(t_o-\epsilon)$ (since vehicles first appear on link 2 at time $t_o$) $= x_3^{14}[t_o-\epsilon+\bar{\tau}_2(t_o-\epsilon)]$ (by (5.25)) for any $\epsilon \leq t_o$. Thus, $x_3^{14}(\cdot)$ must equal zero at all times prior to time $t_o + \bar{\tau}_2(t_o)$, so vehicles must stay on link 2 during the entire period $[t_o, \bar{\tau}_2(t_o)]$. This insures that $x_2^{14}(t)$ propagates consistently with travel time $\bar{\tau}_2(t)$ over link 2 at each instant $t$.

Analogously, constraints (5.26) require that flow on link 1 at time $t$ must result in either:

1. added flows on link 2 or 3 at time $t + \bar{\tau}_1(t)$ (in case flows on links 2 and 3 at time $t$ have not yet cleared by time $t + \bar{\tau}_1(t)$); or

2. added arrivals at the destination at time $t + \bar{\tau}_1(t)$.

Thus, this equation insures that $x_1^{14}(t)$ propagates consistently.

### 5.3.2 Optimality Conditions

The extended Hamiltonian function for the instantaneous DUO route choice program (5.11)-(5.23) is constructed as

$$
\begin{aligned}
\mathcal{H} &= \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t),\omega]d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t),\omega]d\omega \right\} \\
&+ \sum_{rs}\sum_{ap} \lambda_{ap}^{rs}(t)[u_{ap}^{rs}(t) - v_{ap}^{rs}(t)] + \sum_r \sum_{s\neq r}\sum_p \nu_p^{rs}(t)e_p^{rs}(t) \\
&+ \sum_s \sum_{r\neq s} \sigma_r^{rs}(t)[f^{rs}(t) - \sum_{a\in A(r)}\sum_p u_{ap}^{rs}(t)] \\
&+ \sum_{rs}\sum_{j\neq rs}\sum_p \sigma_{jp}^{rs}(t)[\sum_{a\in B(j)} v_{ap}^{rs}(t) - \sum_{a\in A(j)} u_{ap}^{rs}(t)] \\
&+ \sum_r \sum_{s\neq r} \sigma_s^{rs}(t)[\sum_{a\in B(s)}\sum_p v_{ap}^{rs}(t) - e^{rs}(t)] \\
&+ \sum_{rsp}\sum_{j\neq r}\sum_{a\in B(j)} \mu_{ap}^{rs}(t) \left\{ x_{ap}^{rs}(t) + \sum_{b\in\tilde{p}} x_{bp}^{rs}(t) + E_p^{rs}(t) \right. \\
&- \left. \sum_{b\in\tilde{p}} x_{bp}^{rs}[t+\bar{\tau}_a(t)] - E_p^{rs}[t+\bar{\tau}_a(t)] \right\}
\end{aligned}
$$

where $\lambda_{ap}^{rs}(t)$ are Lagrange multipliers associated with the link state equations, $\nu_p^{rs}(t)$ are Lagrange multipliers associated with the destination node state equations, $\sigma_{jp}^{rs}(t)$ are Lagrange multipliers associated with the node flow conservation equations, and $\mu_{ap}^{rs}(t)$ are the Lagrange multipliers associated with the flow propagation equations. For each link $a$ which points from node $l$ to node $m$, the first order necessary conditions of instantaneous DUO route choice program (5.11)-(5.23) include

$$\frac{\partial\mathcal{H}}{\partial u_{ap}^{rs}(t)} = g_{1a}[x_a(t),u_a(t)] + \lambda_{ap}^{rs}(t) - \sigma_{lp}^{rs}(t) \geq 0, \qquad \forall l; a\in A(l), p, r, s, \quad (5.27)$$

and $$u_{ap}^{rs}(t)\frac{\partial\mathcal{H}}{\partial u_{ap}^{rs}(t)} = 0 \qquad \forall a,p,r,s; \quad (5.28)$$

$$\frac{\partial\mathcal{H}}{\partial v_{ap}^{rs}(t)} = g_{2a}[x_a(t),v_a(t)] - \lambda_{ap}^{rs}(t) + \sigma_{mp}^{rs}(t) \geq 0,$$

$$\forall m; a\in B(m), p, r, s, \quad (5.29)$$

and $$v_{ap}^{rs}(t)\frac{\partial\mathcal{H}}{\partial v_{ap}^{rs}(t)} = 0 \qquad \forall a,p,r,s; \quad (5.30)$$

$$\frac{\partial \mathcal{H}}{\partial e_p^{rs}(t)} \geq 0, \qquad \forall p, r, s, \tag{5.31}$$

and

$$e_p^{rs}(t)\frac{\partial \mathcal{H}}{\partial e_p^{rs}(t)} = 0 \qquad \forall p, r, s; \tag{5.32}$$

$$\frac{d\lambda_{ap}^{rs}(t)}{dt} = -\frac{\partial \mathcal{H}}{\partial x_{ap}^{rs}(t)} \qquad \forall a, p, r, s; \tag{5.33}$$

$$\frac{d\nu_p^{rs}(t)}{dt} = -\frac{\partial \mathcal{H}}{\partial E_p^{rs}(t)} \qquad \forall p, r, s; \tag{5.34}$$

$$u_{ap}^{rs}(t) \geq 0, \quad v_{ap}^{rs}(t) \geq 0, \quad x_{ap}^{rs}(t) \geq 0, \qquad \forall a, p, r, s; \tag{5.35}$$

$$e_p^{rs}(t) \geq 0, \quad E_p^{rs}(t) \geq 0, \qquad \forall p, r, s. \tag{5.36}$$

Note that $\sigma_{rp}^{rs}(t) = \sigma_r^{rs}(t)$ when node $l$ equals origin $r$.

### 5.3.3 DUO Equivalence Analysis

Since the objective function of optimal control program (5.11)-(5.23) is convex with respect to the control variables, there is a unique optimal solution. We now show that the set of link states, inflows and exit flows that solves this program also satisfies the travel-time-based instantaneous dynamic user-optimal route choice conditions. This equivalence is demonstrated below by proving that the first order necessary conditions for the optimal control program (5.11)-(5.23) are identical to the instantaneous dynamic user-optimal conditions. The equivalence between the instantaneous DUO route choice conditions and the first order necessary conditions of the optimal control program means that the instantaneous DUO route choice conditions are satisfied at the optimal solution of this program.

Combining equations (5.27)-(5.28) with equations (5.29)-(5.30), the following equations can be derived for each link $a$ which points from node $l$ to node $m$.

$$\frac{\partial \mathcal{H}}{\partial u_{ap}^{rs}(t)} + \frac{\partial \mathcal{H}}{\partial v_{ap}^{rs}(t)} = c_a(t) - \sigma_{lp}^{rs}(t) + \sigma_{mp}^{rs}(t) \geq 0,$$

$$\forall a \in A(l) \cap B(m); p, r, s; \tag{5.37}$$

and

$$u_{ap}^{rs}(t)\frac{\partial \mathcal{H}}{\partial u_{ap}^{rs}(t)} = 0 \qquad \forall a, p, r, s; \tag{5.38}$$

and

$$v_{ap}^{rs}(t)\frac{\partial \mathcal{H}}{\partial v_{ap}^{rs}(t)} = 0 \qquad \forall a, p, r, s. \tag{5.39}$$

For route $p$ between origin node $r$ and destination node $s$, let $i$ denote node $r$ or any intermediate node on this route. Denote route $\tilde{p}$ as $(i, 1, 2, \cdots, n, s)$. The instantaneous travel time $\psi_{\tilde{p}}^{is}(t)$ for the remaining route $\tilde{p}$ between $i$ and $s$ is

$$\psi_{\tilde{p}}^{is}(t) = \sum_{a \in is\tilde{p}} c_a[x_a(t), u_a(t), v_a(t)] \qquad \forall i \in p, r, s. \tag{5.40}$$

Suppose we consider a set of routes $p$ from $r \to i \to s$ and the corresponding set of subroutes $\tilde{p}$. The flow conservation constraint at node $i$ can be revised as

$$\sum_{a \in B(i)} v_{ap}^{rs}(t) = \sum_{a \in A(i)} u_{ap}^{rs}(t) \qquad \forall i, p, r, s; i \neq r, s. \tag{5.41}$$

The fourth term in the Hamiltonian function should be revised as

$$\sum_{rs} \sum_{i \neq rs} \sigma_i^{rs}(t) \sum_p [ \sum_{a \in B(j)} v_{ap}^{rs}(t) - \sum_{a \in A(j)} u_{ap}^{rs}(t)$$

so that $\sigma_{ip}^{rs}(t) = \sigma_i^{rs}(t)$ for the set of subroutes $\tilde{p}$. Note that all derivations from equation (5.27) to equation (5.39) will follow for this set of subroutes $\tilde{p}$. Now if route $\tilde{p}$ is being used at time $t$, $u_a^{rs}(t)$ and $v_a^{rs}(t)$ are both positive by definition. Thus, by (5.37)-(5.39),

$$\begin{aligned} \psi_{\tilde{p}}^{is}(t) &= [\sigma_i^{rs}(t) - \sigma_1^{rs}(t)] + [\sigma_{1p}^{rs}(t) - \sigma_{2p}^{rs}(t)] + \cdots \\ &+ [\sigma_{n-1,p}^{rs}(t) - \sigma_{np}^{rs}(t)] + [\sigma_{np}^{rs}(t) - \sigma_s^{rs}(t)] \\ &= \sigma_i^{rs}(t) - \sigma_s^{rs}(t) \end{aligned}$$

for every route $\tilde{p}$ being used at time $t$. Note that $\sigma_{ip}^{rs}(t) = \sigma_i^{rs}(t)$ in the above equation. Thus, routes being used from $i$ to $s$ at time $t$ have travel times equal to $[\sigma_i^{rs}(t) - \sigma_s^{rs}(t)]$. More generally, we obtain the following for each remaining route $\tilde{p}$ between $i$ and $s$.

$$u_{ap}^{rs}(t)\, v_{ap}^{rs}(t)\, [\psi_{\tilde{p}}^{is}(t) - \sigma_i^{rs}(t) + \sigma_s^{rs}(t)] = 0 \quad \forall a, i, p, r, s; a \in \tilde{p}, i \in p; \tag{5.42}$$

$$\psi_{\tilde{p}}^{is}(t) \geq \sigma_i^{rs}(t) - \sigma_s^{rs}(t) \qquad \forall i \in p, r, s; \tag{5.43}$$

$$u_{ap}^{rs}(t) \geq 0, \qquad v_{ap}^{rs}(t) \geq 0, \qquad \forall a, p, r, s; a \in \tilde{p}. \tag{5.44}$$

Conditions (5.42)-(5.44) hold for each remaining route $\tilde{p}$ between $i$ and $s$, where $i$ is any intermediate node (including the origin) between each O-D pair $(r, s)$ in the network. For route $\tilde{p}$ connecting node $i$ and destination $s$, if each link is being used at time $t$, then $u_{ap}^{rs}(t)$ and $v_{ap}^{rs}(t)$ will be positive, so that the quantities in brackets in equation (5.42) will be zero, i.e., equation (5.43) will hold as an equality. Thus, routes which are being used at time $t$ have travel times equal to $[\sigma_i^{rs}(t) - \sigma_s^{rs}(t)]$.

Equation (5.43) states that the difference of Lagrange multipliers $[\sigma_i^{rs}(t) - \sigma_s^{rs}(t)]$ of node-destination pair $(i, s)$ is less than or equal to the instantaneous travel times on all routes connecting this node-destination pair $(i, s)$. Therefore, this difference of Lagrange multipliers $[\sigma_i^{rs}(t) - \sigma_s^{rs}(t)]$ equals the instantaneous minimal route travel time between node $i$ and destination $s$. For any remaining unused route $\tilde{p}$ between $i$ and $s$, at least one link $a$ in route $\tilde{p}$ is *not* being used at time $t$. Thus the inflow $u_{ap}^{rs}(t)$ or the exit flow $v_{ap}^{rs}(t)$ is equal to zero, so that (5.43) may hold as a strict inequality, i.e., the instantaneous travel time

$\psi_{\tilde{p}}^{is}(t)$ on remaining route $\tilde{p}$ will not be less than the instantaneous minimal route travel time $[\sigma_i^{rs}(t) - \sigma_s^{rs}(t)]$.

Furthermore, if we consider any pair of decision nodes $i$ and $j$ between origin $r$ and destination $s$, the above results also hold for each sub-route between $i$ and $j$. Thus, at the optimal solution to Model 1, flows always use the instantaneous minimal time sub-routes, even if all links on an entire route are not in use at the same time.

Since intermediate node $i$ could be origin node $r$, the above results also hold for routes from $r$ to $s$. With the above interpretation, it is now clear that equations (5.42)-(5.44) state the travel-time-based instantaneous dynamic user-optimal conditions. This optimal control program (5.11)-(5.23) can be referred to as an instantaneous DUO route choice program or an instantaneous DUO equivalent optimal control problem.

# 5.4 Instantaneous Route Choice Model 2

## 5.4.1 Model Formulation

Various DUO route choice models can be formulated with alternative link travel time assumptions. In Model 1, the instantaneous travel time $c_a(t)$ over link $a$ is assumed to be dependent on the number of vehicles $x_a(t)$, the inflow $u_a(t)$, and the exit flow $v_a(t)$ on link $a$ at time $t$. In this section, the instantaneous link travel time $c_a(t)$ is assumed to be dependent only on $x_a(t)$ and $v_a(t)$:

$$c_a(t) = c_a[x_a(t), v_a(t)]. \tag{5.45}$$

The time-dependent link travel time function $c_a[x_a(t), v_a(t)]$ is again assumed to be nonnegative and differentiable for all $x_a(t)$ and $v_a(t)$.

Using optimal control theory, the equivalent optimization program of the instantaneous dynamic user-optimal route choice problem (**Model 2**) is formulated as follows.

$$\min_{u,v,x,e,E} \quad \int_0^T \sum_a \left\{ \int_0^{v_a(t)} c_a[x_a(t), \omega]\, d\omega \right\} dt \tag{5.46}$$

s.t. constraints (5.12)–(5.23).

The objective function is similar to the objective function of the well-known static user-optimal model. The constraints are identical to those of Model 1.

In Model 1, the instantaneous travel time $c_a[x_a(t), u_a(t), v_a(t)]$ over link $a$ is assumed to be the sum of two components: 1) an instantaneous flow-dependent running time $g_{1a}[x_a(t), u_a(t)]$ over link $a$, and 2) an instantaneous queuing delay $g_{2a}[x_a(t), v_a(t)]$. If $g_{1a}[x_a(t), u_a(t)]$ is set equal to zero and the queuing delay $g_{2a}[x_a(t), v_a(t)]$ is extended to include the running time, Model 1 is equivalent to Model 2. We prove in the next section that the unique optimal solution to Model 2 is in a travel-time-based instantaneous DUO state.

## 5.4.2 Optimality Conditions

The extended Hamiltonian for Model 2 is constructed as

$$
\begin{aligned}
\mathcal{H} &= \sum_a \int_0^{v_a(t)} c_a[x_a(t), \omega] d\omega \\
&+ \sum_{rs} \sum_{ap} \lambda_{ap}^{rs}(t)[u_{ap}^{rs}(t) - v_{ap}^{rs}(t)] + \sum_r \sum_{s \neq r} \sum_p \nu_p^{rs}(t) e_p^{rs}(t) \\
&+ \sum_s \sum_{r \neq s} \sigma_r^{rs}(t)[f^{rs}(t) - \sum_{a \in A(r)} \sum_p u_{ap}^{rs}(t)] \\
&+ \sum_{rs} \sum_{j \neq rs} \sum_p \sigma_{jp}^{rs}(t)[\sum_{a \in B(j)} v_{ap}^{rs}(t) - \sum_{a \in A(j)} u_{ap}^{rs}(t)] \\
&+ \sum_r \sum_{s \neq r} \sigma_s^{rs}(t)[\sum_{a \in B(s)} \sum_p v_{ap}^{rs}(t) - e^{rs}(t)] \\
&+ \sum_{rsp} \sum_{j \neq r} \sum_{a \in B(j)} \mu_{ap}^{rs}(t) \Bigg\{ x_{ap}^{rs}(t) + \sum_{b \in \tilde{p}} x_{bp}^{rs}(t) + E_p^{rs}(t) \\
&- \sum_{b \in \tilde{p}} x_{bp}^{rs}[t + \bar{\tau}_a(t)] - E_p^{rs}[t + \bar{\tau}_a(t)] \Bigg\}
\end{aligned}
$$

where $\lambda_{ap}^{rs}(t)$ are Lagrange multipliers associated with the link state equations, $\nu_p^{rs}(t)$ are Lagrange multipliers associated with the destination node state equations, $\sigma_{jp}^{rs}(t)$ are Lagrange multipliers associated with the node flow conservation equations, and $\mu_{ap}^{rs}(t)$ are Lagrange multipliers associated with the flow propagation equations. For each link $a$ which points from node $l$ to node $m$, the first order necessary conditions of Model 2 include

$$\frac{\partial \mathcal{H}}{\partial u_{ap}^{rs}(t)} = \lambda_{ap}^{rs}(t) - \sigma_{lp}^{rs}(t) \geq 0, \qquad \forall l; a \in A(l), p, r, s, \tag{5.47}$$

$$\text{and} \qquad u_{ap}^{rs}(t) \frac{\partial \mathcal{H}}{\partial u_{ap}^{rs}(t)} = 0 \qquad \forall a, p, r, s; \tag{5.48}$$

$$\frac{\partial \mathcal{H}}{\partial v_{ap}^{rs}(t)} = c_a[x(t), v(t)] - \lambda_{ap}^{rs}(t) + \sigma_{mp}^{rs}(t) \geq 0, \qquad \forall m; a \in B(m), p, r, s, \tag{5.49}$$

$$\text{and} \qquad v_{ap}^{rs}(t) \frac{\partial \mathcal{H}}{\partial v_{ap}^{rs}(t)} = 0 \qquad \forall a, p, r, s; \tag{5.50}$$

$$\frac{\partial \mathcal{H}}{\partial e_p^{rs}(t)} \geq 0, \qquad \forall p, r, s, \tag{5.51}$$

$$\text{and} \qquad e_p^{rs}(t) \frac{\partial \mathcal{H}}{\partial e_p^{rs}(t)} = 0 \qquad \forall p, r, s; \tag{5.52}$$

$$\frac{d\lambda_{ap}^{rs}(t)}{dt} = -\frac{\partial \mathcal{H}}{\partial x_{ap}^{rs}(t)}, \qquad \forall a, p, r, s; \tag{5.53}$$

$$\frac{d\nu_{p}^{rs}(t)}{dt} = -\frac{\partial \mathcal{H}}{\partial E_{p}^{rs}(t)} \qquad \forall p, r, s; \tag{5.54}$$

$$u_{ap}^{rs}(t) \geq 0, \qquad v_{ap}^{rs}(t) \geq 0, \qquad x_{ap}^{rs}(t) \geq 0, \qquad \forall a, p, r, s; \tag{5.55}$$

$$e_{p}^{rs}(t) \geq 0, \qquad E_{p}^{rs}(t) \geq 0, \qquad \forall p, r, s. \tag{5.56}$$

The equivalence of the instantaneous dynamic user-optimal conditions and Model 2 is demonstrated by showing that the unique trajectories of link states, inflows and exit flows that solve Model 2 also satisfy the instantaneous dynamic user-optimal conditions. As with Model 1, this equivalence is demonstrated by proving that the first order necessary conditions for Model 2 are identical to the instantaneous dynamic user-optimal conditions.

Combining equations (5.47)-(5.48) with equations (5.49)-(5.50), the same equations as equations (5.42)-(5.44) for Model 1 can be derived for each link $a$ which points from node $l$ to node $m$. Thus, we follow the derivation of optimality conditions for Model 1 and obtain the same equations (5.37)-(5.39) for each remaining route $\tilde{p}$ between intermediate node $i$ and destination $s$. Similarly, we can obtain the same interpretation of the optimality conditions for Model 2 as those for Model 1. It is now clear that the optimality conditions of Model 2 state the instantaneous dynamic user-optimal route choice conditions. Therefore, Model 2 is an equivalent instantaneous DUO route choice program.

## 5.5 Instantaneous Route Choice Model 3

Now, the instantaneous link travel time $c_a(t)$ is assumed to depend only on $x_a(t)$:

$$c_a(t) = c_a[x_a(t)]. \tag{5.57}$$

The time-dependent link travel time function $c_a[x_a(t)]$ is again assumed to be nonnegative and differentiable for all $x_a(t)$. In addition, the actual link travel time $\tau_a(t)$ is introduced, and also assumed to be dependent only on $x_a(t)$, i.e.,

$$\tau_a(t) = \tau_a[x_a(t)]. \tag{5.58}$$

Using optimal control theory, the equivalent optimization program of the instantaneous dynamic user-optimal route choice problem (**Model 3**) is formulated as follows.

$$\min_{u,v,x,e,E} \quad \int_0^T \sum_a \left\{ \int_0^{u_a(t)} c_a[x_a(t)]\, d\omega \right\} dt \tag{5.59}$$

s.t. constraints (5.12)–(5.23).

The constraints are identical to those of Model 1 except that the flow propagation constraints need to be revised. The flow propagation constraints can be expressed directly based on the actual link travel time function $\tau_a(t)$, instead of its estimate $\bar{\tau}_a(t)$ in a diagonalization fashion. The new flow propagation constraints are

$$x_{ap}^{rs}(t) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[t + \tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall a \in B(j); j \neq r; p, r, s. \qquad (5.60)$$

The instantaneous link travel time function $c_a(t)$ in Model 3 is a special case of $c_a[x_a(t), u_a(t), v_a(t)]$ in Model 1 or $c_a[x_a(t), v_a(t)]$ in Model 2. We can formulate the extended Hamiltonian function and derive the first-order necessary conditions. Since the actual link travel time $\tau_a(t)$ is also assumed to be dependent only on $x_a(t)$, the terms in flow propagation constraints (5.60) have no impact on the partial derivative of the Hamiltonian function with respect to control variables $u_{ap}^{rs}(t)$ and $v_{ap}^{rs}(t)$. Following the same process as in Model 1, the same equations as (5.42)-(5.44) for Model 1 can be derived for each link $a$ which points from node $l$ to node $m$.

Thus, we follow the derivation of optimality conditions for Model 1 and obtain the same equations (5.37)-(5.39) for each remaining route $\tilde{p}$ between intermediate node $i$ and destination $s$. Similarly, we can obtain the same interpretation of the optimality conditions for Model 3 as those for Model 1. It is now clear that the optimality conditions of Model 3 state the instantaneous dynamic user-optimal route choice conditions. Therefore, optimal control Model 3 is an equivalent instantaneous DUO route choice program. Since the flow propagation constraints are expressed directly using the actual link travel time function $\tau_a(t)$, instead of its estimate $\bar{\tau}_a(t)$ in a diagonalization fashion, Model 3 is a complete optimal control model. It will be shown in Chapter 12 that it is equivalent to a variational inequality.

## 5.6 A Numerical Example

We illustrate the solution of Model 1 with the 4-link, 4-node test network shown in Figure 5.4. A symmetrical network is intentionally used to demonstrate that the route travel times in the solution are equal. To convert OCP model (5.11)-(5.23) into an NLP, assignment time period $[0, T]$ is subdivided into $K = 5$ small time intervals, and the OCP is reformulated as a discrete time NLP. We then use an algorithm based on the Frank-Wolfe and diagonalization techniques to solve this NLP. This algorithm was coded in FORTRAN and solved on a IBM 3090-300J. The details of the algorithm are presented in Chapter 6; only computational results for a small network are given to illustrate the instantaneous DUO traffic flows.

The following link travel time functions were used in the computations:

$$c_a(k) = g_{1a}(k) + g_{2a}(k)$$

$$g_{1a}(k) = \beta_{1a} + \beta_{2a}[u_a(k)]^2 + \beta_{3a}[x_a(k)]^2$$
$$g_{2a}(k) = \beta_{4a} + \beta_{5a}[v_a(k)]^2 + \beta_{6a}[x_a(k)]^2$$

where time interval $k = 1, 2, \cdots, 5$. Parameter values for each link travel time function are given in Table 5.1, and the trip table is given in Table 5.2. The optimal link flows and corresponding optimal link travel times are given in Table 5.3. The optimal route travel times are given in Table 5.4. In this discrete time example, $x_a(k)$ represents vehicles on the link at the beginning of interval $k$; $u_a(k)$ and $v_a(k)$ represent inflow and exit flow during interval $k$.

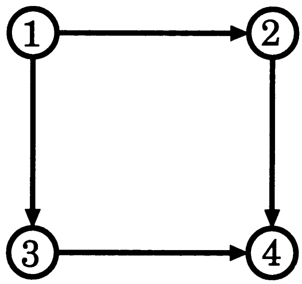



Figure 5.4: Test Network

Table 5.1: Parameters of Link Travel Time Functions

| link $a$ | $\beta_{1a}$ | $\beta_{2a}$ | $\beta_{3a}$ | $\beta_{4a}$ | $\beta_{5a}$ | $\beta_{6a}$ |
|---|---|---|---|---|---|---|
| 1—2 | 1. | 0.001 | 0. | 0. | 0.015 | 0.003 |
| 1—3 | 1. | 0.001 | 0. | 0. | 0.015 | 0.003 |
| 2—4 | 1. | 0.001 | 0. | 0. | 0.015 | 0.003 |
| 3—4 | 1. | 0.001 | 0. | 0. | 0.015 | 0.003 |

Table 5.2: Required Flows from Origin 1 to Destination 4

| Time Interval $k$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Flow/interval | 20. | 0. | 0. | 0. | 0. |

In this example, note from Table 5.4 that travel times on routes 1–2–4 and 1–3–4 are equal during each interval (except for small differences resulting from incomplete convergence). Thus, the results indicate the existence of dynamic user optimality in the intuitive sense that the two routes have equal

Table 5.3: Optimal Numbers of Vehicles, Inflows, Exit Flows and Travel Times

| Interval $k$ | Link $a$ | Vehicles $x_a(k+1)$ | Inflow $u_a(k)$ | Exit Flow $v_a(k)$ | Vehicles $x_a(k)$ | Travel Time $c_a(k)$ |
|---|---|---|---|---|---|---|
| 1 | 1 — 2 | 10.0 | 10.0 | 0.0 | 0.0 | 1.1 |
| 2 | 1 — 2 | 4.5 | 0.0 | 5.5 | 10.0 | 1.8 |
| 3 | 1 — 2 | 0.0 | 0.0 | 4.5 | 4.5 | 1.4 |
| 4 | 1 — 2 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 5 | 1 — 2 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 1 | 1 — 3 | 10.0 | 10.0 | 0.0 | 0.0 | 1.1 |
| 2 | 1 — 3 | 4.4 | 0.0 | 5.6 | 10.0 | 1.8 |
| 3 | 1 — 3 | 0.0 | 0.0 | 4.4 | 4.4 | 1.3 |
| 4 | 1 — 3 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 5 | 1 — 3 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 1 | 2 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 2 | 2 — 4 | 5.5 | 5.5 | 0.0 | 0.0 | 1.0 |
| 3 | 2 — 4 | 4.5 | 4.5 | 5.5 | 5.5 | 1.6 |
| 4 | 2 — 4 | 0.0 | 0.0 | 4.5 | 4.5 | 1.4 |
| 5 | 2 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 1 | 3 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 2 | 3 — 4 | 5.6 | 5.6 | 0.0 | 0.0 | 1.0 |
| 3 | 3 — 4 | 4.4 | 4.4 | 5.6 | 5.6 | 1.6 |
| 4 | 3 — 4 | 0.0 | 0.0 | 4.4 | 4.4 | 1.3 |
| 5 | 3 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |

Table 5.4: Dynamic User-Optimal Route Travel Times

| Interval | Route Travel Times | |
|---|---|---|
| $k$ | Route 1-2-4 | Route 1-3-4 |
| 1 | 2.1 | 2.1 |
| 2 | 2.8 | 2.8 |
| 3 | 3.0 | 2.9 |
| 4 | 2.4 | 2.3 |
| 5 | 2.0 | 2.0 |

travel times in each interval, even though the two routes are only partially used; i.e., we do not have $u_{ap}^{rs}(k) > 0$ and $v_{ap}^{rs}(k) > 0$ for all links on the two routes.

Note from Table 5.3 that the *average* travel time that vehicles incur on link 1–2 from the beginning of interval 2 is 1.8 intervals (some vehicles have smaller travel times and some have larger); 10 vehicles are on the link at the beginning of interval 2. Some vehicles exit during interval 2, which reduces the travel time in subsequent intervals. Thus, 5.5 vehicles exit during interval 2 (decreasing the travel time to 1.4 intervals for interval 3) and 4.5 vehicles exit during interval 3. Note also that 5.5 vehicles are on link 2–4 at the beginning of interval 3, and the average travel time is 1.6 from the beginning of interval 3. Also, 4.5 vehicles enter this link during interval 3, 5.5 vehicles exit during interval 3, and 4.5 vehicles exit during interval 4, reducing the travel time to only 1.0 for interval 5. At the beginning of interval 6 (column $x_a(k+1)$ in rows $k = 5$), no vehicle remains on any link.

## 5.7 Notes

### 5.7.1 Several Formulation Issues

There are many possible dynamic generalizations of the static user-optimal route choice model. Some of them may be suitable for providing instantaneous information and advising vehicles of their best routes in a dynamic route guidance system. Since our instantaneous DUO route choice models have a direct correspondence to the static UO assignment model, other static UO formulations can be expected to have their DUO counterparts. Various optimal control formulations should be investigated and their corresponding solution characteristics, such as the uniqueness of the optimal control strategies, should be studied.

In Chapter 12, our instantaneous DUO route choice models are extended to more realistic situations when both capacity constraints and oversaturation constraints are taken into account. In this section, we discuss several specific problems encountered in our model formulations to date, in part to document dead ends that have been explored and rejected.

- Time Period $[0, T]$

In reality, the "travel period" is infinitely long; that is, the origin-destination flows never become permanently zero. Since we want a finite horizon model, it is inherent that there must be a boundary condition to represent the artificial termination period in our models. If the time period is too short, travelers who do not reach the end of their paths at the end of the time horizon are not "using" their path, and thus they can take literally any path. This can be prevented by using a long enough time horizon so that all flow clears the network. The rolling horizon method could be used to represent more realistic traffic flows which never become permanently zero.

• Penalty and Salvage Costs

In the development of flow propagation constraints, we initially used the penalty cost $D_a[x_a(t)]$ to "force" the vehicle flows to continue propagation. Also, a salvage cost $S_a[x_a(T)]$ was used as the artificial boundary condition for the final time instant to clear the traffic flow in the network. These constraints are standard in optimal control theory.

In reality, traffic flow will not clear and will continue to appear in the network in an infinite time horizon. However, since we are only able to analyze a finite time period, we need a boundary condition to clear the traffic of interest (such as peak hour traffic) at the end of the analysis period. This boundary condition can either be represented by salvage cost in the objective function or be represented by a set of physical constraints at the destination.

Later, we introduced a set of physical constraints at the destination. These constraints state that the cumulative exiting vehicles from destinations equal some amount of cumulative vehicles departing origins. It follows that

$$E^{rs}(T) = F^{rs}(T - \tau_m^{rs}) \qquad \forall r, s \tag{5.61}$$

where $\tau_m^{rs}$ is a prespecified maximum O-D travel time, such as 60 or 120 minutes for a typical peak hour period.

In our computation experience, we found that the penalty and salvage cost terms produced a long-tailed result; i.e., a small amount of traffic remains in the network for an excessively long time. This result suggested the use of physical constraints in the model. Thus, these terms are now represented by the more accurate flow propagation constraints.

### 5.7.2 Properties of Models

The definition of instantaneous DUO is similar to that given in Friesz et al (1989). However, in their model, the route travel costs are equal at every instant of time; in contrast, in our models, instantaneous route travel times equal the minimum instantaneous travel time only at each decision node that has flows to the destination. We show that our alternative definition allows the resulting models to equilibrate flows using route travel times based only on link driving times. This is a fundamental difference in the definition of instantaneous route travel time between their model and our models. We discuss this point further in the following.

Moreover, we discuss some fundamental differences among existing dynamic route choice models. These models diverge at two points: 1) system-optimal vs. user-optimal; and 2) the interpretation of route travel time. Following Merchant and Nemhauser (1978), Carey (1987) presented improved dynamic system-optimal models which considered the minimization of cumulative instantaneous route travel times. However, the dynamic *user-optimal* state was not considered in his models. Friesz et al (1989) presented an instantaneous dynamic user-optimal route choice model which did consider the equilibration of instantaneous unit route travel cost. In their model, however, the instantaneous unit route travel cost is defined as

$$\psi_p^{rs}(t) = \sum_{a \in rsp} \left[ c_a[x_a(t)] + \frac{\dot{\lambda}_a(t)}{g_a'[x_a(t)]} \right] \qquad \forall r, s, p, \tag{5.62}$$

where the numerator of the second term is the time variation (rate of change) of the Lagrange multiplier $\lambda_a(t)$ and the denominator is the derivative of exit flow function $g_a[x_a(t)]$ with respect to link state variable $x_a(t)$. Since both of these could change from problem to problem, their presence may present a difficulty for the physical interpretation of the model. In contrast, our models only use the summation of instantaneous link travel times as the instantaneous route travel time:

$$\psi_p^{rs}(t) = \sum_{a \in rsp} c_a[x_a(t), u_a(t), v_a(t)] \qquad \forall r, s, p. \tag{5.63}$$

This explicit definition of instantaneous route travel time reflects the dynamic route choice behavior and decision criterion of travelers. It is also consistent with the definition of instantaneous route travel time in the DSO route choice models of Merchant-Nemhauser (1978) and Carey (1987). We note that it is not known how to formulate an optimization program which is consistent with our definition of instantaneous route travel time if exit flow functions instead of exit flow variables are used.

# Chapter 6

# A Computational Algorithm for Instantaneous Dynamic User-Optimal Route Choice Models

In this chapter, solution algorithms are considered for solving the instantaneous DUO route choice models presented in Chapter 5. A capability to solve the DUO route choice problem is needed for several reasons. First, it appears that properties of alternative models can only be fully understood by computing solutions to hypothetical and real test problems. Unlike their static counterparts, dynamic models are sufficiently opaque that they are difficult to understand analytically. Second, computational solutions for standard test problems based on actual networks are needed to evaluate how well alternative models describe reality. Third, solutions for large networks are required to evaluate the potential effectiveness of proposed in-vehicle navigation and route guidance systems. Ultimately, such models might be used to guide the operation of such systems; however, the requirements of such systems are so undefined at this time that any discussion of algorithmic requirements is highly speculative.

The objective of this chapter is to describe in detail an algorithm for solving one of the instantaneous DUO route choice models and to illustrate its performance with a toy network. Through the development and implementation of the algorithm, additional insights into the model's properties have been gained. These properties are also discussed.

In Section 6.1, we reformulate the instantaneous DUO route choice model as a discrete-time nonlinear program (NLP). Then the diagonalization technique and the Frank-Wolfe algorithm are employed to solve the NLP. In the diagonalization procedure, the estimated link travel time is updated iteratively. Then we apply the Frank-Wolfe technique to solve the NLP. An expanded time-space network is constructed in Section 6.2 so that each LP subproblem can be decomposed according to O-D pairs and can be viewed as a set of minimal-cost route problems. The flow propagation constraints representing the relationship of link flows and travel times are satisfied in modified minimal-cost route searches in Section 6.3 so that only flow conservation constraints for links and nodes remain. Since the model is convex, the discrete version should be efficient to solve for large networks. A numerical example is given in Section 6.4.

# 6.1 The Algorithm

## 6.1.1 Discrete Instantaneous DUO Route Choice Model

The Frank-Wolfe algorithm (Frank and Wolfe, 1956) is reasonably efficient for solving nonlinear programming problems (NLP) with network constraints, and has been widely used for solving the static UO model on urban networks (LeBlanc et al, 1975). They showed that solving the static UO model simplifies to solving a sequence of minimal-cost route (shortest path) problems and line searches. Thus, the computing times can be reduced by orders of magnitude for large-scale networks, as compared with solving a sequence of linear programming problems.

To convert our instantaneous DUO route choice model into an NLP, the time period $[0, T]$ is subdivided into $K$ small time increments. Each time increment is a unit of time. Then, $u_a(k)$ represents the inflow into link $a$ during interval $k$ and $v_a(k)$ represents the exit flow from link $a$ during interval $k$. To simplify the formulation, we modify the estimated actual travel time on each link in the following way so that each estimated travel time is equal to a multiple of the time increment.

$$\bar{\tau}_a(k) = i \quad \text{if} \quad i - 0.5 \leq \bar{\tau}_a(k) < i + 0.5,$$

where $i$ is an integer and $0 \leq i \leq K$. We note that this round-off method is used only in the flow propagation constraints. More accurate flow propagation constraints can be obtained by making the time intervals smaller. We also note that evaluation of the instantaneous link travel time function and objective function does not have this round-off error so that the subsequent minimal-cost route search does not have this round-off error.

An optimal control program can then be reformulated as a discrete time NLP as follows:

$$\min_{u,v,x,E} \quad Z = \sum_{k=1}^{K} \sum_{a} \left\{ \int_0^{u_a(k)} g_{1a}[x_a(k), \omega] d\omega + \int_0^{v_a(k)} g_{2a}[x_a(k), \omega] d\omega \right\} \tag{6.1}$$

s.t.

$$x_{ap}^{rs}(k+1) = x_{ap}^{rs}(k) + u_{ap}^{rs}(k) - v_{ap}^{rs}(k) \qquad \forall a, p, r, s; k = 1, \cdots, K; \tag{6.2}$$

$$E^{rs}(k+1) = E^{rs}(k) + \sum_{a \in B(s)} \sum_{p} v_{ap}^{rs}(k) \qquad \forall r; s \neq r; k = 1, \cdots, K; \quad (6.3)$$

$$\sum_{a \in A(r)} \sum_{p} u_{ap}^{rs}(k) = f^{rs}(k) \qquad \forall r \neq s; k = 1, \cdots, K; \quad (6.4)$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(k) - \sum_{a \in A(j)} u_{ap}^{rs}(k) = 0 \qquad \forall j, p, r, s; j \neq r, s; k = 1, \cdots, K; \quad (6.5)$$

$$x_{ap}^{rs}(k) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[k + \bar{\tau}_a(k)] - x_{bp}^{rs}(k)\} + \{E_p^{rs}[k + \bar{\tau}_a(k)] - E_p^{rs}(k)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; k = 1, \cdots, K+1; \quad (6.6)$$

$$u_{ap}^{rs}(k) \geq 0, \quad v_{ap}^{rs}(k) \geq 0, \quad x_{ap}^{rs}(k+1) \geq 0, \qquad \forall a, p, r, s; k = 1, \cdots, K; \quad (6.7)$$

$$E_p^{rs}(k+1) \geq 0, \qquad \forall p, r, s; k = 1, \cdots, K; \quad (6.8)$$

$$E_p^{rs}(1) = 0 \qquad \forall p, r, s; \quad (6.9)$$

$$x_{ap}^{rs}(1) = 0, \qquad \forall a, p, r, s. \quad (6.10)$$

### 6.1.2 The Diagonalization/Frank-Wolfe Algorithm

Denote the subproblem variables as $p, q, y, \bar{E}$, corresponding to the main problem variables $u, v, x, E$. Applying the Frank-Wolfe algorithm to the minimization of the discretized DUO route choice program requires, at each iteration, a solution of the following linear program (LP) :

$$\begin{aligned} \min_{p,q,y,\bar{E}} \hat{Z} &= \nabla_u Z(u,v,x,E)\, p^T + \nabla_v Z(u,v,x,E)\, q^T \\ &+ \nabla_x Z(u,v,x,E)\, y^T + \nabla_E Z(u,v,x,E)\, \bar{E}^T \end{aligned} \quad (6.11)$$

s.t.

$$y_{ap}^{rs}(k+1) = y_{ap}^{rs}(k) + p_{ap}^{rs}(k) - q_{ap}^{rs}(k) \qquad \forall a, p, r, s; k = 1, \cdots, K; \quad (6.12)$$

$$\bar{E}^{rs}(k+1) = \bar{E}^{rs}(k) + \sum_{a \in B(s)} \sum_{p} q_{ap}^{rs}(k) \qquad \forall r; s \neq r; k = 1, \cdots, K; \quad (6.13)$$

$$\sum_{a \in A(r)} \sum_{p} p_{ap}^{rs}(k) = f^{rs}(k) \qquad \forall r \neq s; k = 1, \cdots, K; \quad (6.14)$$

$$\sum_{a \in B(j)} q_{ap}^{rs}(k) - \sum_{a \in A(j)} p_{ap}^{rs}(k) = 0 \qquad j, p, r, s; \forall j \neq r, s; k = 1, \cdots, K; \quad (6.15)$$

$$y_{ap}^{rs}(k) = \sum_{b \in \tilde{p}} \{y_{bp}^{rs}[k + \bar{\tau}_a(k)] - y_{bp}^{rs}(k)\} + \{\bar{E}_p^{rs}[k + \bar{\tau}_a(k)] - \bar{E}_p^{rs}(k)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; k = 1, \cdots, K+1; \quad (6.16)$$

$$p_{ap}^{rs}(k) \geq 0, \quad q_{ap}^{rs}(k) \geq 0, \quad y_{ap}^{rs}(k+1) \geq 0, \quad \forall a, p, r, s; k = 1, \cdots, K; \tag{6.17}$$

$$\bar{E}_p^{rs}(k+1) \geq 0, \qquad \forall p, r, s; k = 1, \cdots, K; \tag{6.18}$$

$$\bar{E}_p^{rs}(1) = 0, \qquad \forall p, r, s; \tag{6.19}$$

$$y_{ap}^{rs}(1) = 0, \qquad \forall a, p, r, s. \tag{6.20}$$

Objective function (6.11) is equivalent to:

$$\begin{aligned}\hat{Z} &= \sum_{k=1}^{K} \sum_{r,s} \sum_{ap} \left[ \frac{\partial Z}{\partial u_{ap}^{rs}(k)} p_{ap}^{rs}(k) + \frac{\partial Z}{\partial v_{ap}^{rs}(k)} q_{ap}^{rs}(k) + \frac{\partial Z}{\partial x_{ap}^{rs}(k+1)} y_{ap}^{rs}(k+1) \right] \\ &+ \sum_{k=1}^{K} \sum_{r,s,p} \frac{\partial Z}{\partial E_p^{rs}(k+1)} \bar{E}_p^{rs}(k+1) \end{aligned} \tag{6.21}$$

The components of the gradient of $Z(u, v, x, E)$ with respect to control and state variables $u$, $v$, $x$, $E$ are

$$t_{1a}(k) = \frac{\partial Z(u, v, x, E)}{\partial u_a(k)} = g_{1a}[x_a(k), u_a(k)] \qquad \forall a; k = 1, \cdots, K; \tag{6.22}$$

$$t_{2a}(k) = \frac{\partial Z(u, v, x, E)}{\partial v_a(k)} = g_{2a}[x_a(k), v_a(k)] \qquad \forall a; k = 1, \cdots, K; \tag{6.23}$$

$$\begin{aligned} t_{3a}(k) &= \frac{\partial Z(u, v, x, E)}{\partial x_a(k)} \\ &= \int_0^{u_a(k)} \frac{\partial g_{1a}[x_a(k), \omega]}{\partial x_a(k)} d\omega + \int_0^{v_a(k)} \frac{\partial g_{2a}[x_a(k), \omega]}{\partial x_a(k)} d\omega \\ &\qquad \forall a; k = 2, \cdots, K; \end{aligned} \tag{6.24}$$

$$t_{3a}(K+1) = \frac{\partial Z(u, v, x, E)}{\partial x_a(K+1)} = 0 \qquad \forall a; \tag{6.25}$$

$$t_4^{rs}(k) = \frac{\partial \hat{Z}(u, v, x, E)}{\partial E^{rs}(k)} = 0 \qquad \forall r, s; k = 2, \cdots, K+1. \tag{6.26}$$

The objective function can be rewritten as

$$\hat{Z} = \sum_{k=1}^{K} \sum_{r,s,a,p} \left[ t_{1a}(k) p_{ap}^{rs}(k) + t_{2a}(k) q_{ap}^{rs}(k) + t_{3a}(k+1) y_{ap}^{rs}(k+1) \right] \tag{6.27}$$

Since $g_{1a}$ and $g_{2a}$ are nonnegative and increasing functions, it follows that

$$t_{1a}(k) \geq 0, t_{2a}(k) \geq 0, t_{3a}(k+1) \geq 0, \qquad \forall a; k = 1, \cdots, K. \tag{6.28}$$

This property has a significant impact on the solvability of this model for large networks, since these components will be link cost coefficients in a minimal-cost network subproblem.

Note that there are no capacity constraints on the links. In addition to the flow propagation constraints (6.16) and link definitional constraints (6.12), the only constraints are non-negativity and conservation of flow. Furthermore, the constraints apply to each origin-destination pair independently, so linear program (6.11)-(6.20) can be decomposed by origin-destination pair. The resulting subproblem for each O-D pair $(r, s)$ is given by

$$\min_{p,q,y,\bar{E}} \quad \sum_{k=1}^{K}\sum_{ap} \left[t_{1a}\, p_{ap}^{rs}(k) + t_{2a}\, q_{ap}^{rs}(k) + t_{3a}^{rs}(k+1)\, y_{ap}^{rs}(k+1)\right] \tag{6.29}$$

s.t.

$$y_{ap}^{rs}(k+1) = y_{ap}^{rs}(k) + p_{ap}^{rs}(k) - q_{ap}^{rs}(k) \qquad \forall a, p; k = 1, \cdots, K; \tag{6.30}$$

$$\bar{E}^{rs}(k+1) = \bar{E}^{rs}(k) + \sum_{a \in B(s)} \sum_{p} q_{ap}^{rs}(k) \qquad \forall k = 1, \cdots, K; \tag{6.31}$$

$$\sum_{a \in A(r)} \sum_{p} p_{ap}^{rs}(k) = f^{rs}(k) \qquad \forall k = 1, \cdots, K; \tag{6.32}$$

$$\sum_{a \in B(j)} q_{ap}^{rs}(k) - \sum_{a \in A(j)} p_{ap}^{rs}(k) = 0 \qquad \forall j, p; j \neq r, s; k = 1, \cdots, K; \tag{6.33}$$

$$y_{ap}^{rs}(k) = \sum_{b \in \tilde{p}} \{y_{bp}^{rs}[k + \bar{\tau}_a(k)] - y_{bp}^{rs}(k)\} + \{\bar{E}_p^{rs}[k + \bar{\tau}_a(k)] - \bar{E}_p^{rs}(k)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; k = 1, \cdots, K+1; \tag{6.34}$$

$$p_{ap}^{rs}(k) \geq 0, \quad q_{ap}^{rs}(k) \geq 0, \quad y_{ap}^{rs}(k+1) \geq 0, \quad \forall a; p; k = 1, \cdots, K; \tag{6.35}$$

$$\bar{E}_p^{rs}(k+1) \geq 0, \qquad \forall p; k = 1, \cdots, K; \tag{6.36}$$

$$E_p^{rs}(1) = 0; \qquad \forall p; \tag{6.37}$$

$$y_{ap}^{rs}(1) = 0 \qquad \forall a, p. \tag{6.38}$$

For each O-D flow $f^{rs}(i)$ for each time interval $i = 1, \cdots, K$, the above subproblem can be further decomposed as follows, where each variable with index $i$ denotes the value caused by O-D flow $f^{rs}(i)$ for each interval $i = 1, \cdots, K$.

$$\min_{p,q,y,\bar{E}} \quad \sum_{k=1}^{K}\sum_{ap} \left[t_{1a}\, p_{ap}^{rs}(k,i) + t_{2a}\, q_{ap}^{rs}(k,i) \right.$$
$$\left. + \quad t_{3a}^{rs}\,(k+1) y_{ap}^{rs}(k+1, i)\right] \tag{6.39}$$

s.t.

$$y_{ap}^{rs}(k+1, i) = y_{ap}^{rs}(k, i) + p_{ap}^{rs}(k, i) - q_{ap}^{rs}(k, i) \qquad \forall a, p; k = 1, \cdots, K; \tag{6.40}$$

$$\bar{E}^{rs}(k+1,i) = \bar{E}^{rs}(k,i) + \sum_{a\in B(s)} \sum_{p} q_{ap}^{rs}(k,i) \qquad \forall k = 1,\cdots,K; \quad (6.41)$$

$$\sum_{a\in A(r)} \sum_{p} p_{ap}^{rs}(i,i) = f^{rs}(i) \qquad \forall k = i; \quad (6.42)$$

$$\sum_{a\in B(j)} q_{ap}^{rs}(k,i) - \sum_{a\in A(j)} p_{ap}^{rs}(k,i) = 0 \qquad \forall j,p; j \neq r,s; k = 1,\cdots,K; \quad (6.43)$$

$$\begin{aligned} y_{ap}^{rs}(k,i) &= \sum_{b\in\tilde{p}} \{y_{bp}^{rs}[k+\bar{\tau}_a(k),i] - y_{bp}^{rs}(k,i)\} \\ &+ \{\bar{E}_p^{rs}[k+\bar{\tau}_a(k),i] - \bar{E}_p^{rs}(k,i)\} \end{aligned}$$

$$\forall a \in B(j); j \neq r; p,r,s; k = 1,\cdots,K+1; \quad (6.44)$$

$$p_{ap}^{rs}(k,i) \geq 0, \quad q_{ap}^{rs}(k,i) \geq 0, \quad y_{ap}^{rs}(k+1,i) \geq 0, \quad \forall a,p; k = 1,\cdots,K; \quad (6.45)$$

$$\bar{E}_p^{rs}(k+1,i) \geq 0, \qquad \forall p; k = 1,\cdots,K; \quad (6.46)$$

$$E_p^{rs}(1,i) = 0; \qquad \forall p; \quad (6.47)$$

$$y_{ap}^{rs}(1,i) = 0 \qquad \forall a,p; \quad (6.48)$$

The above LP subproblem for each O-D flow $f^{rs}(i)$ for each time interval $i = 1,\cdots,K$ between each O-D pair $(r,s)$ can be viewed as a one-to-one minimal-cost route problem over an expanded time-space network using an artificial origin (see the next subsection and Section 6.2). It can be solved by determining the minimal-cost routes from the artificial origin to a super destination and completing an all-or-nothing assignment. By revising the costs for some artificial links, the minimal cost route is searched while the flow propagation constraints are automatically satisfied by construction of the links of the expanded time-space networks (see Section 6.2). The flow variables $p_{ap}^{rs}(k)$, $q_{ap}^{rs}(k)$, $y_{ap}^{rs}(k+1)$, $\bar{E}_p^{rs}(k+1)$ are determined by solving the minimal-cost route problems for all artificial origin-destination pairs between each original O-D pair $(r,s)$ for each time interval and assigning the O-D flows to the links.

In this combined algorithm, we define the diagonalization procedure as the outer iteration and the F-W procedure as the inner iteration. Denote the new solution at inner F-W iteration $(n+1)$ as

$$u_a^{(n+1)}(k) = u_a^{(n)}(k) + \alpha^{(n)}[u_a^{(n)}(k) - p_a^{(n)}(k)] \qquad \forall a; k = 1,\cdots,K; \quad (6.49)$$

$$v_a^{(n+1)}(k) = v_a^{(n)}(k) + \alpha^{(n)}[v_a^{(n)}(k) - q_a^{(n)}(k)] \qquad \forall a; k = 1,\cdots,K; \quad (6.50)$$

$$x_a^{(n+1)}(k) = x_a^{(n)}(k) + \alpha^{(n)}[x_a^{(n)}(k) - y_a^{(n)}(k)] \quad \forall a; k = 1,\cdots,K+1; \quad (6.51)$$

where $\alpha^{(n)}$ is the optimal step size of the one-dimensional search problem in the F-W algorithm. The one-dimensional search problem is to find step size $\alpha^{(n)}$ that solves

$$\min_{0\leq\alpha^{(n)}\leq 1} \quad \sum_{k=1}^{K}\sum_{a}\left\{\int_{0}^{u_a^{(n+1)}(k)} g_{1a}[x_a^{(n+1)}(k),\omega]d\omega + \int_{0}^{v_a^{(n+1)}(k)} g_{2a}[x_a^{(n+1)}(k),\omega]d\omega\right\} \tag{6.52}$$

where $u_a^{(n+1)}(k)$, $v_a^{(n+1)}(k)$, $x_a^{(n+1)}(k)$ must be substituted using the above definitional equations.

The algorithm for solving our instantaneous DUO route choice model is illustrated in the flowchart in Figure 6.1 and is summarized as follows:

**Step 0: Initialization.**

Find an initial feasible solution $\{x_a^{(1)}(k)\}$, $\{u_a^{(1)}(k)\}$, $\{v_a^{(1)}(k)\}$, $\{E^{(1)}(k)\}$. Set the outer iteration counter $m = 1$.

**Step 1: Diagonalization.**

Find a new estimate of the actual link travel time $\bar{\tau}_a^{(n)}(k)$ and solve the instantaneous DUO program. Set the inner iteration counter $n = 1$.

[Step 1.1]: *Update.* Calculate $t_{1a}(k)$, $t_{2a}(k)$ and $t_{3a}(k)$ using equations (6.22)-(6.25).

[Step 1.2]: *Direction Finding.* Based on $\{t_{1a}(k)\}$, $\{t_{2a}(k)\}$ and $\{t_{3a}(k)\}$ and satisfying the flow propagation constraints (6.44), search the minimal-cost route forward from each artificial origin to the super destination over an expanded time-space network for each O-D pair $(r, s)$. Perform an all-or-nothing assignment, yielding subproblem solution $\{p_a(k)\}, \{q_a(k)\}, \{y_a(k)\}, \{\bar{E}^{rs}(k)\}$.

[Step 1.3]: *Line Search.* Find the optimal step size $\alpha^{(n)}$ that solves the one-dimensional search problem.

[Step 1.4]: *Move.* Find a new solution by combining $\{u_a(k)\}$, $\{v_a(k)\}$, $\{x_a(k)\}$, $\{E^{rs}(k)\}$ and $\{p_a(k)\}$, $\{q_a(k)\}$, $\{y_a(k)\}$, $\{\bar{E}^{rs}(k)\}$.

[Step 1.5]: *Convergence Test for Inner Iterations.* If $n$ equals a prespecified number, go to step 2; otherwise, set $n = n+1$ and go to step 1.1.

**Step 2: Convergence Test for Outer Iterations.**

If $\bar{\tau}_a^{(m)}(k) \simeq \bar{\tau}_a^{(m+1)}(k)$, stop. The current solution, $\{u_a(k)\}$, $\{v_a(k)\}$, $\{x_a(k)\}$, $\{E^{rs}(k)\}$, is in a near instantaneous DUO state; otherwise, set $m = m + 1$ and go to step 1.

The number of inner iterations $n$ and the number of outer iterations $m$ are interrelated. If we set $m$ larger, then $n$ should be set smaller accordingly, and vice

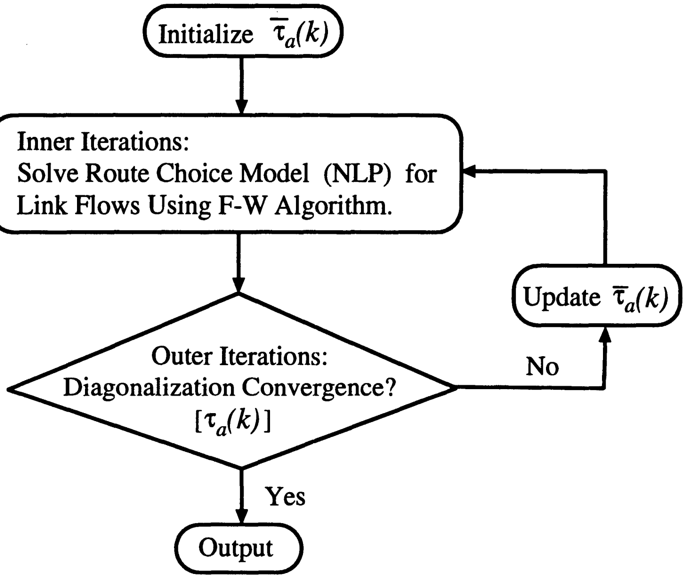



Figure 6.1: Flowchart of the Solution Algorithm

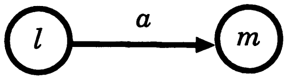
versa. In general, $n = 1$ to 3 is sufficient to speed up convergence. Similar experience with a diagonalization and F-W algorithm for solving static UO traffic equilibrium problems was reported by Mahmassani and Mouskos (1988). In order to speed up convergence, an incremental assignment technique is suggested for finding a good starting solution before the diagonalization procedure. Since the linear subproblem can be decomposed by each artificial origin-destination pair, this problem is a good candidate for solution with parallel computing techniques.

To speed up the serial computing speed, we can construct a super origin which connects all artificial origins. Thus, we can search the minimal-cost route forward from the super origin to the super destination over an expanded time-space network for each physical destination $s$. Therefore, the number of iterations within Step 1.2 can be reduced by the order of the number of origins. A significant saving of total CPU time can be achieved for a large network with many O-D pairs, although a marginal increase of CPU time is incurred due to the slight enlargement of the expanded time-space network. (A super origin is created and is connected with all artificial origins.)

### 6.1.3 Solving the LP Subproblem Using an Expanded Time-Space Network

In addition to flow propagation constraints, there are 3 types of constraints in the LP subproblem: link flow state equations, node flow conservation equations

Original Network:




Expanded Time-Space Network:

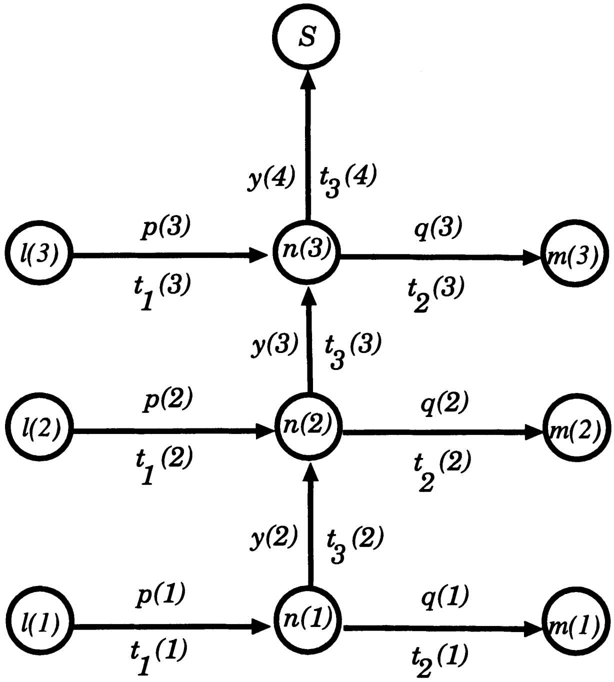



Figure 6.2: Expansion for Link $a$ for 3 Time Periods

and nonnegativity/boundary constraints. Since there are 3 variables associated with each physical link, we replace each link with 3 separate artificial links for each time period, by adding artificial nodes to define the new links. Figure 6.2 shows the expansion for link $a$ which points from $l$ to $m$ for 3 time periods. The initial state for link $a$ is assumed to be $x_a(1) = 0$. Note that a total of 9 nodes are required for the expansion of each physical link. The next section discusses how these new links and nodes can be numbered. If node $m$ is a destination, the expansion is shown in Figure 6.3. Figure 6.4 shows the expanded network for an example problem with 3 links, 3 nodes, 3 time periods, 3 origin (node 1) and 2 destinations (nodes 2 and 3). The initial state for each link is also assumed to be $x_a(1) = 0$ $(a = 1, 2, 3)$ and the instantaneous O-D trips are given as $f^{12}(k)$, $f^{13}(k)$, $k = 1, 2, 3$.

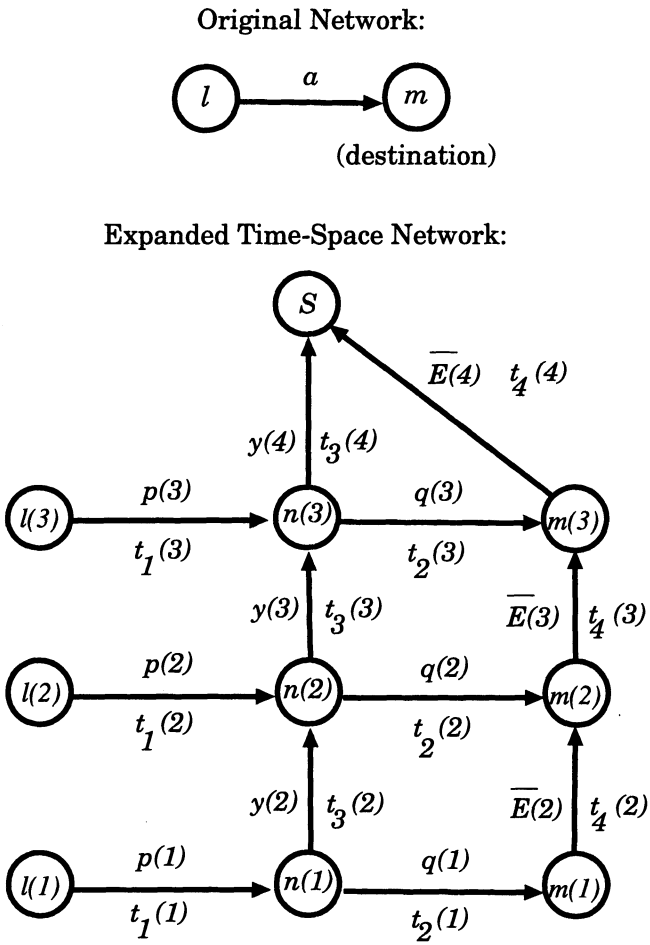



Figure 6.3: Expansion for Link $a$ for 3 Time Periods ($m$ is a destination)

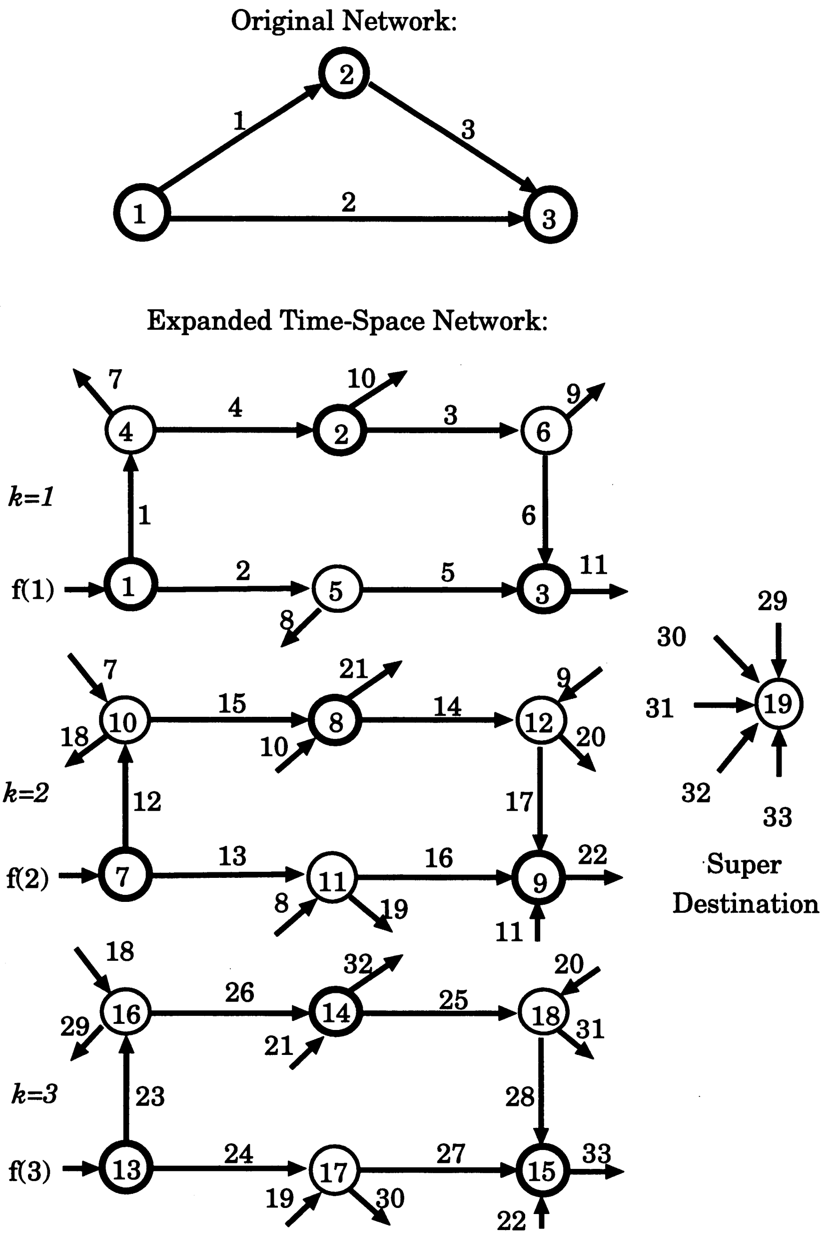



Figure 6.4: Expansion for 3-Link Network for 3 Time Periods

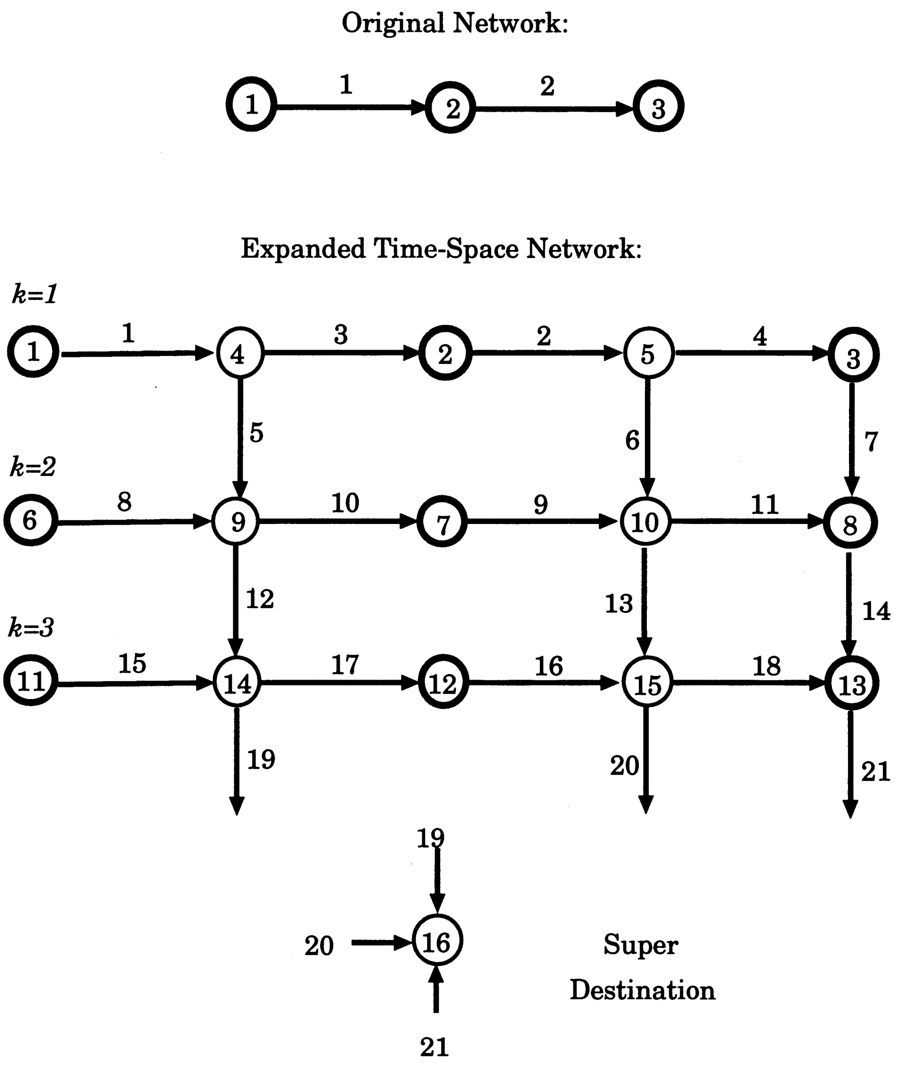



Figure 6.5: Expansion of Example Network (2 Links)

## 6.2 Time-Space Network Expansion

In the expansion of the time-space network, each physical node is expanded into additional nodes only (i.e., no extra links are introduced). However, since there are 3 separate variables per link, each link is expanded into 3 links, which requires additional nodes as well. Since the LP problem can be decomposed by each artificial origin-destination pair for all O-D pairs, we initially only consider a one-to-one network. The subproblem variables $p_{ap}^{rs}(k,i)$, $q_{ap}^{rs}(k,i)$, $y_{ap}^{rs}(k+1,i)$, $\bar{E}_p^{rs}(k+1,i)$ are viewed as flows over the artificial links of the expanded time-space network, which is constructed as follows.

1. **Nodes**: Each node $j$ is expanded to $K$ nodes, one for each time period. We denote these as $j(k)$, $k = 1, \cdots, K$, with new node numbers $j(k) = (N+A)(k-1)+j$. For example, we consider the time-space expansion of a 3-node, 2-link network for 3 time periods (Figure 6.5). The physical nodes in the original network are numbered as 1, 2, 3 and physical links are numbered as 1 = (1,2) and 2 = (2,3). Node 1 is an origin and node 3 is a destination. Thus, $N = 3$ and $A = 2$. For the expansion of physical nodes 1, 2, 3, the new node numbers are 1, 2, 3 for period 1; 6, 7, 8 for period 2; and 11, 12, 13 for period 3.

2. **Links**: Since there are 3 variables for each link, each link $a=(l,m)$ is expanded to $K$ nodes and $3K$ links, as follows:
   *a*) $K$ transshipment nodes $n_a(k)$, $k = 1, \cdots, K$, with new node numbers $n_a(k) = (N+A)(k-1)+N+a$. For the expansion of physical link 1 and 2 in Figure 6.5, the new transshipment node numbers are 4 and 5 for period 1, 9 and 10 for period 2, and 14 and 15 for period 3.
   *b*) $K$ links $(l(k), n_a(k))$, $k = 1, \cdots, K$, where each new link $\hat{a}$ has link number $[3A(k-1)+a]$, flow $p_a(k,i)$ and cost $t_{1a}(k)$. In Figure 6.5, these new links are 1 = (1,4) and 2 = (2,5) for period 1; 8 = (6,9) and 9 = (7,10) for period 2; and 15 = (11,14) and 16 = (12,15) for period 3.
   *c*) $K$ links $(n_a(k), m(k))$, $k = 1, \cdots, K$, where each new link $\hat{a}$ has link number $[3A(k-1)+A+a]$, flow $q_a(k,i)$ and cost $t_{2a}(k)$. In Figure 6.5, these new links are 3 = (4,2) and 4 = (5,3) for period 1; 10 = (9,7) and 11 = (10,8) for period 2; and 17 = (14,12) and 18 = (15,13) for period 3.
   *d*) $(K-1)$ links $(n_a(k), n_a(k+1))$, $k = 1, \cdots, K-1$, where each new link $\hat{a}$ has link number $[3A(k-1)+2A+a]$, flow $y_a^{rs}(k+1,i)$ and cost $t_{3a}(k+1)$ for corresponding O-D pair $(r,s)$. In Figure 6.5, these new links are 5 = (4,9) and 6 = (5,10) for period 1; 12 = (9,14) and 13 = (10,15) for period 2;
   *e*) one link $\hat{a} = (n_a(K), \mathcal{S})$ with new link number $[3A(K-1)+2A+a]$, flow $y_a^{rs}(K+1,i)$ and cost $\hat{t}_{\hat{a}}^{rs}$ for corresponding O-D pair $(r,s)$; in Figure 6.5, these new links are 19 = (14,16) and 20 = (15,16) for period 3.

3. **Origin Nodes**: Each origin node $r$ is expanded to $K$ origin nodes $r(k)$, $k = 1, \cdots, K$. In Figure 6.5, origin 1 is expanded to nodes 1, 6 and 11.

4. **Destination Nodes:** The destination node $s$ is expanded to:
   *a*) one super destination $\mathcal{S}$; in Figure 6.5, the physical destination node 3 is expanded to super destination 16;
   *b*) $K$ nodes $s(k)$, $k = 1, \cdots, K$; (nodes 3, 8 and 13 in Figure 6.5);
   *c*) $K$ links $\hat{a} = (s(k), s(k+1))$, $k = 1, \cdots, K-1$ and $\hat{a} = (s(K), \mathcal{S})$, with new link number $[3AK + k]$, flow $\bar{E}^{rs}(k+1, i)$ and cost $t_4^{rs}(k+1) = 0$ for corresponding artificial origin-destination pair for all O-D pair $(r, s)$; (links 7 = (3,8), 14 = (8,13), 21 = (13,16) for periods 1,2,3).

In summary, the expanded time-space network has $[(3A+1)K]$ links and $[(N+A)K+1]$ nodes. The expanded time-space network of the example (Figure 6.5) has 21 links and 16 nodes.

Link state equation (6.40) for each physical link $a$ can be viewed as a set of flow conservation equations for artificial nodes $n_a(k)$, $k = 1, \cdots, K$ in the expanded time-space network. For example, in Figure 6.2, the conservation of flow constraint for node $n(3)$ is $p(3, i) + y(3, i) = q(3, i) + y(4, i)$. This is equivalent to link state equation (6.40). Together with node flow conservation equations and nonnegativity constraints, these equations constitute the constraints for a one-to-one minimal-cost route problem with flow propagation constraints. Since the cost functions for the artificial links over the expanded time-space network are nonnegative, the original LP subproblem is transformed into a one-to-one minimal-cost route problem with flow propagation constraints.

The following explanation describes how the travel cost $\hat{t}_{\hat{a}}^{rs}$ for artificial link $\hat{a} = (n_a(K), \mathcal{S})$ is determined for corresponding O-D pair $(r, s)$. In the expanded time-space network, each dummy node $n_a(K)$, which is expanded from each link $a$ for the last time interval $K$, has a link pointing to the super destination $\mathcal{S}$. However, it is not always true that every physical link in the original network will lead to destination $s$. In order to represent this situation, an indicator variable $\delta_{\hat{a}}^s$ for each new link $\hat{a}$ in the expanded time-space network is defined as follows:

$$\delta_{\hat{a}}^s = \begin{cases} 1 & \text{if the expanded link } \hat{a} \text{ can be reached backwards} \\ & \text{from dummy node } s(K) \text{ expanded from destination } s \\ 0 & \text{otherwise} \end{cases}$$

Then, the travel cost $\hat{t}_{\hat{a}}^{rs}$ for dummy link $\hat{a} = (n_a(K), \mathcal{S})$ can be determined as:

$$\hat{t}_{\hat{a}}^{rs} = \begin{cases} t_{3a}(K+1) = 0 & \text{if} \quad \delta_{\hat{a}}^s = 1 \\ +\infty & \text{if} \quad \delta_{\hat{a}}^s = 0 \end{cases}$$

In this way, the property that flows cannot move backwards from destination $s$ to some links $\{a\}$ is also guaranteed in the expanded time-space network. For example, in Figure 6.4, when node 2 is a destination, link 27 and 28 are not reachable from node 14 (which is expanded from node 2 for period 3). Thus, $\delta_{27}^2 = 0$ and $\delta_{28}^2 = 0$ so that $\hat{t}_{27} = \infty$ and $\hat{t}_{28} = \infty$ when node 2 is a destination.

We now assume a *many-to-many* network. Since the original LP subproblem can be decomposed according to O-D pairs, the LP subproblem can be viewed as a set of minimal-cost route problems with propagation constraints. Minimal-cost routes are searched forward from each artificial origin to the super destination. Suppose there are $S$ destinations in a many-to-many network. Each destination node $s$ can be expanded to: one super destination $\mathcal{S}$; $K$ nodes $s(k)$, $k = 1, \cdots, K$; and $K$ links $\hat{a} = (s(k), s(k+1))$, $k = 1, \cdots, K-1$ and $\hat{a} = (s(K), \mathcal{S})$, with new link number $[3AK+(s-1)K+k]$, flow $\bar{E}^{rs}(k+1, i)$ and cost $\hat{t}_{\hat{a}}^{rs}(k+1)$.

For the many-to-many case, we use the same network notation for the expanded time-space network as for the one-to-one network so that the computational code can be simplified. Therefore, we use one super destination $\mathcal{S}$ to represent the super destinations expanded from all destinations $\{s\}$ and define the cost $\hat{t}_{\hat{a}}^{rs}(K+1)$ for each dummy link $\hat{a} = (s(K), s(k+1))$ $k = 1, \cdots, K-1$ and $\hat{a} = (s(K), \mathcal{S})$ as follows:

$$\hat{t}_{\hat{a}}^{rs}(k+1) = \begin{cases} t_{4}^{rs}(k+1) = 0 & \text{if minimal cost route between any artificial} \\ & \text{O-D pair } (r(k), \mathcal{S}) \text{ is searched} \\ +\infty & \text{otherwise} \end{cases}$$

where artificial O-D pair $(r(k), \mathcal{S})$ is expanded from original O-D pair $rs$ for time interval $k$. The above cost setting reflects that when considering artificial O-D pair $(r(k), \mathcal{S})$ for any O-D pair $rs$, other destinations $\hat{s}$ become intermediate nodes. There is no cumulative effect for these nodes $\hat{s}$. In summary, the expanded time-space network has $[(3A+S)K]$ links and $[(N+A)K+1]$ nodes for a many-to-many network.

## 6.3 Flow Propagation Constraints in Minimal-Cost Route Searches

In searching for the minimal-cost route between each artificial origin-destination pair $(r(k), \mathcal{S})$, flow propagation constraint (6.44) is automatically satisfied by temporally adjusting the costs for $2K$ artificial links $\hat{a} = (n_a(k), n_a(k+1))$, $k = 1, \cdots, K-1$, $\hat{a} = (n_a(K), \mathcal{S})$ and $\hat{a} = (n_a(k), m(k))$, $k = 1, \cdots, K$. These artificial links are expanded from original link $a=(l, m)$.

The temporal cost adjustment procedure begins when an artificial node $l(k)$ is being searched in the minimal-cost route search. This procedure is as follows:

*if* $i - 0.5 \leq \bar{\tau}_a(k) < i + 0.5$ $(i = 0, 1, \cdots, K)$,
*then the feasible subroute will be* $[l(k), n(k), \cdots, n(k+i), m(k+i)]$
*for the time-space subnetwork expanded from link* $a=(l, m)$.

This subroute is guaranteed to have the minimal cost by setting the costs of other artificial links in the time-space subnetwork (except links on the above subroute) temporally equal to infinity.

In Figure 6.3, we assume the minimal-cost route is searched from $l(1)$ to $\mathcal{S}$. If $\bar{\tau}_a(1) = 2$, the feasible subroute is $[l(1), n(1), n(2), n(3), m(3)]$. This cost adjustment procedure should be performed for each artificial origin-destination pair $(r(k), \mathcal{S})$.

## 6.4 Computational Experience

The algorithm was coded in FORTRAN and solved on a IBM 3090-300J and on a CRAY Y-MP/464. For the purpose of illustration, we give computational results for a small test problem with 12 arcs and 9 nodes. The network is shown in Figure 6.6. This problem required approximately 5 seconds on the IBM for 20 incremental iterations for the initial solution, 3 inner F-W iterations per outer iteration and 40 outer diagonalization iterations to converge. We have also solved a problem with 60 links, 36 nodes, 9 time intervals, 4 origins and 4 destinations on the CRAY using CFT77. The CPU time was approximately 68 seconds. The following link travel cost functions were used in the computations.

$$c_a(k) = g_{1a}(k) + g_{2a}(k)$$

$$g_{1a}(k) = \beta_{1a} + \beta_{2a}[u_a(k)]^2 + \beta_{3a}[x_a(k)]^2$$

$$g_{2a}(k) = \beta_{4a} + \beta_{5a}[v_a(k)]^2 + \beta_{6a}[x_a(k)]^2$$

The parameters for each link travel cost function are given in Table 6.1. The number of vehicles $x_a(1)$ on each link $a$ at initial time $k = 1$ is assumed to equal 0. Two O-D pairs and 7 time intervals are considered, and the corresponding trip table for each time interval is given in Table 6.2. The optimal link flow trajectories and corresponding optimal link travel costs are given in Table 6.3.

Table 6.3 shows that 10.0+10.0 = 20.0 vehicles enter the network during interval 1 and 5.0+5.0 = 10.0 enter during interval 2 (on links 1-2 and 1-4). For node 5, 18.1 vehicles enter links 2-5 and 4-5 during intervals 2-4. No vehicles remain on these 2 links at the end of interval 4, and 3.1 vehicles enter links 5-6 and 5-8 pointing out of node 5 continuing to node 9 during intervals 2-5. Thus, 18.1 − 3.1 = 15.0 vehicles exit the network at node 5 during intervals 1-5 so that the requirements for O-D pair 1-5 for intervals 1-5 are met. At node 9, 3.5 + 3.1 + 0.7 + 3.6 + 3.1 + 1.0 = 15.0 vehicles exit links 6-9 and 8-9 into node 9 during intervals 5-7. Note that no vehicles remain on links 6-9 and 8-9 at the end of interval 7 and no vehicles remain on other links on the network at the end of interval 6.

Inspection of Table 6.3 reveals that several links are being *used* in the sense defined in Chapter 5; also a few portions of routes are being used. In no case is an entire route *used* because of the short duration of the O-D flows. Even so, it is interesting to ask whether the instantaneous route travel times are equal.

Table 6.4 provides a comparison of instantaneous route travel times for the 2 O-D flows. The travel times for the 2 routes used from node 1 to node 5 are equal in each time interval. The travel times for the 6 routes from node 1 to node 9 are equal in each interval except intervals 3, 4 and 5. In interval 3, note that routes 1-4-7-8-9 and 1-2-3-6-9 have the minimal route travel costs for O-D pair 1-9, since they avoid the congestion at destination node 5. The other routes for O-D pair 1-9 have higher costs in interval 3 because of flows exiting the network from node 5. Moreover, the predominant flow from node 1 to node 9 avoids node 5 in interval 3. Likewise in intervals 3, 4 and 5, routes 1-4-7-8-9 and 1-2-3-6-9 with predominant flows have equal cost; because these flows are already on routes avoiding congestion at node 5, these 2 routes have slightly higher costs than the unused routes through node 5 in intervals 4 and 5. This simple example illustrates the inherent complexity of the dynamic route choice model, as compared with its static counterpart.

## 6.5 Notes

An algorithm for solving the instantaneous dynamic user-optimal route choice model was presented in this chapter. One significant aspect of our algorithm is that by using a time-space expanded network, the Frank-Wolfe LP subproblem requires only the solution of minimal-cost route problems for each O-D pair. This expansion technique allows standard algorithms for *static* traffic assignment to solve dynamic route choice models. Thus, the DUO route choice model and solution algorithm have an elegant correspondence with the static UO route choice model and its Frank-Wolfe algorithm. Therefore, it may be possible to extend other static UO formulations and solution algorithms to dynamic versions. This algorithm should be tested on a large-scale transportation network; other efficient algorithms, such as the algorithm of Leventhal et al (1973), also need to be investigated.

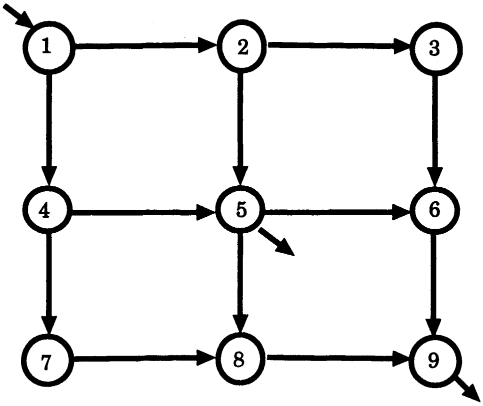



Figure 6.6: Test Network

Table 6.1: Parameters of Link Cost Functions

| link a | $\beta_{1a}$ | $\beta_{2a}$ | $\beta_{3a}$ | $\beta_{4a}$ | $\beta_{5a}$ | $\beta_{6a}$ |
|---|---|---|---|---|---|---|
| 1—2 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 2—3 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 1—4 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 2—5 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 3—6 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 4—5 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 5—6 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 4—7 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 5—8 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 6—9 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 7—8 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 8—9 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |

Table 6.2: O-D Trip Table for Each Time Interval $k$

| Time Interval $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $f^{15}(k)$ | 10. | 5. | 0. | 0. | 0. | 0. | 0. |
| $f^{19}(k)$ | 10. | 5. | 0. | 0. | 0. | 0. | 0. |

Table 6.3: Optimal Trajectories of Link Flows and Travel Times

| Interval k | Link a | Vehicles $x_a(k+1)$ | Inflow $u_a(k)$ | Exit Flow $v_a(k)$ | Vehicles $x_a(k)$ | Travel Time $c_a(k)$ |
|---|---|---|---|---|---|---|
| 1 | 1—2 | 10.0 | 10.0 | 0.0 | 0.0 | 1.10 |
| 2 | 1—2 | 7.9 | 5.0 | 7.1 | 10.0 | 1.99 |
| 3 | 1—2 | 1.7 | 0.0 | 6.2 | 7.9 | 1.70 |
| 4 | 1—2 | 0.0 | 0.0 | 1.7 | 1.7 | 1.05 |
| 5 | 1—2 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 6 | 1—2 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 1 | 2—3 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 2—3 | 3.0 | 3.0 | 0.0 | 0.0 | 1.01 |
| 3 | 2—3 | 2.6 | 2.6 | 3.0 | 3.0 | 1.16 |
| 4 | 2—3 | 0.4 | 0.4 | 2.6 | 2.6 | 1.12 |
| 5 | 2—3 | 0.0 | 0.0 | 0.4 | 0.4 | 1.00 |
| 6 | 2—3 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 1 | 1—4 | 10.0 | 10.0 | 0.0 | 0.0 | 1.10 |
| 2 | 1—4 | 7.9 | 5.0 | 7.1 | 10.0 | 1.99 |
| 3 | 1—4 | 1.6 | 0.0 | 6.2 | 7.9 | 1.71 |
| 4 | 1—4 | 0.0 | 0.0 | 1.6 | 1.6 | 1.04 |
| 5 | 1—4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 6 | 1—4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 1 | 2—5 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 2—5 | 4.2 | 4.2 | 0.0 | 0.0 | 1.02 |
| 3 | 2—5 | 3.6 | 3.6 | 4.2 | 4.2 | 1.31 |
| 4 | 2—5 | 1.2 | 1.2 | 3.6 | 3.6 | 1.22 |
| 5 | 2—5 | 0.0 | 0.0 | 1.2 | 1.2 | 1.03 |
| 6 | 2—5 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 1 | 3—6 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 3—6 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 3 | 3—6 | 3.0 | 3.0 | 0.0 | 0.0 | 1.01 |
| 4 | 3—6 | 2.6 | 2.6 | 3.0 | 3.0 | 1.16 |
| 5 | 3—6 | 0.4 | 0.4 | 2.6 | 2.6 | 1.12 |
| 6 | 3—6 | 0.0 | 0.0 | 0.4 | 0.4 | 1.00 |
| 1 | 4—5 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 4—5 | 4.2 | 4.2 | 0.0 | 0.0 | 1.02 |
| 3 | 4—5 | 3.6 | 3.6 | 4.2 | 4.2 | 1.32 |
| 4 | 4—5 | 1.3 | 1.3 | 3.6 | 3.6 | 1.23 |
| 5 | 4—5 | 0.0 | 0.0 | 1.3 | 1.3 | 1.03 |
| 6 | 4—5 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |

Table 6.3: Optimal Trajectories of Link Flows and Travel Times (continued)

| Interval k | Link a | Vehicles $x_a(k+1)$ | Inflow $u_a(k)$ | Exit Flow $v_a(k)$ | Vehicles $x_a(k)$ | Travel Time $c_a(k)$ |
|---|---|---|---|---|---|---|
| 1 | 5—6 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 5—6 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 3 | 5—6 | 0.5 | 0.5 | 0.0 | 0.0 | 1.00 |
| 4 | 5—6 | 0.5 | 0.5 | 0.5 | 0.5 | 1.01 |
| 5 | 5—6 | 0.2 | 0.2 | 0.5 | 0.5 | 1.00 |
| 6 | 5—6 | 0.0 | 0.0 | 0.2 | 0.2 | 1.00 |
| 1 | 4—7 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 4—7 | 2.9 | 2.9 | 0.0 | 0.0 | 1.01 |
| 3 | 4—7 | 2.6 | 2.6 | 2.9 | 2.9 | 1.15 |
| 4 | 4—7 | 0.3 | 0.3 | 2.6 | 2.6 | 1.11 |
| 5 | 4—7 | 0.0 | 0.0 | 0.3 | 0.3 | 1.00 |
| 6 | 4—7 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 1 | 5—8 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 5—8 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 3 | 5—8 | 0.7 | 0.7 | 0.0 | 0.0 | 1.00 |
| 4 | 5—8 | 0.5 | 0.5 | 0.7 | 0.7 | 1.01 |
| 5 | 5—8 | 0.7 | 0.7 | 0.5 | 0.5 | 1.01 |
| 6 | 5—8 | 0.0 | 0.0 | 0.7 | 0.7 | 1.01 |
| 1 | 6—9 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 6—9 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 3 | 6—9 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 4 | 6—9 | 3.5 | 3.5 | 0.0 | 0.0 | 1.01 |
| 5 | 6—9 | 3.1 | 3.1 | 3.5 | 3.5 | 1.22 |
| 6 | 6—9 | 0.7 | 0.7 | 3.1 | 3.1 | 1.16 |
| 7 | 6—9 | 0.0 | 0.0 | 0.7 | 0.7 | 1.01 |
| 1 | 7—8 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 7—8 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 3 | 7—8 | 2.9 | 2.9 | 0.0 | 0.0 | 1.01 |
| 4 | 7—8 | 2.6 | 2.6 | 2.9 | 2.9 | 1.15 |
| 5 | 7—8 | 0.3 | 0.3 | 2.6 | 2.6 | 1.11 |
| 6 | 7—8 | 0.0 | 0.0 | 0.3 | 0.3 | 1.00 |
| 1 | 8—9 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 8—9 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 3 | 8—9 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 4 | 8—9 | 3.6 | 3.6 | 0.0 | 0.0 | 1.01 |
| 5 | 8—9 | 3.1 | 3.1 | 3.6 | 3.6 | 1.23 |
| 6 | 8—9 | 1.0 | 1.0 | 3.1 | 3.1 | 1.17 |
| 7 | 8—9 | 0.0 | 0.0 | 1.0 | 1.0 | 1.02 |

Table 6.4: Comparison of Instantaneous Route Travel Times

| Interval | Routes from 1 to 5 | |
|---|---|---|
| k | 1-2-5 | 1-4-5 |
| 1 | 2.10 | 2.10 |
| 2 | 3.01 | 3.01 |
| 3 | 3.01 | 3.03 |
| 4 | 2.27 | 2.27 |
| 5 | 2.03 | 2.03 |
| 6 | 2.00 | 2.00 |

| Interval | Routes from 1 to 9 | | | | | |
|---|---|---|---|---|---|---|
| k | 1-4-7-8-9 | 1-4-5-8-9 | 1-4-5-6-9 | 1-2-3-6-9 | 1-2-5-6-9 | 1-2-5-8-9 |
| 1 | 4.10 | 4.10 | 4.10 | 4.10 | 4.10 | 4.10 |
| 2 | 5.00 | 5.01 | 5.01 | 5.00 | 5.01 | 5.01 |
| 3 | 4.87 | 5.03 | 5.03 | 4.87 | 5.01 | 5.01 |
| 4 | 4.31 | 4.29 | 4.29 | 4.34 | 4.29 | 4.29 |
| 5 | 4.34 | 4.27 | 4.25 | 4.34 | 4.25 | 4.27 |
| 6 | 4.17 | 4.18 | 4.16 | 4.16 | 4.16 | 4.18 |
| 7 | 4.02 | 4.02 | 4.01 | 4.01 | 4.01 | 4.02 |

# Chapter 7

# An Ideal Dynamic User-Optimal Route Choice Model

In this chapter, we present an *ideal* dynamic user-optimal route choice model for a network with multiple origin-destination pairs. The model extends our previous *instantaneous* DUO route choice model in an important respect: route equilibrium is based on actual travel times rather than instantaneous travel times at the time of the choice. In Section 7.1, additional network flow constraints and the definition of ideal DUO state are presented. The equivalent equality constraints of the ideal DUO route choice conditions are developed in Section 7.2. Then, an optimal control formulation of the travel-time-based ideal DUO route choice problem is presented in Section 7.3. In Section 7.4, this model is reformulated as a discrete time NLP. Subsequently, penalty and diagonalization/Frank-Wolfe methods are suggested to solve this NLP.

## 7.1 Additional Network Flow Constraints and Definition of the Ideal DUO State

For the formulation of travel-time-based ideal DUO route choice model, we need to add more route flow conservation constraints. Assume there are $P$ routes from origin $r$ to destination $s$ (these can be generated as needed). Denote indicator parameters as

$$\delta_{ap}^{rs} = \begin{cases} 1 & \text{if link } a \text{ is on route } p \text{ between O-D pair } (r, s) \\ 0 & \text{otherwise.} \end{cases}$$

Flow conservation at origin node $r$ relates departure rates ($f^{rs}(t)$ and $f_p^{rs}(t)$) to the flow entering each link emanating from the origin. These flow conservation equations for origin $r$ can be expressed as

$$f_p^{rs}(t) = \sum_{a \in A(r)} \delta_{ap}^{rs} u_{ap}^{rs}(t) \qquad \forall p, r, s; r \neq s; \tag{7.1}$$

$$\sum_p f_p^{rs}(t) = f^{rs}(t) \qquad \forall r, s; r \neq s. \tag{7.2}$$

Denote the *cumulative* number of vehicles departing from origin $r$ to destination $s$ from time 0 to $t$ as the state variable $F^{rs}(t)$. Also, $F_p^{rs}(t)$ denotes the cumulative number of departing vehicles from origin $r$ toward destination $s$ along route $p$ by time $t$. Then, we have an additional state equation for each origin $r$

$$\frac{dF_p^{rs}(t)}{dt} = f_p^{rs}(t) \qquad \forall p, r \neq s, s. \tag{7.3}$$

Also, at initial time $t = 0$,

$$F_p^{rs}(0) = 0, \qquad \forall p, r, s. \tag{7.4}$$

Denote the instantaneous flow rate *arriving at* destination node $s$ from origin node $r$ at time $t$ as $e^{rs}(t)$, which is also a control variable. Control variable $e_p^{rs}(t)$ denotes the arrival rate on route $p$. Flow conservation at destination node $s$ relates arriving flows ($e^{rs}(t)$ and $e_p^{rs}(t)$) to the flows exiting each link leading to destination $s$ at time $t$. Thus, the flow conservation equations for destination $s$ can be expressed as

$$e_p^{rs}(t) = \sum_{a \in B(s)} \delta_{ap}^{rs} v_{ap}^{rs}(t) \qquad \forall p, r, s; s \neq r; \tag{7.5}$$

$$\sum_p e_p^{rs}(t) = e^{rs}(t) \qquad \forall r, s; s \neq r. \tag{7.6}$$

Denote the *cumulative* number of vehicles arriving at destination $s$ from origin $r$ by time $t$ as the state variable $E^{rs}(t)$; $E_p^{rs}(t)$ denotes the cumulative number of vehicles arriving at destination $s$ from origin $r$ along route $p$ by time $t$. Thus, we have an additional state equation for each destination $s$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r, s \neq r. \tag{7.7}$$

At the initial time $t = 0$,

$$E_p^{rs}(0) = 0 \qquad \forall p, r, s. \tag{7.8}$$

These variables must be nonnegative at all times:

$$E_p^{rs}(t) \geq 0, \quad F_p^{rs}(t) \geq 0, \quad e_p^{rs}(t) \geq 0, \quad f_p^{rs}(t) \geq 0 \quad \forall p, r, s. \tag{7.9}$$

Now we define the ideal DUO route choice problem. The dynamic user-optimal route choice problem is to find the dynamic trajectories of link states and inflow and exit flow control variables, given the network, the link travel time functions and the time-dependent O-D departure rate requirements. In Chapter 5, we defined the link-time-based instantaneous DUO route choice state as follows.

**Link-Time-Based Instantaneous DUO State:** *If, for each O-D pair at each decision node at each instant of time, the instantaneous travel times to the destination over all routes that are being used equal the minimal instantaneous route travel time, the dynamic traffic flow over the network is in a link-time-based instantaneous dynamic user-optimal state.*

In this earlier definition of DUO, the instantaneous user-optimal travel times for all routes that are being used are equal at each decision node at each instant of time. The corresponding model provides the currently prevailing traffic information to travelers. However, route flows with the same departure time and the same origin-destination may actually experience somewhat different route travel times, because the route time may subsequently change due to changing network traffic conditions, even though at each decision node the flows select the route that is currently best. Therefore, in this chapter we propose an alternative definition of DUO that reflects the *ideal* route choice behavior of travelers. The formulation of the problem is based on the underlying choice criterion that each traveler uses the route that minimizes his/her *actual* travel time when departing from the origin or any intermediate node to his/her destination.

**Travel-Time-Based Ideal DUO State:** *If, for each O-D pair at each instant of time, the* actual *travel times experienced by travelers departing at the same time are equal and minimal, the dynamic traffic flow over the network is in a travel-time-based* ideal *dynamic user-optimal state.*

The above definition can also be called a predictive DUO model, since the actual route travel time is predicted using the corresponding route choice model. This model assumes each traveler will have perfect information about the future network conditions and will comply with the guidance instructions based on ideal DUO route choice conditions. Travelers will not regret what decisions they have made before their journeys.

In this ideal DUO route choice problem, a route $p$ between $r$ and $s$ is defined as being used at time $t$ if $f_p^{rs}(t) > 0$. This is a less restrictive definition than for the case of instantaneous DUO in Chapter 5 and is consistent with the general definition of used routes in the variational inequality models in Chapter 13. We ensure that the above ideal DUO route choice conditions are satisfied through explicit equality constraint conditions in the following section.

## 7.2 Equivalent Equality Constraints for Ideal DUO Route Choice Conditions

Define $\eta_p^{rs*}(t)$ as the travel time *actually* experienced over route $p$ by vehicles departing origin $r$ toward destination $s$ at time $t$. Also denote $\pi^{rs}(t)$ as the minimal travel time experienced by vehicles departing from origin $r$ to destination $s$ at time $t$. $\pi^{rs}(t)$ is a functional of all link flow variables at time $\omega \geq t$, i.e., $\pi^{rs}(t) = \pi^{rs}[u(\omega), v(\omega), x(\omega)|\omega \geq t]$. This functional is neither a state variable nor a control variable, and it is not fixed; moreover, it is not available in closed form. Nevertheless, it can be evaluated when $u(\omega)$, $v(\omega)$ and $x(\omega)$ are temporarily fixed, as in a Frank-Wolfe algorithm, which is all that is required for solving the model.

The travel-time-based ideal DUO route choice conditions can then be expressed as follows:

$$\eta_p^{rs^*}(t) \geq \pi^{rs^*}(t) \qquad \forall p, r, s; \tag{7.10}$$

$$f_p^{rs^*}(t)\left[\eta_p^{rs^*}(t) - \pi^{rs^*}(t)\right] = 0 \qquad \forall p, r, s; \tag{7.11}$$

$$f_p^{rs}(t) \geq 0 \qquad \forall p, r, s. \tag{7.12}$$

The asterisk in the above equations denotes that the flow variables are the optimal solutions under the travel-time-based ideal DUO state. For any O-D pair $(r, s)$, if there is a positive inflow over route $p$, i.e., $f_p^{rs^*}(t) \geq 0$, equation (7.11) requires that

$$\eta_p^{rs^*}(t) = \pi^{rs^*}(t) \qquad \forall p, r, s. \tag{7.13}$$

Thus, route inflow $f_p^{rs^*}(t)$ uses the minimal actual travel time $\pi^{rs^*}(t)$. If the inflow over route $p$ is zero, i.e., $f_p^{rs^*}(t) = 0$, equation (7.11) requires that $[\eta_p^{rs^*}(t) - \pi^{rs^*}(t)]$ be either zero or positive (by equation (7.10)). In other words, route $p$ has either the minimal travel time or higher travel time at time $t$. On the other hand, if route $p$ has higher travel time at time $t$, i.e., $\eta_p^{rs^*}(t) > \pi^{rs^*}(t)$, equation (7.11) requires that route $p$ has zero inflow at time $t$ ($f_p^{rs^*}(t) = 0$). By transforming the above inequality DUO route choice conditions into equivalent equality constraints for cumulative departures/arrivals and route flows, we formulate an optimal control program in Section 7.3.

Next, we discuss two different approaches for computing the minimal travel time $\pi^{rs}(t)$. The first method is to compute link travel times and use a recursive formula to compute the route travel time $\eta_p^{rs}(t)$ for all allowable routes. Assume route $p$ consists of nodes $(r, 1, 2, \cdots, i, \cdots, s)$. Denote $\eta_p^{ri}(t)$ as the travel time *actually* experienced over route $p$ from origin $r$ to node $i$ by vehicles departing origin $r$ at time $t$. Then, a recursive formula for route travel time $\eta_p^{rs}(t)$ is:

$$\eta_p^{ri}(t) = \eta_p^{r(i-1)}(t) + \tau_a[t + \eta_p^{r(i-1)}(t)] \qquad \forall p, r, i; i = 1, 2, \cdots, s;$$

where link $a = (i-1, i)$. Then, $\pi^{rs}(t) = \min_p \eta_p^{rs}(t)$. This is the conventional approach discussed in Chapter 4.

A second method for computing $\pi^{rs}(t)$ avoids enumerating routes; therefore, we prefer this approach in this chapter. In a travel-time-based ideal DUO state, the cumulative number of vehicles departing from origin $r$ by time $t$ must equal the number of vehicles arriving at destination $s$ by time $t+\pi^{rs}(t)$, regardless of their routes taken. It follows that

$$\int_o^t f^{rs}(\omega)d\omega = \int_o^{t+\pi^{rs}(t)} e^{rs}(\omega)d\omega \qquad \forall r, s; \tag{7.14}$$

or

$$F^{rs}(t) = E^{rs}[t+\pi^{rs}(t)] \qquad \forall r, s. \tag{7.15}$$

The relationship of $F^{rs}(t)$ and $E^{rs}(t)$ to $\pi^{rs}(t)$ is shown in Figure 7.1. Recall that the flow propagation constraints ensure that $e^{rs}(t)$ cannot become prematurely positive. The asterisk is ignored in the following derivation of the equality constraints for the travel-time-based ideal DUO route choice conditions.

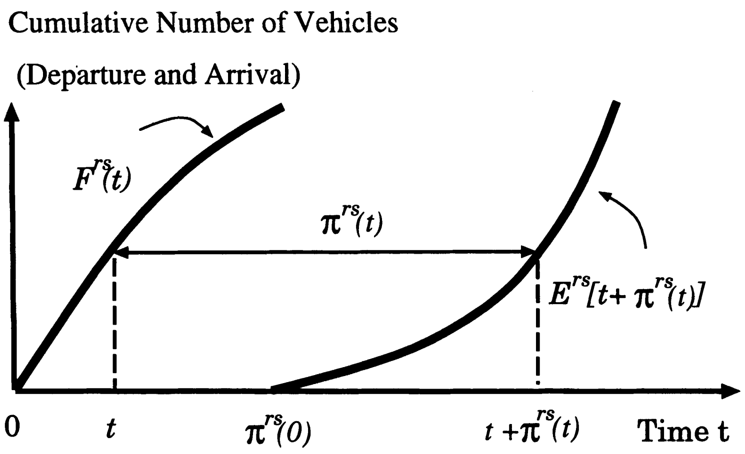



Figure 7.1: Relationship of $F^{rs}(t)$ and $E^{rs}(t)$ to $\pi^{rs}(t)$

In the route choice problem, the cumulative number of departing vehicles $F^{rs}(t)$ is given, since the $f^{rs}(t)$ are exogenous in the route choice problem. Also, the cumulative number of arriving vehicles $E^{rs}(t)$ can be computed by equations (7.5)-(7.7). Then, from (7.15), the following numerical procedure within a diagonalization technique can be used to compute $\pi^{rs}(t)$. This involves estimating $\pi^{rs}(t)$ and calculating $F^{rs}(t)$ and $E^{rs}(t)$. If $F^{rs}(t) \neq E^{rs}[t+\pi^{rs}(t)]$, we increase or decrease $\pi^{rs}(t)$ until convergence occurs. We discuss this further in Section 7.4.

Next we consider the constraints for departure rate and arrival rate over *each route*. Consider a small time interval $[t, t+\Delta t]$. The number of vehicles departing along each route $p$ during this interval is

$$F_p^{rs}(t+\Delta t) - F_p^{rs}(t) \qquad \forall p, r, s.$$

The travel-time-based ideal DUO route choice conditions require that all vehicles departing during time interval $[t, t+\Delta t]$ spend the minimal O-D travel times and arrive at destination $s$ within time interval $[t+\pi^{rs}(t), (t+\Delta t)+\pi^{rs}(t+\Delta t)]$, no matter which route they choose. The number of arrivals along each route $p$ during this interval is

$$E_p^{rs}[(t+\Delta t)+\pi^{rs}(t+\Delta t)] - E_p^{rs}(t+\pi^{rs}(t)) \qquad \forall p, r, s. \tag{7.16}$$

Therefore, under the travel-time-based ideal DUO route choice conditions, the total number of vehicles departing from origin $r$ along route $p$ during time interval $[t, t+\Delta t]$ must equal the total number of vehicles arriving at destination $s$ along route $p$ during time interval $[t+\pi^{rs}(t), (t+\Delta t)+\pi^{rs}(t+\Delta t)]$. It follows that

$$F_p^{rs}(t+\Delta t) - F_p^{rs}(t) = E_p^{rs}[(t+\Delta t)+\pi^{rs}(t+\Delta t)] - E_p^{rs}(t+\pi^{rs}(t)) \qquad \forall p, r, s. \tag{7.17}$$

A simple network with 2 links and 2 nodes illustrates the above equations (see Figure 7.2). Suppose there are 4 vehicles departing from the origin in the period 8:00–8:01 AM. Thus, the departure time is $t = 8{:}00$ and the time interval $\Delta t$ equals 1 minute. Assume $\pi^{rs}(t) = 10$ minutes and $\pi^{rs}(t+\Delta t) = 11$ minutes. Then, these 4 vehicles must arrive at the destination in the period 8:10 to 8:12, regardless of the routes taken. Also, assume that 2 of the 4 vehicles use link 1, and the 2 other vehicles use link 2. The travel-time-based ideal DUO route choice conditions require for each route (link 1 or 2) that the number of arrivals in 8:10–8:12 AM must equal the number of departures in 8:00–8:01 AM, regardless of the routes taken.

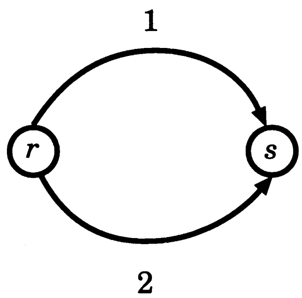



Figure 7.2: Two-Link Network

Consider the limit of the following equation taken from equation (7.17):

$$\frac{E_p^{rs}[(t+\Delta t)+\pi^{rs}(t+\Delta t)] - E_p^{rs}(t+\pi^{rs}(t))}{\Delta t} = \frac{F_p^{rs}(t+\Delta t) - F_p^{rs}(t)}{\Delta t} \qquad \forall p, r, s. \tag{7.18}$$

Assume that the actual travel time $\pi^{rs}(t)$ is differentiable with respect to $t$. Expanding $\pi^{rs}(t+\Delta t)$ in a Taylor series about the point $t$ gives

$$\pi^{rs}(t+\Delta t) = \pi^{rs}(t) + \dot{\pi}^{rs}(t)\ \Delta t + \mathbf{O}(\Delta t) \qquad \forall r, s, \tag{7.19}$$

where $\dot{\pi}^{rs}(t)$ denotes the derivative of $\pi^{rs}(t)$ with respect to time $t$, and $\mathbf{O}(\Delta t)$ denotes terms in the expansion of order two and greater in $\Delta t$. These latter terms are smaller in magnitude than $\Delta t$ as $\Delta t$ approaches zero. Using equation (7.19) to substitute for $\pi^{rs}(t+\Delta t)$ in the first term in equation (7.18), we have

$$\begin{aligned} & E_p^{rs}[(t+\Delta t) + \pi^{rs}(t+\Delta t)] \\ &= E_p^{rs}\ \{(t+\Delta t) + (\pi^{rs}(t) + \dot{\pi}^{rs}(t)\ \Delta t + \mathbf{O}(\Delta t))\} \\ &= E_p^{rs}\ \{[t+\pi^{rs}(t)] + [1+\dot{\pi}^{rs}(t)]\ \Delta t + \mathbf{O}(\Delta t)\} \qquad \forall p, r, s. \end{aligned}$$

Also, using $X \equiv t + \pi^{rs}(t)$ and $\Delta X \equiv [1 + \dot{r}^{rs}(t)]\ \Delta t + \mathbf{O}(\Delta t)$ in a Taylor series expansion of $F(X + \Delta X)$ about the point $X$ for the right-hand-side of the above equation, we have

$$\begin{aligned} & E_p^{rs}\ \{[t+\pi^{rs}(t)] + [1+\dot{r}^{rs}(t)]\ \Delta t + \mathbf{O}(\Delta t)\} = \\ & \qquad E_p^{rs}[t+\pi^{rs}(t)] + \dot{E}_p^{rs}[t+\pi^{rs}(t)]\ \ \{[1+\dot{\pi}^{rs}(t)]\ \Delta t + \mathbf{O}(\Delta t)\} \\ & + \quad \mathbf{O}\{[1+\dot{\pi}^{rs}(t)]\ \Delta t + \mathbf{O}(\Delta t)\} \qquad \forall p, r, s. \end{aligned}$$

Using

$$\dot{E}_p^{rs}[t+\pi^{rs}(t)] = e_p^{rs}[t+\pi^{rs}(t)]$$

we obtain

$$\begin{aligned} & E_p^{rs}[(t+\Delta t) + \pi^{rs}(t+\Delta t)] = \\ & \qquad E_p^{rs}[t+\pi^{rs}(t)] + e_p^{rs}[t+\pi^{rs}(t)]\ \ \{[1+\dot{\pi}^{rs}(t)]\ \Delta t + \mathbf{O}(\Delta t)\} \\ & + \quad \mathbf{O}\{[1+\dot{\pi}^{rs}(t)]\ \Delta t + \mathbf{O}(\Delta t)\} \qquad \forall p, r, s. \end{aligned} \tag{7.20}$$

Thus, by substituting the right-hand-side of equation (7.20) into equation (7.18), the left-hand-side of equation (7.18) becomes

$$e_p^{rs}[t+\pi^{rs}(t)]\left\{[1+\dot{\pi}^{rs}(t)] + \frac{\mathbf{O}(\Delta t)}{\Delta t}\right\} + \frac{\mathbf{O}\{[1+\dot{\pi}^{rs}(t)]\ \Delta t + \mathbf{O}(\Delta t)\}}{\Delta t} \tag{7.21}$$

When $\Delta t \to 0$, $\mathbf{O}(\Delta t) \to 0$ and $\{[1+\dot{\pi}^{rs}(t)]\ \Delta t + \mathbf{O}(\Delta t)\} \to 0$ so that $\mathbf{O}\{[1+\dot{\pi}^{rs}(t)]\ \Delta t + \mathbf{O}(\Delta t)\} \to 0$. Allowing $\Delta t \to 0$ in (7.21), by definition of the derivative, the left-hand-side equals $e_p^{rs}[t+\pi^{rs}(t)]\ [1+\dot{\pi}^{rs}(t)]$, and the right-hand-side of (7.18) equals $f_p^{rs}(t)$. It follows that as $\Delta t \to 0$, the following constraints are equivalent to (7.17)

$$e_p^{rs}[t+\pi^{rs}(t)] \cdot [1+\dot{\pi}^{rs}(t)] = f_p^{rs}(t) \qquad \forall p, r, s. \tag{7.22}$$

The term $\dot{\pi}^{rs}(t)$ is the rate of change of minimal O-D actual travel time. When $\dot{\pi}^{rs}(t) = 0$, the minimal O-D actual travel time is constant. Constraints (7.22) require that any route departure flow $f_p^{rs}(t)$ at time $t$ must use the O-D specific minimal travel time $\pi^{rs}(t)$ so as to arrive at the destination at time $t + \pi^{rs}(t)$. Furthermore, since cumulative departures at time $t$ equal cumulative arrivals at time $t + \pi^{rs}(t)$ for each O-D pair $rs$, all departure flows necessarily use minimal time routes at any time $t$. Thus, using constraints (7.15) and (7.22), the route choice model will generate traffic flows which satisfy travel-time-based ideal DUO route choice conditions (7.10)-(7.12). These two constraints (7.15) and (7.22) are one of the main contributions of this chapter. They make our ideal DUO route choice model distinct from other formulations.

Furthermore, we demonstrate in the following that constraints (7.22) can be derived from constraints (7.15) and definitional constraints as follows:

$$\sum_p F_p^{rs}(t) = F^{rs}(t), \quad \sum_p E_p^{rs}[t + \pi^{rs}(t)] = E^{rs}[t + \pi^{rs}(t)], \quad \forall r, s. \quad (7.23)$$

Note the second definitional constraint is included in the definitional constraint

$$\sum_p E_p^{rs}(t) = E^{rs}(t), \qquad \forall r, s \qquad (7.24)$$

when the assignment period $T$ is long enough to clear the traffic flow considered in the analysis. Substituting the left-hand-sides of equations (7.23) into equation (7.15),

$$\sum_p \{F_p^{rs}(t) - E_p^{rs}[t + \pi^{rs}(t)]\} = 0, \qquad \forall r, s. \qquad (7.25)$$

By definition, $\pi^{rs}(t) \leq \eta_p^{rs}(t)$ for all routes $p$. By the flow propagation constraint, for each route $p$ and each O-D pair $rs$ at any time $t$, the cumulative number of vehicles having arrived at the destination $s$ by time $[t + \pi^{rs}(t)]$ must be equal to or smaller than the cumulative number of vehicles that has departed from origin $r$ by time $t$. It follows that

$$F_p^{rs}(t) \geq E_p^{rs}[t + \pi^{rs}(t)], \quad \text{or} \quad F_p^{rs}(t) - E_p^{rs}[t + \pi^{rs}(t)] \geq 0, \quad \forall r, s, p \quad (7.26)$$

where time $t$ applies to any instant from 0 to $T$. Combining equations (7.25) and (7.26),

$$F_p^{rs}(t) - E_p^{rs}[t + \pi^{rs}(t)] = 0, \qquad \forall r, s, p. \qquad (7.27)$$

Taking derivatives of the above equations by using the chain rule (note that since $F_p^{rs}(t)$ is the integral of $f_p^{rs}(t)$, $\dot{F}_p^{rs}(t) = f_p^{rs}(t)$), we have

$$f_p^{rs}(t) - e_p^{rs}[t + \pi^{rs}(t)]\,[1 + \dot{\pi}^{rs}(t)] = 0 \qquad \forall r, s, p. \qquad (7.28)$$

Thus, the above constraint is redundant when constraints (7.15) and (7.23) are enforced. Therefore, we obtain equality constraints (7.15) and (7.23) which are equivalent to the travel-time-based ideal DUO route choice conditions.

## 7.3 An Optimal Control Model of Ideal DUO Route Choice

Using optimal control theory, the travel-time-based ideal dynamic user-optimal route choice problem is formulated as follows.

$$\min_{u,v,x,e,E,f_p^{rs},F_p^{rs};\pi} \int_0^T \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t),\omega]d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t),\omega]d\omega \right\} dt \qquad (7.29)$$

s.t.

Relationships between state and control variables:

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \qquad (7.30)$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall r, s, p; \qquad (7.31)$$

$$\frac{dF_p^{rs}(t)}{dt} = f_p^{rs}(t) \qquad \forall r, s, p; \qquad (7.32)$$

Flow conservation constraints:

$$f_p^{rs}(t) = \sum_{a\in A(r)} \delta_{ap}^{rs} u_{ap}^{rs}(t) \qquad \forall p, r, s; \qquad (7.33)$$

$$e_p^{rs}(t) = \sum_{a\in B(s)} \delta_{ap}^{rs} v_{ap}^{rs}(t) \qquad \forall p, r, s; \qquad (7.34)$$

$$\sum_{a\in B(j)} v_{ap}^{rs}(t) = \sum_{a\in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \qquad (7.35)$$

Constraints equilibrating *actual* route travel times:

$$F^{rs}(t) = E^{rs}[t + \pi^{rs}(t)] \qquad \forall r, s; \qquad (7.36)$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b\in \tilde{p}} \{x_{bp}^{rs}[t + \tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall r, s, p, j; a \in B(j); j \neq r; \qquad (7.37)$$

Definitional constraints:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \qquad (7.38)$$

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \qquad \sum_{rs} x_a^{rs}(t) = x_a(t), \qquad \forall a, r, s; \tag{7.39}$$

$$\sum_p E_p^{rs}(t) = E^{rs}(t), \qquad \sum_p F_p^{rs}(t) = F^{rs}(t), \qquad \forall r, s; \tag{7.40}$$

$$\sum_p f_p^{rs}(t) = f^{rs}(t), \qquad \sum_p e_p^{rs}(t) = e^{rs}(t), \qquad \forall r, s; \tag{7.41}$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \qquad u_{ap}^{rs}(t) \geq 0, \qquad v_{ap}^{rs}(t) \geq 0 \qquad \forall a, p, r, s; \tag{7.42}$$

$$e_p^{rs}(t) \geq 0, \quad f_p^{rs}(t) \geq 0, \quad E_p^{rs}(t) \geq 0, \quad F_p^{rs}(t) \geq 0 \qquad \forall p, r, s; \tag{7.43}$$

Boundary conditions:

$$E_p^{rs}(0) = 0, \qquad F_p^{rs}(0) = 0 \quad \forall p, r, s; \qquad x_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s. \tag{7.44}$$

In program (7.29)-(7.44), the *route-specific* departure variables $f_p^{rs}(t)$ and $F_p^{rs}(t)$ must be determined. The objective function is similar to the objective function of the well-known static user-optimal (UO) model. We note that other objective functions can also be used since constraints (7.36) and (7.40) enforce ideal DUO route choice.

The first three constraints (7.30)-(7.32) are state equations for each link $a$ and for cumulative effects at origins and destinations. equations (7.33)-(7.35) are flow conservation constraints at each node including origins and destinations. Equations (7.36) and (7.40) are constraints which equilibrate *actual* route travel times for departure route flows. Other constraints include flow propagation constraints, definitional constraints, nonnegativity, and boundary conditions. In summary, the control variables are $u_{ap}^{rs}(t)$, $v_{ap}^{rs}(t)$, $e_p^{rs}(t)$, and $f_p^{rs}(t)$; the state variables are $x_{ap}^{rs}(t)$, $E_p^{rs}(t)$, and $F_p^{rs}(t)$; the functionals are $\pi^{rs}(t)$, which must be determined by diagonalization as discussed in Section 7.4. Note that constraints (7.36) and (7.40) apply to the above optimization program. Those two constraints guarantee that the optimal control problem will generate traffic flows satisfying the ideal DUO route choice conditions, given any O-D departure flows.

## 7.4 Solution Algorithm

As with the instantaneous DUO route choice model, we propose an algorithm to solve the ideal DUO route choice model. This ideal DUO route choice model has nonlinear constraints (7.36) which equilibrate *actual* route travel times for departing route flows. By placing these constraints as penalty terms in the objective function, we obtain a similar formulation to the instantaneous DUO route choice model. Since the revised objective function involves link flow and O-D flow variables, we can avoid enumerating routes in computing the objective function. Then using the diagonalization and Frank-Wolfe techniques, the resulting NLP program can be solved.

### 7.4.1 Discrete Formulation of the Ideal Model

To convert our ideal DUO route choice problem into an NLP, time period $[0, T]$ is subdivided into $K$ small time intervals. (The time intervals are not necessarily equal.) To simplify the formulation, we modify the estimated actual travel time on each link in the following way so that each estimated travel time is equal to a multiple of the time increment.

$$\tau_a(k) = i \quad \text{if} \quad i - 0.5 \leq \tau_a(k) < i + 0.5,$$

where $i$ is an integer and $0 \leq i \leq K$. We note that the above approximation applies to flow propagation constraints and the computation of route travel times.

In the resulting discrete time problem, $x_a(k)$ represents vehicles on the link at the beginning of interval $k$; $u_a(k)$ and $v_a(k)$ represent inflow and exit flows during interval $k$. Let $\tau_a(k)$ denote the travel time for vehicles entering link $a$ at the beginning of interval $k = [k, k+1]$, and let $\pi^{rs}(k)$ be the average minimal $r - s$ travel time for vehicles departing origin $r$ during interval $k$. Let $f^{rs}(k)$ denote the O-D departure flow during interval $k$.

The optimal control program can then be reformulated as a discrete time NLP as follows:

$$\min_{u,v,x,e,E,f,F,\pi} \quad Z = \sum_{k=1}^{K} \sum_{a} \left\{ \int_0^{u_a(k)} g_{1a}[x_a(k), \omega] d\omega + \int_0^{v_a(k)} g_{2a}[x_a(k), \omega] d\omega \right\} \tag{7.45}$$

s.t.

$$x_{ap}^{rs}(k+1) = x_{ap}^{rs}(k) + u_{ap}^{rs}(k) - v_{ap}^{rs}(k) \qquad \forall a, p, r, s; k = 1, \cdots, K; \tag{7.46}$$

$$F_p^{rs}(k+1) = F_p^{rs}(k) + f_p^{rs}(k) \qquad \forall p, r, s; k = 1, \cdots, K; \tag{7.47}$$

$$E_p^{rs}(k+1) = E_p^{rs}(k) + e_p^{rs}(k) \qquad \forall p, r, s; k = 1, \cdots, K; \tag{7.48}$$

$$f_p^{rs}(k) = \sum_{a \in A(r)} \delta_{ap}^{rs} u_{ap}^{rs}(k) \qquad \forall p, r, s; k = 1, \cdots, K; \tag{7.49}$$

$$e_p^{rs}(k) = \sum_{a \in B(s)} \delta_{ap}^{rs} v_{ap}^{rs}(k) \qquad \forall p, r, s; k = 1, \cdots, K; \tag{7.50}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(k) - \sum_{a \in A(j)} u_{ap}^{rs}(k) = 0 \qquad \forall j, p, r, s; j \neq r, s; k = 1, \cdots, K; \tag{7.51}$$

$$x_{ap}^{rs}(k) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[k + \tau_a(k)] - x_{bp}^{rs}(k)\} + \{E_p^{rs}[k + \tau_a(k)] - E_p^{rs}(k)\}$$
$$\forall a \in B(j); j \neq r; p, r, s; k = 1, \cdots, K+1; \tag{7.52}$$

$$F^{rs}(k) = E^{rs}[k + \pi^{rs}(k)] \qquad \forall r, s; k = 1, \cdots, K; \tag{7.53}$$

$$u_{ap}^{rs}(k) \geq 0, \;\; v_{ap}^{rs}(k) \geq 0, \;\; x_{ap}^{rs}(k+1) \geq 0, \;\; \forall a, p, r, s; k = 1, \cdots, K; \tag{7.54}$$

$$E_p^{rs}(k+1) \geq 0, \qquad F_p^{rs}(k+1) \geq 0, \qquad \forall p, r, s; k = 1, \cdots, K; \tag{7.55}$$

$$E_p^{rs}(1) = 0 \qquad F_p^{rs}(1) = 0 \qquad \forall p, r, s; \tag{7.56}$$

$$x_{ap}^{rs}(1) = 0, \qquad \forall a, p, r, s. \tag{7.57}$$

In addition, we require the following definitional constraints.

$$\sum_{rsp} u_{ap}^{rs}(k) = u_a(k), \;\; \sum_{rsp} v_{ap}^{rs}(k) = v_a(k), \;\; \sum_{rsp} x_{ap}^{rs}(k) = x_a(k), \;\; \forall a; \tag{7.58}$$

$$\sum_p e_p^{rs}(k) = e^{rs}(k), \quad \sum_p E_p^{rs}(k) = E^{rs}(k), \quad \forall r, s; \tag{7.59}$$

$$\sum_p f_p^{rs}(k) = f^{rs}(k), \quad \sum_p F_p^{rs}(k) = F^{rs}(k), \quad \forall r, s. \tag{7.60}$$

Nonlinear constraints (7.53) may not hold strictly as equalities because of cumulative round-off errors of link flow variables over routes after time discretization.

### 7.4.2 The Penalty Method

The penalty method has had very few applications to route choice models. Inouye (1986) used the barrier method to solve the static UO problem with explicit link capacity constraints. He combined capacity constraints into the objective function and then used the Frank-Wolfe algorithm to solve the modified problem. In this section, we apply the penalty method to the constraints associated with the relationship between travel times and link flows. We place equality constraints (7.53) in the objective function as penalty terms. Then, only flow conservation and flow propagation equations remain, and the diagonalization/Frank-Wolfe technique can be used to solve the modified program. The penalty function $d^{rs}(k)$ replacing constraint (7.53) is defined as

$$d^{rs}(k) = \{F^{rs}(k) - E^{rs}[k + \pi^{rs}(k)]\}^2 \qquad \forall r, s; k = 2, \cdots, K+1.$$

We then reformulate the discrete time NLP as follows:

$$\min_{u,v,x,e,E,f,F,\pi} \hat{Z} = \sum_{k=1}^{K} \sum_a \{ \int_0^{u_a(k)} g_{1a}[x_a(k), \omega] d\omega + \int_0^{v_a(k)} g_{2a}[x_a(k), \omega] d\omega \} + \sum_{k=2}^{K+1} \sum_{r,s} \mu^{(n)} d^{rs}(k) \tag{7.61}$$

s.t.

$$x_{ap}^{rs}(k+1) = x_{ap}^{rs}(k) + u_{ap}^{rs}(k) - v_{ap}^{rs}(k) \qquad \forall a, p, r, s; k = 1, \cdots, K; \tag{7.62}$$

$$F_p^{rs}(k+1) = F_p^{rs}(k) + f_p^{rs}(k) \qquad \forall r; s \neq r; p; k = 1, \cdots, K; \tag{7.63}$$

$$E_p^{rs}(k+1) = E_p^{rs}(k) + e_p^{rs}(k) \qquad \forall r; s \neq r; p; k = 1, \cdots, K; \tag{7.64}$$

$$\sum_p f_p^{rs}(k) = f^{rs}(k), \qquad \forall r, s; \tag{7.65}$$

$$f_p^{rs}(k) - \sum_{a \in A(r)} u_{ap}^{rs}(k) = 0 \qquad \forall p, r, s; k = 1, \cdots, K; \tag{7.66}$$

$$f_p^{rs}(k) = \sum_{a \in A(r)} \delta_{ap}^{rs} u_{ap}^{rs}(k) \qquad \forall p, r, s; k = 1, \cdots, K; \tag{7.67}$$

$$e_p^{rs}(k) = \sum_{a \in B(s)} \delta_{ap}^{rs} v_{ap}^{rs}(k) \qquad \forall p, r, s; k = 1, \cdots, K; \tag{7.68}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(k) - \sum_{a \in A(j)} u_{ap}^{rs}(k) = 0 \qquad \forall j \neq r, s; k = 1, \cdots, K; \tag{7.69}$$

$$x_{ap}^{rs}(k) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[k + \tau_a(k)] - x_{bp}^{rs}(k)\} + \{E_p^{rs}[k + \tau_a(k)] - E_p^{rs}(k)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; k = 1, \cdots, K+1; \tag{7.70}$$

$$u_{ap}^{rs}(k) \geq 0, \quad v_{ap}^{rs}(k) \geq 0, \quad x_{ap}^{rs}(k+1) \geq 0, \quad \forall a, p, r, s; k = 1, \cdots, K; \tag{7.71}$$

$$e_p^{rs}(k) \geq 0, \qquad f_p^{rs}(k) \geq 0, \qquad \forall p, r, s; k = 1, \cdots, K; \tag{7.72}$$

$$E_p^{rs}(k+1) \geq 0, \qquad F_p^{rs}(k+1) \geq 0, \qquad \forall p, r, s; k = 1, \cdots, K; \tag{7.73}$$

$$E_p^{rs}(1) = 0 \qquad F_p^{rs}(1) = 0 \qquad \forall p, r, s; \tag{7.74}$$

$$x_{ap}^{rs}(1) = 0, \qquad \forall a, p, r, s. \tag{7.75}$$

In the objective function, the last term is the penalty term associated with equality constraint equation (7.53). The penalty coefficient $\mu^{(n)}$ is a large positive number and increases with iteration $n$. It is well known that this penalty function $d^{rs}$ has continuous first partial derivatives.

$$\frac{\partial d^{rs}(k)}{\partial E^{rs}[k + \pi^{rs}(k)]} = 2\,\{F^{rs}(k) - E^{rs}[k + \pi^{rs}(k)]\} \geq 0 \;\; \forall r, s; k = 2, \cdots, K+1.$$

### 7.4.3 The Diagonalization/Frank-Wolfe Algorithm

We then apply the *diagonalization technique* as introduced in Chapter 6. In this procedure, the actual travel times over each link $a$, $\tau_a(k)$, are temporarily fixed as $\bar{\tau}_a(k)$ and are updated iteratively. At each iteration, since each $\bar{\tau}_a(k)$ is temporarily fixed, the minimal O-D travel time functional $\pi^{rs}(k)$ can be computed and is also temporarily fixed as $\bar{\pi}^{rs}(k)$. After solving the route choice problem for fixed $\bar{\tau}_a(k)$, link travel times corresponding to the solution obtained for $x_a(k)$, $u_a(k)$ and $v_a(k)$ are compared to functions $\tau_a(k)$. If link travel times corresponding to the solution are different from $\tau_a(k)$, their values are reset to these travel times and the process is repeated. After link travel time $\tau_a(k)$ is updated, minimal O-D travel time $\pi^{rs}(k)$ can be updated so that the penalty terms can be updated. Then we can proceed to the next diagonalization iteration. Given the robust nature of the diagonalization technique, we expect that the solution will converge to the ideal DUO solution.

We use the same diagonalization/Frank-Wolfe algorithm presented in Chapter 6 for solving the instantaneous DUO traffic assignment problem. Denote the subproblem variables as $p, q, y, \bar{E}, \bar{e}, \bar{f}, \bar{F}$, corresponding to the main problem variables $u, v, x, E, e, f, F$. To distinguish the notation, we use $l$ to denote a route in the subproblem. Applying the Frank-Wolfe algorithm to the minimization of the discretized ideal DUO program requires, at each iteration, a solution of the following linear program (LP):

$$\min_{p,q,y,\bar{E},\bar{e},\bar{f},\bar{F}} \hat{Z} = \nabla_u \hat{Z} \cdot p^T + \nabla_v \hat{Z} \cdot q^T + \nabla_x \hat{Z} \cdot y^T + \nabla_E \hat{Z} \cdot \bar{E}^T \tag{7.76}$$

s.t.

$$y_{al}^{rs}(k+1) = y_{al}^{rs}(k) + p_{al}^{rs}(k) - q_{ap}^{rs}(k) \quad \forall a, l, r, s; k = 1, \cdots, K; \tag{7.77}$$

$$\bar{F}_l^{rs}(k+1) = \bar{F}_l^{rs}(k) + \bar{f}_l^{rs}(k) \quad \forall l, r, s; k = 1, \cdots, K; \tag{7.78}$$

$$\bar{E}_l^{rs}(k+1) = \bar{E}_l^{rs}(k) + \bar{e}_l^{rs}(k) \quad \forall l, r, s; k = 1, \cdots, K; \tag{7.79}$$

$$\sum_l f_l^{rs}(k) = f^{rs}(k) = 0 \quad \forall r, s; k = 1, \cdots, K; \tag{7.80}$$

$$f_l^{rs}(k) - \sum_{a \in A(r)} p_{al}^{rs}(k) = 0 \quad \forall l, r, s; k = 1, \cdots, K; \tag{7.81}$$

$$\bar{f}_l^{rs}(k) = \sum_{a \in A(r)} \delta_{al}^{rs} u_{al}^{rs}(k) \quad \forall l, r, s; k = 1, \cdots, K; \tag{7.82}$$

$$\bar{e}_l^{rs}(k) = \sum_{a \in B(s)} \delta_{al}^{rs} v_{al}^{rs}(k) \quad \forall l, r, s; k = 1, \cdots, K; \tag{7.83}$$

$$\sum_{a \in B(j)} q_{al}^{rs}(k) - \sum_{a \in A(j)} p_{al}^{rs}(k) = 0 \quad \forall j, l, r, s; j \neq r, s; k = 1, \cdots, K; \tag{7.84}$$

$$y_{al}^{rs}(k) = \sum_{b\in\tilde{I}} \{y_{bl}^{rs}[k+\bar{\tau}_a(k)] - y_{bl}^{rs}(k)\} + \{\bar{E}_l^{rs}[k+\bar{\tau}_a(k)] - \bar{E}_l^{rs}(k)\}$$

$$\forall a \in B(j); j \neq r; l, r, s; k = 1, \cdots, K+1; \quad (7.85)$$

$$y_{al}^{rs}(k+1) \geq 0, \quad p_{al}^{rs}(k) \geq 0, \quad q_{al}^{rs}(k) \geq 0, \quad \forall a, l, r, s; k = 1, \cdots, K; \quad (7.86)$$

$$\bar{e}_l^{rs}(k) \geq 0, \qquad \bar{f}_l^{rs}(k) \geq 0, \qquad \forall l, r, s; k = 1, \cdots, K; \quad (7.87)$$

$$\bar{E}_l^{rs}(k+1) \geq 0, \qquad \bar{F}_l^{rs}(k+1) \geq 0, \qquad \forall l, r, s; k = 1, \cdots, K; \quad (7.88)$$

$$\bar{E}_l^{rs}(1) = 0, \qquad \forall l, r, s; \quad (7.89)$$

$$y_{al}^{rs}(1) = 0, \qquad \forall a, l, r, s. \quad (7.90)$$

The objective function (7.76) is equivalent to:

$$\begin{aligned} \hat{\hat{Z}} &= \sum_{k=1}^{K}\sum_{r,s}\sum_{a,l}\left[\frac{\partial \hat{Z}}{\partial u_{al}^{rs}(k)} p_{al}^{rs}(k) + \frac{\partial \hat{Z}}{\partial v_{al}^{rs}(k)} q_{al}^{rs}(k)\right. \\ &+ \left.\frac{\partial \hat{Z}}{\partial x_{al}^{rs}(k+1)} y_{al}^{rs}(k+1)\right] + \sum_{k=2}^{K+1}\sum_{r,s,l} \frac{\partial \hat{Z}}{\partial E_l^{rs}(k)} \bar{E}_l^{rs}(k) \end{aligned} \quad (7.91)$$

The components of the gradient of $\hat{Z}$ with respect to the control and state variables $u$, $v$, $x$, $E$ are

$$t_{1a}(k) = \frac{\partial \hat{Z}}{\partial u_a(k)} = g_{1a}[x_a(k), u_a(k)] \qquad \forall a; k = 1, \cdots, K; \quad (7.92)$$

$$t_{2a}(k) = \frac{\partial \hat{Z}}{\partial v_a(k)} = g_{2a}[x_a(k), v_a(k)] \qquad \forall a; k = 1, \cdots, K; \quad (7.93)$$

$$\begin{aligned} t_{3a}(k) &= \frac{\partial \hat{Z}}{\partial x_a(k)} \\ &= \int_0^{u_a(k)} \frac{\partial g_{1a}[x_a(k), \omega]}{\partial x_a(k)} d\omega + \int_0^{v_a(k)} \frac{\partial g_{2a}[x_a(k), \omega]}{\partial x_a(k)} d\omega \\ &\qquad \forall a; k = 2, \cdots, K; \end{aligned} \quad (7.94)$$

$$t_{3a}(K+1) = \frac{\partial \hat{Z}}{\partial x_a(K+1)} = 0 \qquad \forall a; \quad (7.95)$$

$$t_4^{rs}(k) = \frac{\partial \hat{Z}}{\partial E^{rs}(k)} = 0 \qquad \forall r, s; k = 2, \cdots, \bar{\pi}^{rs}(1) + 1; \quad (7.96)$$

$$\begin{aligned} t_4^{rs}[k+\bar{\pi}^{rs}(k)] &= \frac{\partial \hat{Z}}{\partial E^{rs}[k+\bar{\pi}^{rs}(k)]} \\ &= 2\mu^{(n)} \cdot \{F^{rs}(k) - E^{rs}[k+\bar{\pi}^{rs}(k)]\} \\ &\qquad \forall r, s; k = 2, \cdots, K+1; \end{aligned} \quad (7.97)$$

The objective function can be rewritten as

$$\begin{aligned}\hat{Z} &= \sum_{k=1}^{K}\sum_{r,s}\sum_{a,l}[t_{1a}(k)p_{al}^{rs}(k) + t_{2a}(k)q_{al}^{rs}(k) + t_{3a}^{rs}(k+1)y_{al}^{rs}(k+1)] \\ &+ \sum_{k=2}^{K+1}\sum_{r,s,l} t_4^{rs}(k)\bar{E}_l^{rs}(k) \qquad (7.98)\end{aligned}$$

Since $g_{1a}$ and $g_{2a}$ are nonnegative and increasing functions, it follows that

$$t_{1a}(k), t_{2a}(k) \geq 0, \qquad \forall a; k = 1,\cdots,K; \qquad (7.99)$$

$$t_{3a}^{rs}(k+1) \geq 0, \qquad \forall a, r, s; k = 1,\cdots,K. \qquad (7.100)$$

$$t_4^{rs}(k+1) \geq 0, \qquad \forall r, s; k = 1,\cdots,K. \qquad (7.101)$$

This has a significant impact on the solvability of this model for large networks, since these components will be link cost coefficients in a minimal-cost network problem. Note that there are no capacity constraints on the links; the only constraints are non-negativity and conservation of flow. Furthermore, the constraints apply to each origin-destination pair independently, so linear program (7.76)-(7.90) can be decomposed by origin-destination pair. The resulting subproblem for each O-D pair (r,s) is given by

$$\begin{aligned}\min_{p,q,y,\bar{E}} \quad & \sum_{k=1}^{K}\sum_{al}[t_{1a}p_{al}^{rs}(k) + t_{2a}q_{al}^{rs}(k) + t_{3a}^{rs}(k+1)y_{al}^{rs}(k+1)] \\ + \quad & \sum_{k=2}^{K+1}\sum_{l} t_4^{rs}(k)\bar{E}_l^{rs}(k) \qquad (7.102)\end{aligned}$$

s.t. constraints (7.77)-(7.90).

In addition to the fourth term, the above program is identical to the LP subproblem of the discrete time instantaneous DUO route choice model. The fourth term involves flow variables associated with the origin and destination nodes only. Thus, the above LP subproblem for each O-D pair $(r, s)$ can be viewed as a many-to-one minimal-cost route problem over an expanded time-space network using artificial origins. It can be solved by determining the minimal-cost routes from all artificial origins to a super destination and completing an all-or-nothing assignment. Flow variables $p_{al}^{rs}(k)$, $q_{al}^{rs}(k)$, $y_{al}^{rs}(k+1)$, $\bar{E}_l^{rs}(k+1)$, are determined by solving the minimal-cost route problem for each O-D pair $(r, s)$ and assigning the O-D flows to the links.

In this combined algorithm, we define the diagonalization procedure as the outer iteration and the F-W procedure as the inner iteration. Denote the new solution at inner F-W iteration $(n+1)$ as

$$u_a^{(n+1)}(k) = u_a^{(n)}(k) + \alpha^{(n)}[u_a^{(n)}(k) - p_a^{(n)}(k)], \qquad \forall a; k = 1,\cdots,K; \qquad (7.103)$$

$$v_a^{(n+1)}(k) = v_a^{(n)}(k) + \alpha^{(n)}[v_a^{(n)}(k) - q_a^{(n)}(k)], \quad \forall a; k = 1, \cdots, K; \quad (7.104)$$

$$x_a^{(n+1)}(k) = x_a^{(n)}(k) + \alpha^{(n)}[x_a^{(n)}(k) - y_a^{(n)}(k)], \quad \forall a; k = 1, \cdots, K+1; \quad (7.105)$$

$$E^{(n+1)}(k) = E^{(n)}(k) + \alpha^{(n)}[E^{(n)}(k) - \bar{E}^{(n)}(k)], \quad \forall r, s; k = 1, \cdots, K+1; \quad (7.106)$$

where $\alpha^{(n)}$ is the optimal step size of the one-dimensional search problem in the F-W algorithm. The one-dimensional search problem is to find step size $\alpha^{(n)}$ that solves

$$\min_{0 \leq \alpha^{(n)} \leq 1} \quad \sum_{k=1}^{K} \sum_{a} \left\{ \int_0^{u_a^{(n+1)}(k)} g_{1a}[x_a^{(n+1)}(k), \omega] d\omega + \int_0^{v_a^{(n+1)}(k)} g_{2a}[x_a^{(n+1)}(k), \omega] d\omega \right\} + \sum_{k=2}^{K+1} \sum_{r,s} \mu^{(n)} d^{rs} [E^{(n+1)}(k)] \quad (7.107)$$

where $u_a^{(n+1)}(k)$, $v_a^{(n+1)}(k)$, $x_a^{(n+1)}(k)$, $E^{(n+1)}(k)$ are replaced by the definitional equations in the above.

## 7.4.4 Summary of the Algorithm

The algorithm for solving our ideal DUO route choice model is illustrated in the flowchart in Figure 7.3 and is summarized as follows.

**Step 0: Initialization.**

Find an initial feasible solution $\{x_a^{(1)}(k)\}$, $\{u_a^{(1)}(k)\}$, $\{v_a^{(1)}(k)\}$, $\{E^{(1)}(k)\}$. Set the outer iteration counter $m = 1$.

**Step 1: Diagonalization.**

Find a new estimate of actual link travel time $\tau_a^{(n)}(k)$ and solve the DUO program. Set the inner iteration counter $n = 1$.

[Step 1.1]: *Update.* Calculate $t_{1a}(k)$, $t_{2a}(k)$, $t_{3a}(k)$ and $t_4^{rs}(k)$ using equations (7.92)-(7.97).

[Step 1.2]: *Direction Finding.* Based on $\{t_{1a}(k)\}$, $\{t_{2a}(k)\}$, $\{t_{3a}(k)\}$ and $\{t_4^{rs}(k)\}$, search the minimal-cost route forward from each artificial origin to the super destination over an expanded time-space network for each O-D pair $(r, s)$. Perform an all-or-nothing assignment, yielding subproblem solution $\{p_a(k)\}$, $\{q_a(k)\}$, $\{y_a(k)\}$, $\{\bar{E}^{rs}(k)\}$.

[Step 1.3]: *Line Search.* Find the optimal step size $\alpha^{(n)}$ that solves the one dimensional search problem.

[Step 1.4]: *Move.* Find a new solution by combining $\{u_a(k)\}$, $\{v_a(k)\}$, $\{x_a(k)\}$, $\{E^{rs}(k)\}$ and $\{p_a(k)\}$, $\{q_a(k)\}$, $\{y_a(k)\}$, $\{\bar{E}^{rs}(k)\}$.

[Step 1.5]: *Convergence Test for Inner Iterations.* If $n$ equals the prespecified number, go to step 2; otherwise, set $n = n+1$ and go to step 1.1.

**Step 2: Convergence Test for Outer Iterations.**
If $\tau_a^{(m)}(k) \simeq \tau_a^{(m+1)}(k)$, and the penalty term $\sum_k \sum_{r,s} \mu^{(n)} d^{rs}(k) \simeq 0$, stop. The current solution, $\{u_a(k)\}, \{v_a(k)\}, \{x_a(k)\}, \{E^{rs}(k)\}$, is in a near ideal DUO state; otherwise, set $m = m+1$ and go to step 1.

In order to speed up convergence, an incremental assignment technique is suggested for finding a good starting solution before applying the diagonalization procedure.

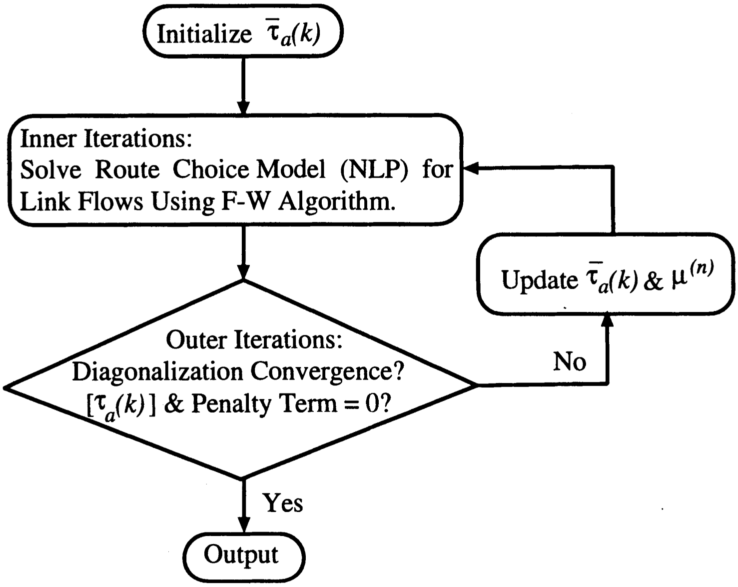



Figure 7.3: Flowchart of the Solution Algorithm

## 7.5 Notes

In our model, a continuous flow of traffic is implicitly assumed – when vehicles approach a decision node (intersection), drivers can continue straight ahead or turn left or right without physically blocking the road. In reality, this may not be literally true. For example, suppose two vehicles traveling on a one-lane road arrive at an intersection. The DUO route choice conditions in our model may require the first vehicle to wait and then make a left turn, but allow the second vehicle to go straight ahead without waiting. In this situation, the dynamic physical conditions will be violated since in reality the second vehicle would have to wait for the first one. This implies that our model literally represents only networks with turning lanes of sufficient length to accommodate all turning vehicles.

An advantage of the proposed model is that the flow propagation constraints can serve as a black box. If accurate link travel time functions are not available, the model still works because the flow propagation constraints can be easily replaced by appropriate traffic flow models. The resulting formulation is a set of nonlinear dynamic equations, or a nonlinear complementarity problem.

In the following, we provide some insights into the instantaneous and ideal DUO route choice models. Our dynamic route choice models are optimization oriented. Ultimately, some optimal control measures are necessary to help to achieve our goal. Those controls may include optimal routing, signal optimization, and other means. Consequently, these control problems can be modeled using optimal control theory. In traffic prediction problems, including dynamic route choice models, many uncertain events can occur in the future, such as traffic accidents, illegal double parking on streets, etc. In other words, disturbances for future traffic might be large enough to cause the traffic prediction and traffic control models based on predicted future travel times (or actual travel times as defined in this book) to fail. In those cases, optimal control models based on instantaneous travel times might provide more meaningful results since feedback is taken into account in the optimal control models and drivers can adjust their routes en-route using current traffic information. In terms of optimal control theory and the uncertainty characteristic of our problem, instantaneous travel time prediction might be the only choice in some situations. This is one of the reasons why feedback control theory was developed.

Of course, the ideal dynamic user-optimal route choice model, which equalibrates actual route travel times in the future, is also useful when some future disturbance (incidents and other future events) are predictable, such as the increased traffic flow from a baseball stadium after a game. In conclusion, the best solution is to provide both instantaneous and ideal user-optimal route choice models for use in various appropriate situations. When a future disturbance is more predicable, we can use the ideal dynamic user-optimal route choice model. When a future disturbance is less predicable, we may prefer the instantaneous model.

# Chapter 8

# Stochastic Dynamic User-Optimal Route Choice Models

Dynamic route guidance systems are being developed in order to inform and guide drivers regarding their best departure times and routes so as to avoid congestion delay. However, drivers may or may not rely on the information provided by the route guidance system to adjust their departure times and routes. Furthermore, drivers without navigation systems do not have perfect information on the network traffic and must use their own experience and perception of current traffic conditions to make travel decisions. Thus, there is a need to develop dynamic route choice models under imperfect information as well as perfect information.

To date, many dynamic route choice models have been presented, regarding both route and departure time choices (see Chapter 1 for an overview). However, very few existing models can be applied to the situation with imperfect traffic information or indeterministic traffic conditions. Moreover, deterministic dynamic user-optimal (DUO) route choice models assign flows to minimal travel time routes only. Those deterministic models can not represent the realistic traffic flow dispersion across different travel time routes. In response to this gap, this chapter presents two stochastic dynamic user-optimal (SDUO) route choice models which are stochastic extensions of our deterministic DUO route choice models presented in Chapters 5 and 7. We note that there is another approach to represent traffic flow dispersion across different travel time routes. This approach stratifies vehicles into different groups and uses group-specific travel disutilities to determine dynamic user-optimal routes. We will discuss this approach in Chapters 12-13.

The perceptions of travel times have also been studied using stochastic static route choice models. Stochastic route choice models have been explored extensively under the assumption of static traffic conditions (Daganzo and Sheffi, 1977; Fisk, 1980; Sheffi and Powell, 1983, etc.). Recently, Cassetta (1991) studied the variation of dynamic route choice from day to day.

Vythoulkas (1990) also extended stochastic static route choice models into the dynamic route choice framework.

The SDUO route choice problem is to find dynamic trajectories of link states and inflow and exit flow control variables, given time-dependent O-D departure rates, the network, link travel time functions and some assumptions about the randomness of imperfect traffic information. Basically, SDUO problems can be classified according to the following:

1. random components in traveler's perceptions of travel times;
2. randomness of origin-destination flows;
3. randomness of the link traffic states.

In this chapter, we will include only the random component in travelers' perception of travel times. Then route choice is a process of selection among alternative routes, for which the perceived time-dependent travel times include a random error. We note that the proposed models in this chapter represent only our initial effort to tackle this very difficult problem. Logit-type dynamic route choice models are used in our formulations in this chapter because of their advantages of mathematical tractability, although logit-type models have IIA (Independence of Irrelevant Alternatives) properties which cause inaccuracy in route choice dispersion. In subsequent studies, we will examine more realistic, but more complicated distributions of route choice dispersion.

In this chapter, the stochastic route choice assumption is introduced into dynamic route choice problems so that our SDUO models are dynamic generalizations of conventional stochastic static user-optimal (SUO) models under the assumption of dispersed travel choice. Stochastic dynamic route choice models are better representations than deterministic dynamic route choice models because travelers' route choice dispersion is taken into account. We concentrate our analysis on the modeling aspects of SDUO route choice formulation in this chapter. Solution algorithms for solving the two models are presented in the next chapter.

Compared with deterministic dynamic route choice models, the proposed stochastic dynamic route choice models represent at least a better approximation to real world conditions because in SDUO route choice models, O-D departure flows are dispersed across different travel time routes. On the contrary, deterministic DUO route choice models assign O-D departure flows to minimal travel time routes only. Furthermore, SDUO models provide better representations of travelers' route choice behavior than their static counterparts because time-dependent traffic flows and travel times are explicitly taken into account.

In this chapter, we first present definitions of travel times in stochastic dynamic problems in Section 8.1. The constraints and the instantaneous SDUO model are then described in Section 8.2. The formulation of the instantaneous SDUO route choice problem is based on the underlying choice criterion that each traveler uses the route that minimizes his/her perceived instantaneous travel time when departing from the origin or any intermediate node to his/her destination. Subsequently, the equivalence of the model with instantaneous SDUO route choice is demonstrated by proving the equivalence of the optimality conditions of the model with the instantaneous SDUO route choice conditions. The solution of this instantaneous SDUO route choice model will result in instantaneous stochastic network flows at each decision node based on a logit function of mean instantaneous travel times of alternative routes. Here, we use logit-type of distribution because of closed form properties for choice probability. We note that this is a first step toward a more realistic distribution. It is also shown that our instantaneous DUO route choice model in Chapter 5 is a particular case of the instantaneous SDUO route choice model when the variance of instantaneous route travel time perception is zero.

In Section 8.3, we present an ideal SDUO route choice model based on stochastic flow loading with a logit function of *mean* actual travel times experienced by drivers over alternative routes for each O-D pair. It is shown that our previous deterministic ideal DUO route choice model in Chapter 7 is a special case of the ideal SDUO route choice model when the variance of the perceived actual route travel time is zero. In Section 8.4, some properties of SDUO route choice models are discussed.

# 8.1 Definitions of Travel Times in Stochastic Situations

Recall that in an ATIS system, there are two kinds of travel time information which can be provided to travelers: current information and future predictions. Current travel time information can be obtained using the currently prevailing instantaneous link travel times. Correspondingly, future travel time information can be obtained using predicted actual link travel times.

The *instantaneous* travel time at time $t$ is defined as the travel time that is experienced by vehicles traversing link $a$ when prevailing traffic conditions remain unchanged. Let $C_a(t)$ denote the instantaneous travel time on link $a$ at time $t$ as perceived by a traveler randomly chosen from the population of travelers. $C_a(t)$ is a random variable that is assumed to have mean $c_a(t)$. The *mean* or *measured* (as opposed to *perceived*) *instantaneous* travel time $c_a(t)$ over link $a$ at time $t$ is assumed to be dependent on the number of vehicles $x_a(t)$, the inflow $u_a(t)$ and the exit flow $v_a(t)$ on link $a$ at time $t$. In this model, we assume the mean instantaneous travel time $c_a(t)$ on link $a$ is the sum of two components: 1) an instantaneous flow-dependent cruise time $g_{1a}[x_a(t), u_a(t)]$ over link $a$; and 2) an instantaneous queuing delay $g_{2a}[x_a(t), v_a(t)]$. It follows that

$$c_a(t) = g_{a1}[x_a(t), u_a(t)] + g_{2a}[x_a(t), v_a(t)]. \tag{8.1}$$

The two components $g_{1a}[x_a(t), u_a(t)]$ and $g_{2a}[x_a(t), v_a(t)]$ are assumed to be nonnegative and differentiable with respect to $x_a(t), u_a(t)$ and $x_a(t), v_a(t)$, respectively.

Consider the flow which originates at node $r$ at time $t$ and is destined for node $s$. There is a set of routes $\{p\}$ between O-D pair $rs$. Denote the minimal free flow travel time from node $i$ to destination node $s$ as $\sigma_o^{is}$. In this chapter, an *efficient route* between $rs$ is redefined to include only links $a = (i, j)$ such that node $j$ is closer to destination $s$ than node $i$, i.e., $\sigma_o^{is} > \sigma_o^{js}$ (Dial, 1971). In the following, all route related constraints are defined using efficient routes. By using efficient routes, the cyclic flow problem in some dynamic assignment models can be prevented. Define the mean or measured instantaneous travel time $\psi_p^{rs}(t)$ for each route $p$ between $rs$ as

$$\psi_p^{rs}(t) = \sum_{a \in rsp} c_a[x_a(t), u_a(t), v_a(t)] \qquad \forall r, s, p; \tag{8.2}$$

the summation is over all links $a$ in route $p$ from origin $r$ to destination $s$. Thus, the mean instantaneous route travel time is that experienced by vehicles if prevailing traffic conditions do not vary until vehicles reach their destination. This instantaneous route travel time provides a first approximation to the time-dependent vehicle travel time.

Next, define $\tau_a(t)$ as the mean actual travel time over link $a$ for vehicles entering link $a$ at time $t$. Let $T_a(t)$ denote the perceived actual travel time for flows entering link $a$ at time $t$. $T_a(t)$ is a random variable with mean $\tau_a(t)$. As described in Chapter 16, an actual link travel time function has a similar form to an instantaneous link travel time function. We refer readers to Section 16.1 for a detailed discussion of the difference between instantaneous and actual link travel times. Similarly, define $\eta_p^{rs}(t)$ as the mean actual travel time experienced over route $p$ by vehicles departing origin $r$ toward destination $s$ at time $t$. Once the mean actual link travel time $\tau_a(t)$ is determined, the mean actual route travel time $\eta_p^{rs}(t)$ can be computed using the following recursive formula:

$$\eta_p^{rj}(t) = \eta_p^{ri}(t) + \tau_a[t + \eta_p^{ri}(t)] \qquad \forall r, s, p; i, j \in p \tag{8.3}$$

where link $a = (i, j)$ is on route $p$.

The new notation for this chapter is summarized as follows:

| | | |
|---|---|---|
| $C_a(t)$ | = | perceived instantaneous travel time for link $a$ at time $t$ |
| $c_a(t)$ | = | mean instantaneous travel time for link $a$ at time $t$ |
| $\Psi_p^{rs}(t)$ | = | perceived instantaneous route travel time for route $p$ between $rs$ at time $t$ |

| | | |
|---|---|---|
| $\psi_p^{rs}(t)$ | = | mean instantaneous route travel time for route $p$ between $rs$ at time $t$ |
| $T_a(t)$ | = | perceived actual travel time over link $a$ for flows entering link $a$ at time $t$ |
| $\tau_a(t)$ | = | mean actual travel time over link $a$ for flows entering link $a$ at time $t$ |
| $\bar{\tau}_a(t)$ | = | estimated mean actual travel time over link $a$ for flows entering link $a$ at time $t$ |
| $\Omega_p^{rs}(t)$ | = | perceived actual travel time for flows departing origin $r$ toward destination $s$ over route $p$ at time $t$ |
| $\eta_p^{rs}(t)$ | = | mean actual travel time experienced by flows departing origin $r$ toward destination $s$ over route $p$ at time $t$ |
| $\sigma_o^{is}$ | = | minimal free flow travel time from node $i$ to destination node $s$ |
| $\theta$ | = | route choice dispersion parameter |

# 8.2 Instantaneous SDUO Route Choice Model

## 8.2.1 Model Formulation

The formulation of the instantaneous SDUO route choice problem is based on the underlying choice criterion that each traveler uses the route that minimizes his/her perceived instantaneous travel time when departing from the origin or any intermediate node to his/her destination. In the following, we extend a similar analysis for stochastic static route choice given by Sheffi (1985) to the case of instantaneous dynamic route choice.

For each origin-destination pair, there are many alternative routes $p$, each with some instantaneous travel time for each time instant. Due to variations in perceptions and exogenous factors, these instantaneous route travel times are perceived differently by each traveler. We represent the perceived instantaneous route travel time as a random variable for each time instant $t$. Given his or her perception of instantaneous route travel time at each time instant $t$, each traveler is assumed to choose the route with minimal perceived instantaneous travel time. Given a probability density function for the instantaneous route travel time at each time $t$, the instantaneous stochastic dynamic user-optimal route choice problem is to determine how many travelers will use each route at each decision node at each instant of time.

Let $\Psi_p^{rs}(t)$ denote the perceived instantaneous route travel time on route $p$ between origin $r$ and destination $s$ at each time $t$. Then, we assume that

$$\Psi_p^{rs}(t) = \psi_p^{rs}(t) + \xi_p^{rs}(t) \qquad \forall r, s, p, \tag{8.4}$$

where $\xi_p^{rs}(t)$ is a random error term associated with the route $p$ under consideration. Furthermore, assume that the mean of the error $E[\xi_p^{rs}(t)] = 0$, or $E[\Psi_p^{rs}(t)] = \psi_p^{rs}(t)$. In other words, the average perceived instantaneous route travel time is the mean instantaneous route travel time.

The share of travelers choosing route $p$ at each time $t$, $P_p^{rs}(t)$, is

$$P_p^{rs}(t) = Pr[\Psi_p^{rs}(t) \leq \Psi_q^{rs}(t), \ \forall \text{ routes } q \text{ between } r \text{ and } s] \quad \forall r, s, p. \tag{8.5}$$

In other words, the probability that a given route for each O-D pair is chosen at time $t$ is the probability that its instantaneous route travel time is perceived to be the lowest of all the alternative routes. Once the distribution of perceived instantaneous route travel times is specified, the probability of selecting each alternative route can be calculated and the route flow can be assigned as follows:

$$f_p^{rs}(t) = f^{rs}(t)P_p^{rs}(t) \qquad \forall r, s, p \tag{8.6}$$

where $f_p^{rs}(t)$ is a control variable specifying the departure rate on route $p$ from $r$ to $s$ at time $t$. In this chapter, we assume that

$$\Psi_p^{rs}(t) = \psi_p^{rs}(t) - \frac{1}{\theta}\epsilon_p^{rs}(t) \qquad \forall r, s, p, \tag{8.7}$$

where $\theta$ is a nonnegative parameter that scales the perceived instantaneous route travel time, and $\epsilon_p^{rs}(t)$ is a random error term associated with the route $p$ under consideration (Sheffi, 1985). We assume that the errors $\epsilon_p^{rs}(t)$ are identically and independently distributed (i.i.d.) Gumbel variates for each time instant $t$. Other distributions of the perceived instantaneous route travel times will be discussed in subsequent studies. Based on random utility theory, the logit route choice probability can be expressed as

$$P_p^{rs}(t) = \frac{\exp[-\theta\psi_p^{rs}(t)]}{\sum_l \exp[-\theta\psi_l^{rs}(t)]} \qquad \forall r, s, p. \tag{8.8}$$

As $\theta \rightarrow \infty$, $var[\Psi_p^{rs}(t)] \rightarrow 0$ and the perceived instantaneous travel time between the O-D equals the mean instantaneous travel time. In this case, travelers will choose the minimal-time route at each time instant, as in the deterministic instantaneous DUO route choice model in Chapter 5. We note that the above analysis for each O-D pair also applies for each decision node-destination pair.

Analogous to the deterministic instantaneous DUO route choice model, we can use a conventional definition of used links and routes. For any link $a$ and any O-D pair $rs$, link $a$ is defined as being used at time $t$ if $u_a^{rs}(t) > 0$. Furthermore, a route $p$ between $r$ and $s$ is defined as being used at time $t$ if $u_{ap}^{rs}(t) > 0$, where link $a$ is the first link on route $p$ from $r$ to $s$. The above definition will be used in general variational inequality models for instantaneous SDUO route choice problems. In this chapter, we are formulating optimal control models which are equivalent to the instantaneous SDUO route choice conditions. These optimal control models are related to the general variational inequality models. Thus, we use the following relaxed definition of used links and routes in this optimal control model. For a link $a$ on any route from origin $r$ to destination $s$, link $a$ is being used at time $t$ if $u_a^{rs}(t) > 0$ and $v_a^{rs}(t) > 0$.

Furthermore, a route $p$ between $r$ and $s$ is being used at time $t$ if $u_a^{rs}(t) > 0$ and $v_a^{rs}(t) > 0$ for all links $a \in rsp$. We note that this definition only applies to this instantaneous SDUO route choice model.

The instantaneous stochastic dynamic user-optimal (SDUO) route choice problem is to determine vehicle flows at each instant of time on each link resulting from drivers using perceived minimal-time routes under the currently prevailing travel times. The instantaneous stochastic dynamic user-optimal state is the following dynamic generalization of the conventional stochastic static user-optimal (SUO) state.

> **Instantaneous SDUO State:** *If, for each O-D pair at each decision node at each instant of time, the perceived instantaneous travel times for all routes that are being used equal the minimal perceived instantaneous route travel time, the dynamic traffic flow over the network is in an instantaneous stochastic dynamic user-optimal state.*

This definition is the stochastic generalization of the one given in Chapter 5. It assumes that all drivers make their route choice decisions using their perceptions of the current prevailing O-D travel times so that traffic flows are more dispersed over alternative routes than the deterministic dynamic route choice models.

As in deterministic instantaneous DUO route choice models in Chapter 5, we write the flow propagation constraints using *estimates* of the mean actual link travel times. These link time estimates must be updated in an iterative diagonalization solution procedure. Denote $\bar{\tau}_a(t)$ as the estimated mean actual travel time over link $a$ for flows entering link $a$ at time $t$. Then, the flow propagation constraints are as follows:

$$x_{ap}^{rs}(t) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[t + \bar{\tau}_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \bar{\tau}_a(t)] - E_p^{rs}(t)\}$$

$$\forall a \in p; p, r, s. \qquad (8.9)$$

Using optimal control theory, an equivalent optimization model of the instantaneous stochastic dynamic user-optimal route choice problem is formulated as follows.

$$\min_{u,v,x,e,E} \int_0^T \left\{ \sum_a \left[ \int_0^{u_a(t)} g_{1a}[x_a(t), \omega] d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t), \omega] d\omega \right] \right.$$
$$+ \frac{1}{\theta} \left[ \sum_{rs} \sum_p \sum_a u_{ap}^{rs}(t) \ln u_{ap}^{rs}(t) \right.$$
$$\left. \left. - \sum_{rs} \sum_p \sum_{a \notin B(s)} v_{ap}^{rs}(t) \ln v_{ap}^{rs}(t) \right] \right\} dt \qquad (8.10)$$

s.t.

Relationship between state and control variables:

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \tag{8.11}$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r; s \neq r; \tag{8.12}$$

Flow conservation constraints:

$$f^{rs}(t) = \sum_{a \in A(r)} \sum_p u_{ap}^{rs}(t) \qquad \forall r \neq s; s; \tag{8.13}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \tag{8.14}$$

$$e^{rs}(t) = \sum_{a \in B(s)} \sum_p v_{ap}^{rs}(t) \qquad \forall r, s \neq r; \tag{8.15}$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[t + \bar{\tau}_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \bar{\tau}_a(t)] - E_p^{rs}(t)\}$$

$$\forall a \in p; p, r, s; \tag{8.16}$$

Definitional constraints:

$$\sum_p u_{ap}^{rs}(t) = u_a^{rs}(t), \qquad \sum_{rs} u_a^{rs}(t) = u_a(t), \qquad \forall a; \tag{8.17}$$

$$\sum_p v_{ap}^{rs}(t) = v_a^{rs}(t), \qquad \sum_{rs} v_a^{rs}(t) = v_a(t), \qquad \forall a; \tag{8.18}$$

$$\sum_p x_{ap}^{rs}(t) = x_a^{rs}(t), \qquad \sum_{rs} x_a^{rs}(t) = x_a(t), \qquad \forall a; \tag{8.19}$$

$$\sum_p e_p^{rs}(t) = e^{rs}(t), \qquad \sum_p E_p^{rs}(t) = E^{rs}(t), \qquad \forall r, s; \tag{8.20}$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \quad u_{ap}^{rs}(t) \geq 0, \quad v_{ap}^{rs}(t) \geq 0 \quad \forall a, p, r, s; \tag{8.21}$$

$$e_p^{rs}(t) \geq 0, \quad E_p^{rs}(t) \geq 0, \qquad \forall p, r, s; \tag{8.22}$$

Boundary conditions:

$$E_p^{rs}(0) = 0, \quad \forall p, r, s; \qquad x_{ap}^{rs}(0) = 0, \quad \forall a, p, r, s; \tag{8.23}$$

where the expressions $u(t)\ln u(t)$ and $v(t)\ln v(t)$ are assigned the value zero at $u(t) = 0$ and $v(t) = 0$, respectively. The first two terms of the objective function are similar to the objective function of the well-known stochastic static UO model. The third and fourth terms of the objective function are also similar to the route flow entropy terms in the objective function of the stochastic UO model. In the last term of the objective function, the entropy function applies to all except the last link on every route between every O-D pair.

The first two constraints (8.11)-(8.12) are state equations for flows on each link $a$ and for the arrivals at destinations. Equations (8.13)-(8.15) are flow conservation constraints at each node including origins and destinations. The other constraints include flow propagation constraints, definitional constraints, nonnegativity, and boundary conditions. In summary, the control variables are $u_{ap}^{rs}(t)$, $v_{ap}^{rs}(t)$, $e_p^{rs}(t)$; the state variables are $x_{ap}^{rs}(t)$, $E_p^{rs}(t)$. We prove in the next section that the optimal solution to the model is in an instantaneous SDUO state.

Generally we assume the route choice dispersion parameter $\theta$ to be non-negative. As $\theta \to \infty$, the third and fourth terms in the objective function will vanish and the solution will approach the deterministic instantaneous user-optimal solution.

## 8.2.2 Optimality Conditions and Equivalence Analysis

### Optimality Conditions

We construct the extended Hamiltonian as follows.

$$\begin{aligned}
\mathcal{H} &= \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t),\omega]d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t),\omega]d\omega \right\} \\
&+ \frac{1}{\theta}\left[ \sum_{rs}\sum_p\sum_a u_{ap}^{rs}(t)\ln u_{ap}^{rs}(t) - \sum_{rs}\sum_p\sum_{a\notin B(s)} v_{ap}^{rs}(t)\ln v_{ap}^{rs}(t) \right] \\
&+ \sum_{rs}\sum_{ap}\lambda_{ap}^{rs}(t)[u_{ap}^{rs}(t) - v_{ap}^{rs}(t)] + \sum_r\sum_{s\neq r}\sum_p \nu_p^{rs}(t)e_p^{rs}(t) \\
&+ \sum_s\sum_{r\neq s}\sigma_r^{rs}(t)[f^{rs}(t) - \sum_{a\in A(r)}\sum_p u_{ap}^{rs}(t)] \\
&+ \sum_{rs}\sum_{j\neq rs}\sum_p \sigma_{jp}^{rs}(t)[\sum_{a\in B(j)} v_{ap}^{rs}(t) - \sum_{a\in A(j)} u_{ap}^{rs}(t)] \\
&+ \sum_r\sum_{s\neq r}\sigma_s^{rs}(t)[\sum_{a\in B(s)}\sum_p v_{ap}^{rs}(t) - e^{rs}(t)] \\
&+ \sum_{rs}\sum_p\sum_{a\in p}\mu_{ap}^{rs}(t)\left\{ x_{ap}^{rs}(t) + \sum_{b\in\tilde{p}} x_{bp}^{rs}(t) + E_p^{rs}(t) \right.
\end{aligned}$$

$$- \sum_{b \in \tilde{p}} x_{bp}^{rs}[t + \bar{\tau}_a(t)] - E_p^{rs}[t + \bar{\tau}_a(t)] \Bigg\}$$

where $\lambda_{ap}^{rs}(t)$ are Lagrange multipliers associated with the link state equations, $\nu_p^{rs}(t)$ are Lagrange multipliers associated with the destination node state equations, $\sigma_{jp}^{rs}(t)$ are Lagrange multipliers associated with the node flow conservation equations, and $\mu_{ap}^{rs}(t)$ are Lagrange multipliers associated with the flow propagation equations. For each link $a$ which points from node $l$ to node $m$, the first order necessary conditions of the instantaneous SDUO route choice program (8.10)-(8.23) include

$$\frac{\partial \mathcal{H}}{\partial u_{ap}^{rs}(t)} = g_{1a}[x_a(t), u_a(t)] + \frac{1}{\theta}\left[\ln u_{ap}^{rs}(t) + 1\right] + \lambda_{ap}^{rs}(t) - \sigma_{lp}^{rs}(t) \geq 0,$$

$$\forall a \in A(l), p, r, s, \qquad (8.24)$$

$$\text{and} \qquad u_{ap}^{rs}(t) \frac{\partial \mathcal{H}}{\partial u_{ap}^{rs}(t)} = 0 \qquad \forall a \in A(l), p, r, s; \qquad (8.25)$$

$$\frac{\partial \mathcal{H}}{\partial v_{ap}^{rs}(t)} = g_{2a}[x_a(t), v_a(t)] - \frac{1}{\theta}\left[\ln v_{ap}^{rs}(t) + 1\right] - \lambda_{ap}^{rs}(t) + \sigma_{mp}^{rs}(t) \geq 0,$$

$$\forall a \in B(m), m \neq s; p, r, s, \qquad (8.26)$$

$$\frac{\partial \mathcal{H}}{\partial v_{ap}^{rs}(t)} = g_{2a}[x_a(t), v_a(t)] - \lambda_{ap}^{rs}(t) + \sigma_{mp}^{rs}(t) \geq 0,$$

$$\forall a \in B(m), m = s; p, r, s, \qquad (8.27)$$

$$\text{and} \qquad v_{ap}^{rs}(t) \frac{\partial \mathcal{H}}{\partial v_{ap}^{rs}(t)} = 0 \qquad \forall a, p, r, s; \qquad (8.28)$$

$$\frac{\partial \mathcal{H}}{\partial e_p^{rs}(t)} \geq 0, \qquad \forall p, r, s, \qquad (8.29)$$

$$\text{and} \qquad e_p^{rs}(t) \frac{\partial \mathcal{H}}{\partial e_p^{rs}(t)} = 0 \qquad \forall p, r, s; \qquad (8.30)$$

$$\frac{d\lambda_{ap}^{rs}(t)}{dt} = -\frac{\partial \mathcal{H}}{\partial x_{ap}^{rs}(t)} \qquad \forall a, p, r, s; \qquad (8.31)$$

$$\frac{d\nu_p^{rs}(t)}{dt} = -\frac{\partial \mathcal{H}}{\partial E_p^{rs}(t)} \qquad \forall p, r, s; \qquad (8.32)$$

$$x_{ap}^{rs}(t) \geq 0, \qquad u_{ap}^{rs}(t) \geq 0, \qquad v_{ap}^{rs}(t) \geq 0, \qquad \forall a, p, r, s; \qquad (8.33)$$

$$e_p^{rs}(t) \geq 0, \qquad E_p^{rs}(t) \geq 0, \qquad \forall p, r, s. \qquad (8.34)$$

Note that $\sigma_{rp}^{rs}(t) = \sigma_r^{rs}(t)$ when node $l$ equals origin $r$.

**Equivalence Analysis**

We now show that the set of link states, inflows and exit flows that solves this program also satisfies the instantaneous stochastic dynamic user-optimal route choice conditions. This equivalence is demonstrated below by proving that the first order necessary conditions for optimal control program (8.10)-(8.23) are identical to the instantaneous stochastic dynamic user-optimal route choice conditions. The equivalence between the instantaneous SDUO route choice conditions and the first order necessary conditions of the optimal control program means that the instantaneous SDUO route choice conditions are satisfied at the optimal solution of this program.

Combining equations (8.24)-(8.25) with equations (8.26)-(8.28), the following equations can be derived for each link $a$ pointing from node $l$ to node $m$.

$$\frac{\partial \mathcal{H}}{\partial u_{ap}^{rs}(t)} + \frac{\partial \mathcal{H}}{\partial v_{ap}^{rs}(t)} = c_a(t) - \sigma_{lp}^{rs}(t) + \sigma_{mp}^{rs}(t) + \frac{1}{\theta}\left[\ln u_{ap}^{rs}(t) - \ln v_{ap}^{rs}(t)\right] \geq 0,$$

$$\forall a \in A(l) \cap B(m); p, r, s; \qquad (8.35)$$

$$\frac{\partial \mathcal{H}}{\partial u_{ap}^{rs}(t)} + \frac{\partial \mathcal{H}}{\partial v_{ap}^{rs}(t)} = c_a(t) - \sigma_{lp}^{rs}(t) + \sigma_{s}^{rs}(t) + \frac{1}{\theta}\left[\ln u_{ap}^{rs}(t) + 1\right] \geq 0,$$

$$\forall a \in A(l) \cap B(s); p, r, s. \qquad (8.36)$$

where node $m$ equals destination $s$ so that $\sigma_{mp}^{rs}(t) = \sigma_{s}^{rs}(t)$.

Furthermore, if $u_{ap}^{rs}(t) > 0$ and $v_{ap}^{rs}(t) > 0$, by (8.24)-(8.28),

$$c_a(t) = \sigma_{lp}^{rs}(t) - \sigma_{mp}^{rs}(t) - \frac{1}{\theta}\left[\ln u_{ap}^{rs}(t) - \ln v_{ap}^{rs}(t)\right]$$

$$\forall a \in A(l) \cap B(m); p, r, s; \qquad (8.37)$$

$$c_a(t) = \sigma_{lp}^{rs}(t) - \sigma_{s}^{rs}(t) - \frac{1}{\theta}\left[\ln u_{ap}^{rs}(t) + 1\right]$$

$$\forall a \in A(l) \cap B(s); p, r, s. \qquad (8.38)$$

For route $p$ between origin node $r$ and destination node $s$, let $i$ denote node $r$ or any intermediate node on this route. Denote route $\tilde{p}$ as a sequence of nodes $(i, 1, 2, \cdots, n, s)$ and also as a sequence of links $(1, 2, \cdots, k)$. The mean instantaneous travel time $\psi_{\tilde{p}}^{is}(t)$ for the remaining route $\tilde{p}$ between $i$ and $s$ is

$$\psi_{\tilde{p}}^{is}(t) = \sum_{a \in is\tilde{p}} c_a[x_a(t), u_a(t), v_a(t)] \qquad \forall i \in p, r, s. \qquad (8.39)$$

Consider a set of routes $p$ which are from $r \to i \to s$ and the corresponding set of subroutes $\tilde{p}$. The flow conservation constraint at node $i$ can be revised as

$$\sum_{a \in B(i)} v_{ap}^{rs}(t) = \sum_{a \in A(i)} u_{ap}^{rs}(t) \qquad \forall i, p, r, s; i \neq r, s. \qquad (8.40)$$

The fourth term in the Hamiltonian function should be revised as

$$\sum_{rs}\sum_{i\neq rs}\sigma_i^{rs}(t)\sum_p[\sum_{a\in B(j)} v_{ap}^{rs}(t) - \sum_{a\in A(j)} u_{ap}^{rs}(t)$$

so that $\sigma_{ip}^{rs}(t) = \sigma_i^{rs}(t)$ for the set of subroutes $\tilde{p}$. Note that all derivations from equation (8.24) to equation (8.38) follow for this set of subroutes $\tilde{p}$.

Now if $u_{ap}^{rs}(t) > 0$ and $v_{ap}^{rs}(t) > 0$, route $\tilde{p}$ is being used at time $t$. Thus, by equations (8.37)-(8.38),

$$\begin{aligned}
\psi_{\tilde{p}}^{is}(t) &= [\sigma_i^{rs}(t) - \sigma_1^{rs}(t)] + [\sigma_{1p}^{rs}(t) - \sigma_{2p}^{rs}(t)] + \cdots \\
&+ [\sigma_{n-1,p}^{rs}(t) - \sigma_{np}^{rs}(t)] + [\sigma_{np}^{rs}(t) - \sigma_s^{rs}(t)] \\
&- \frac{1}{\theta}[\ln u_{1\tilde{p}}^{rs}(t) - \ln v_{1\tilde{p}}^{rs}(t)] - \frac{1}{\theta}[\ln u_{2\tilde{p}}^{rs}(t) - \ln v_{2\tilde{p}}^{rs}(t)] - \cdots \\
&- \frac{1}{\theta}[\ln u_{(k-1)\tilde{p}}^{rs}(t) - \ln v_{(k-1)\tilde{p}}^{rs}(t)] - \frac{1}{\theta}[\ln u_{k\tilde{p}}^{rs}(t) + 1] \\
&= \sigma_i^{rs}(t) - \sigma_s^{rs}(t) - \frac{1}{\theta}[\ln u_{1\tilde{p}}^{rs}(t) + 1]
\end{aligned} \tag{8.41}$$

for every route $\tilde{p}$ being used at time $t$. In equation (8.41), for any intermediate node $j$ between upstream link $a$ and downstream link $b$,

$$v_{a\tilde{p}}^{rs}(t) = u_{b\tilde{p}}^{rs}(t) \tag{8.42}$$

Thus, in the second line of equation (8.41), all terms except $\ln u_{1\tilde{p}}^{rs}(t)$ are canceled. Reorganizing equation (8.41), we have

$$u_{1\tilde{p}}^{rs}(t) = \exp[-\theta\psi_{\tilde{p}}^{is}(t)]\exp[\theta\sigma_i^{rs}(t) - \theta\sigma_s^{rs}(t) - 1] \tag{8.43}$$

Summing the above equation over all subroutes $\tilde{q}$,

$$\sum_{\tilde{q}} u_{1\tilde{q}}^{rs}(t) = \exp[\theta\sigma_i^{rs}(t) - \theta\sigma_s^{rs}(t) - 1]\sum_{\tilde{q}}\exp[-\theta\psi_{\tilde{q}}^{is}(t)] \tag{8.44}$$

so that

$$\exp[\theta\sigma_i^{rs}(t) - \theta\sigma_s^{rs}(t) - 1] = \frac{\sum_{\tilde{q}} u_{1\tilde{q}}^{rs}(t)}{\sum_{\tilde{q}}\exp[-\theta\psi_{\tilde{q}}^{is}(t)]} \tag{8.45}$$

Combining equations (8.43) and (8.45) for any subroute $\tilde{p}$ of the set $\{\tilde{q}\}$,

$$u_{1\tilde{p}}^{rs}(t) = \sum_{\tilde{q}} u_{1\tilde{q}}^{rs}(t)\frac{\exp[-\theta\psi_{\tilde{p}}^{is}(t)]}{\sum_{\tilde{q}}\exp[-\theta\psi_{\tilde{q}}^{is}(t)]} \tag{8.46}$$

where $\sum_{\tilde{q}} u_{1\tilde{q}}^{rs}(t)$ is the total flow entering the first links on subroutes $\tilde{q}$ from origin $r$ arriving at node $i$ and departing toward destination $s$ at each time instant $t$; link 1 is the first link on subroute $\tilde{p}$ from $i$ to $s$.

Equation (8.46) holds for each remaining route $\tilde{p}$ between $i$ and $s$, where $i$ is any intermediate node (including the origin) between each O-D pair $rs$

in the network. For route $\tilde{p}$ connecting node $i$ and destination $s$, flows for each O-D pair are assigned according to the above logit function of measured instantaneous route travel times.

When $\theta \to \infty$, the variance of the perceived instantaneous route travel times is zero. Thus, by equations (8.7) and (8.41), we have

$$\Psi_{\tilde{p}}^{is}(t) \to \psi_{\tilde{p}}^{is}(t) = \sigma_i^{rs}(t) - \sigma_s^{rs}(t) \tag{8.47}$$

Since we assume $u_a^{rs}(t) > 0$ and $v_a^{rs}(t) > 0$, route $\tilde{p}$ is being used by definition. Thus, the routes which are being used at time $t$ have minimal instantaneous travel times equal to $[\sigma_i^{rs}(t) - \sigma_s^{rs}(t)]$. This is the case of deterministic instantaneous DUO route choice.

The deviations from the instantaneous DUO can be attributed to drivers' ignorance of (or non-compliance with) current travel time information provided. This is also a learning process. The value, $\theta = 0$, corresponds to total driver ignorance (no compliance) in route travel times in which case either route is equally probable. As drivers increase their compliance to the provided prevailing travel time information at each decision node, the flow on the shortest instantaneous travel time route will increase until an instantaneous DUO state is attained.

We note that the above analysis applies to any two decision nodes $i$ and $j$ between $r - s$. Since the intermediate node $i$ could be the origin node $r$, the above results also hold for routes from $r$ to $s$. It follows that

$$f_p^{rs}(t) = u_{1p}^{rs}(t) = f^{rs}(t)\frac{\exp[-\theta\psi_p^{rs}(t)]}{\sum_q \exp[-\theta\psi_q^{rs}(t)]} \tag{8.48}$$

With the above interpretation, it is now clear that equations (8.46) state the instantaneous stochastic dynamic user-optimal route choice conditions.

## 8.3 Ideal SDUO Route Choice Model

The ideal stochastic dynamic user-optimal (SDUO) route choice problem is to determine vehicle flows at each instant of time on each link resulting from drivers using minimal perceived actual travel time routes. In the following, we first discuss additional network flow constraints for the ideal SDUO route choice model.

### 8.3.1 Constraints for Mean Actual Route Travel Times

Denote the cumulative number of departing vehicles from origin $r$ to destination $s$ over route $p$ as a control variable $F_p^{rs}(t)$. We have

$$\frac{dF_p^{rs}(t)}{dt} = f_p^{rs}(t) \qquad \forall p, r \neq s, s. \tag{8.49}$$

At the initial time $t = 0$,

$$F_p^{rs}(0) = 0, \qquad \forall p, r, s. \tag{8.50}$$

Denote the indicator parameters $\delta_{ap}^{rs}$ as

$$\delta_{ap}^{rs} = \begin{cases} 1 & \text{if link } a \text{ is on route } p \text{ between O-D pair } rs \\ 0 & \text{otherwise.} \end{cases}$$

The flow conservation equations for origin $r$ can be rewritten as

$$f_p^{rs}(t) = \sum_{a \in A(r)} \delta_{ap}^{rs} \, u_{ap}^{rs}(t) \qquad \forall p, r, s; \tag{8.51}$$

$$\sum_p f_p^{rs}(t) = f^{rs}(t) \qquad \forall r \neq s, s. \tag{8.52}$$

The flow conservation equations for destination $s$ can be rewritten as

$$e_p^{rs}(t) = \sum_{a \in B(s)} \delta_{ap}^{rs} \, v_{ap}^{rs}(t) \qquad \forall p, r, s; \tag{8.53}$$

$$\sum_p e_p^{rs}(t) = e^{rs}(t) \qquad \forall r, s \neq r. \tag{8.54}$$

In the above equations, $f_p^{rs}(t)$, $u_{ap}^{rs}(t)$, $e_p^{rs}(t)$ and $v_{ap}^{rs}(t)$ are all control variables.

For any route $p$, the cumulative number of vehicles departing from origin $r$ by time $t$ must equal the number of vehicles arriving at destination $s$ over route $p$ by time $t + \eta_p^{rs}(t)$. It follows that

$$\int_o^t f_p^{rs}(\omega) d\omega = \int_o^{t+\eta_p^{rs}(t)} e_p^{rs}(\omega) d\omega \qquad \forall r, s, p; \tag{8.55}$$

or

$$F_p^{rs}(t) = E_p^{rs}[t + \eta_p^{rs}(t)] \qquad \forall r, s, p. \tag{8.56}$$

The cumulative number of departing vehicles $F_p^{rs}(t)$ and the cumulative number of arriving vehicles $E_p^{rs}(t)$ can be computed by using flow conservation equations at origins and destinations. Thus, the mean actual route travel time $\eta_p^{rs}(t)$ can be determined from the above equation, as shown in Figure 8.1. Taking the derivatives of the above equation with respect to time $t$, it follows that

$$e_p^{rs}[t + \eta_p^{rs}(t)][1 + \dot{\eta}_p^{rs}(t)] = f_p^{rs}(t) \qquad \forall p, r, s. \tag{8.57}$$

The term $\dot{\eta}_p^{rs}(t)$ is the rate of change of mean actual route travel time. When $\dot{\eta}_p^{rs}(t) = 0$, the mean actual route travel time is constant. However, this constraint for departure/arrival flow rates is redundant, because it can be derived using equation (8.56). This constraint is mentioned here so that the comparison of the ideal SDUO route choice model with the deterministic ideal DUO route choice model in Chapter 7 can be made in subsequent sections.

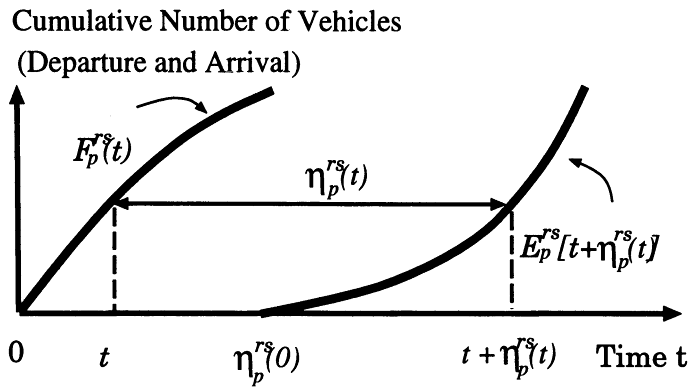



Figure 8.1: Relation of $F_p^{rs}(t)$ and $E_p^{rs}(t)$ to $\eta_p^{rs}(t)$

### 8.3.2 Definition of Ideal SDUO and Logit-Based Route Flow Constraints

For each origin-destination pair, there are many alternative efficient routes $p$, each with some mean actual travel time for flow departing from each origin to each destination at each time $t$. Due to variations in perception and exogenous factors, actual route travel times are perceived differently by each traveler. We assume the perceived actual route travel time to be a random variable for each time instant $t$. Given his or her perception of actual route travel time at each time instant $t$, each traveler is assumed to choose the route with minimal perceived actual travel time. Let $\Omega_p^{rs}(t)$ denote the perceived actual route travel time for flows departing origin $r$ toward destination $s$ on route $p$ at each time $t$. Then, we assume that

$$\Omega_p^{rs}(t) = \eta_p^{rs}(t) + \xi_p^{rs}(t) \qquad \forall r, s, p, \tag{8.58}$$

where $\xi_p^{rs}(t)$ is a random error term associated with the route $p$ under consideration. Furthermore, assume that $E[\xi_p^{rs}(t)] = 0$, i.e., $E[\Omega_p^{rs}(t)] = \eta_p^{rs}(t)$. In other words, the average perceived actual route travel time is the mean actual route travel time. Based on a probability density function for the actual route travel time at each time $t$, the ideal SDUO route choice problem is to determine how many travelers will use each route at each origin node at each instant of time. The share of travelers choosing route $p$ at time $t$, $P_p^{rs}(t)$, is given by

$$P_p^{rs}(t) = Pr[\Omega_p^{rs}(t) \leq \Omega_q^{rs}(t), \ \forall \text{ routes } q \text{ between } r \text{ and } s] \quad \forall r, s, p. \tag{8.59}$$

In other words, the probability that a given efficient route is chosen at time $t$ is the probability that its actual route travel time is perceived to be the lowest of all the alternative routes. Once the distribution of the perceived route travel times is specified, the probability of selecting each alternative route can be calculated and the time-dependent route flow can be assigned as follows:

$$f_p^{rs}(t) = f^{rs}(t)P_p^{rs}(t) \qquad \forall r, s, p. \tag{8.60}$$

In this section, we assume that

$$\Omega_p^{rs}(t) = \eta_p^{rs}(t) - \frac{1}{\theta}\epsilon_p^{rs}(t) \qquad \forall r, s, p, \tag{8.61}$$

where $\theta$ is a nonnegative parameter that scales the perceived actual route travel time, and $\epsilon_p^{rs}(t)$ is a random error term associated with the route $p$ under consideration. We assume that $\epsilon_p^{rs}(t)$ are identically and independently distributed (i.i.d.) Gumbel variates for each time instant. Other distributions of the perceived actual route travel times will be discussed in subsequent studies. Following random utility theory, a logit route choice probability can be expressed as

$$P_p^{rs}(t) = \frac{\exp[-\theta\eta_p^{rs}(t)]}{\sum_q \exp[-\theta\eta_q^{rs}(t)]} \qquad \forall r, s, p. \tag{8.62}$$

In this model, a route $p$ between $rs$ is defined as being used at time $t$ if $f_p^{rs}(t) > 0$. Thus, for each route $p$ being used at time $t$, the logit-based stochastic route flow is

$$f_p^{rs}(t) = f^{rs}(t)\frac{\exp[-\theta\eta_p^{rs}(t)]}{\sum_q \exp[-\theta\eta_q^{rs}(t)]} \qquad \forall r, s, p. \tag{8.63}$$

When $\theta \rightarrow \infty$, $var[\Omega_p^{rs}(t)] \rightarrow 0$, and the perceived actual travel time between O-D equals the mean actual travel time. In this case, travelers will choose the minimal actual travel time route at each time instant, as in the deterministic ideal DUO route choice model in Chapter 7. We then propose a definition of the ideal SDUO state which is the stochastic generalization of our previous deterministic *ideal* DUO state in Chapter 7.

> **Ideal SDUO State:** *If, for each O-D pair at each instant of time, the actual travel times perceived by travelers departing at the same time over used routes are equal and minimal, the dynamic traffic flow over the network is in an ideal stochastic dynamic user-optimal state.*

This definition represents another kind of perception of travel times by travelers. The above definition can also be called as a predictive (or anticipatory) SDUO since the mean actual route travel time is a predicted mean route travel time. Because we assume a Gumbel distribution of the perception errors, the equilibration of the perceived actual route travel times is ensured by the logit route flow constraints (8.63). Furthermore, because the mean actual route travel times $\eta_p^{rs}(t)$ can be determined by constraint conditions (8.56), the logit route flow constraints (8.63) are placed in the model directly without showing that the optimality conditions of the model are consistent with the ideal SDUO route choice conditions. Further discussion of the logit route flow constraints will be given later.

### 8.3.3 Model Formulation

Using optimal control theory, an optimization program of the ideal stochastic dynamic user-optimal route choice problem is formulated as follows.

$$\min_{u,v,x,e,E,f,F;\tau} \int_0^T \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t), \omega] d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t), \omega] d\omega \right\} dt \tag{8.64}$$

s.t.

Relationship between state and control variables:

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \tag{8.65}$$

$$\frac{dF_p^{rs}(t)}{dt} = f_p^{rs}(t) \qquad \forall p, r \neq s, s; \tag{8.66}$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r, s \neq r; \tag{8.67}$$

Flow conservation constraints:

$$f_p^{rs}(t) = \sum_{a \in A(r)} \delta_{ap}^{rs} \; u_{ap}^{rs}(t) \qquad \forall p, r \neq s, s; \tag{8.68}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \tag{8.69}$$

$$e_p^{rs}(t) = \sum_{a \in B(s)} \delta_{ap}^{rs} \; v_{ap}^{rs}(t) \qquad \forall p, r, s \neq r; \tag{8.70}$$

Logit route flow constraints:

$$f_p^{rs}(t) = f^{rs}(t) \frac{\exp[-\theta \eta_p^{rs}(t)]}{\sum_q \exp[-\theta \eta_q^{rs}(t)]} \qquad \forall r, s, p; \tag{8.71}$$

Constraints for mean actual route travel times:

$$F_p^{rs}(t) = E_p^{rs}[t + \eta_p^{rs}(t)] \qquad \forall r, s, p; \tag{8.72}$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[t + \tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall a \in p; p, r, s; \tag{8.73}$$

Definitional constraints:

$$\sum_p u_{ap}^{rs}(t) = u_a^{rs}(t), \qquad \sum_{rs} u_a^{rs}(t) = u_a(t), \qquad \forall a; \tag{8.74}$$

$$\sum_p v_{ap}^{rs}(t) = v_a^{rs}(t), \qquad \sum_{rs} v_a^{rs}(t) = v_a(t), \qquad \forall a; \tag{8.75}$$

$$\sum_p x_{ap}^{rs}(t) = x_a^{rs}(t), \qquad \sum_{rs} x_a^{rs}(t) = x_a(t), \qquad \forall a; \tag{8.76}$$

$$\sum_p e_p^{rs}(t) = e^{rs}(t), \qquad \sum_p E_p^{rs}(t) = E^{rs}(t), \qquad \forall r, s; \tag{8.77}$$

$$\sum_p f_p^{rs}(t) = f^{rs}(t), \qquad \sum_p F_p^{rs}(t) = F^{rs}(t), \qquad \forall r, s; \tag{8.78}$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \qquad u_{ap}^{rs}(t) \geq 0, \qquad v_{ap}^{rs}(t) \geq 0 \qquad \forall a, p, r, s; \tag{8.79}$$

$$e_p^{rs}(t) \geq 0, \quad f_p^{rs}(t) \geq 0, \quad E_p^{rs}(t) \geq 0, \quad F_p^{rs}(t) \geq 0, \qquad \forall p, r, s; \tag{8.80}$$

Boundary conditions:

$$F_p^{rs}(0) = 0, \quad E_p^{rs}(0) = 0 \quad \forall p, r, s; \qquad x_a^{rs}(0) = 0, \quad \forall a, r, s; \tag{8.81}$$

In this program, the two terms of the objective function are similar to the objective function of the well-known static user-optimal (UO) route choice model. The first three constraints (8.65)-(8.67) are state equations for flows on each link $a$, departures at origins, and arrivals at destinations. Equations (8.68)-(8.70) are flow conservation constraints at each node including origins and destinations. Equations (8.71) are logit route flow constraints which can be explicitly written as constraints in the model because the mean actual travel times $\eta_p^{rs}(t)$ are functionals and can be computed directly by equations (8.72). In contrast to this ideal SDUO route choice model, the stochastic route flow conditions in the instantaneous SDUO route choice model in Section 8.2 are satisfied indirectly as optimality conditions and need not be written as constraints directly in the model. Equations (8.72) are constraints for mean actual route travel times. The other constraints include flow propagation constraints, definitional constraints, nonnegativity and boundary conditions.

In summary, in this program the control variables are $u_{ap}^{rs}(t)$, $v_{ap}^{rs}(t)$, $e_p^{rs}(t)$, and $f_p^{rs}(t)$; the state variables are $x_{ap}^{rs}(t)$, $E_p^{rs}(t)$, and $F_p^{rs}(t)$; and the functionals are $\eta_p^{rs}(t)$. Compared with the deterministic ideal DUO route choice model in Chapter 7, stochastic route flow constraints (8.71) are added. Constraints (8.72) for the mean actual travel times replace similar constraints in the deterministic model.

### 8.3.4 Analysis of Dispersed Route Choice

Generally we assume $\theta$ to be nonnegative. In the following we will show that when $\theta \to \infty$, the solution approaches that of deterministic dynamic user-optimal route choice model based on actual route travel times in Chapter 7. Rewriting the logit-based route flow constraints (8.71), we have

$$\frac{f_p^{rs}(t)}{\exp[-\theta\eta_p^{rs}(t)]} = \frac{f^{rs}(t)}{\sum_m \exp[-\theta\eta_m^{rs}(t)]} \qquad \forall r, s, p. \tag{8.82}$$

Now assume that there is another route $q$ from $r$ to $s$ which has flow $f_q^{rs}(t)$ at each instant of time. Thus, it follows that

$$\frac{f_q^{rs}(t)}{\exp[-\theta\ \eta_q^{rs}(t)]} = \frac{f^{rs}(t)}{\sum_m \exp[-\theta\ \eta_m^{rs}(t)]} \qquad \forall r, s, q. \tag{8.83}$$

Comparing the above two equations, we have

$$\frac{f_p^{rs}(t)}{\exp[-\theta\ \eta_p^{rs}(t)]} = \frac{f_q^{rs}(t)}{\exp[-\theta\ \eta_q^{rs}(t)]} \qquad \forall p, q, r, s. \tag{8.84}$$

Assume that flows $f_p^{rs}(t)$ and $f_q^{rs}(t)$ are positive. Taking the logarithms of the above equation, it follows that

$$\ln f_p^{rs}(t) + \theta\ \eta_p^{rs}(t) = \ln f_q^{rs}(t) + \theta\ \eta_q^{rs}(t) \qquad \forall p, q, r, s. \tag{8.85}$$

Dividing the above equation by $\theta$, we have

$$\frac{1}{\theta} \ln f_p^{rs}(t) + \eta_p^{rs}(t) = \frac{1}{\theta} \ln f_q^{rs}(t) + \eta_q^{rs}(t) \qquad \forall p, q, r, s. \tag{8.86}$$

As $\theta \to \infty$, the first terms of both the left-hand-side and the right-hand-side will approach zero. Thus, we have

$$\eta_p^{rs}(t) = \eta_q^{rs}(t) \qquad \forall p, q, r, s. \tag{8.87}$$

The above equation demonstrates that for any O-D pair $rs$, any route carrying positive O-D departure flow has equal mean actual travel time. This is one of the optimality conditions for the deterministic ideal dynamic user-optimal route choice model in Chapter 7. Together with constraints (8.72) for the mean actual route travel times, these constraints for the ideal SDUO route choice model reduce to the corresponding optimality conditions for the deterministic ideal DUO route choice model when $\theta \to \infty$. Therefore we conclude that the optimality conditions of model (8.64)-(8.81) state the required ideal SDUO route choice conditions.

## 8.4 Notes

We have formulated two stochastic dynamic route choice problems as equivalent optimal control programs. The solutions of the optimal control models result in dispersed dynamic route choice governed by a logit distribution incorporating both mean instantaneous route travel times and mean actual route travel times. In stochastic dynamic route choice models, the O-D departure flows are dispersed across different travel time routes so that stochastic dynamic models represent more realistic travel choice behavior than deterministic models. This is one of the major merits of stochastic dynamic route choice models over deterministic dynamic route choice models.

In reality, drivers may rely on current information or predictive information to choose routes. Thus, drivers can be stratified into different groups. In each group, drivers may have different compliance to current information or future predictive information. In other words, a more realistic model could be a combined instantaneous and ideal SDUO route choice model with multiple groups of travelers.

We are investigating other SDUO route choice models based on different types of distributions for travel time perception errors. However, realistic distributions of travel time perception errors should depend on real data collected from IVHS operational tests such as the ADVANCE Project. Our models only provide a theoretical approach toward a better understanding of stochastic dynamic travel choices and travel time predictions.

It is expected that other SDUO route choice models based on different types of distributions of travel time perception errors will be developed. Figure 8.2 describes how realistic SDUO route choice models might be developed. Our future research also includes calibration of the route choice dispersion parameters and development of more general models incorporating stochastic mode and departure time choice. Using realistic link travel time functions in Chapter 16, we plan to implement our models on larger networks.

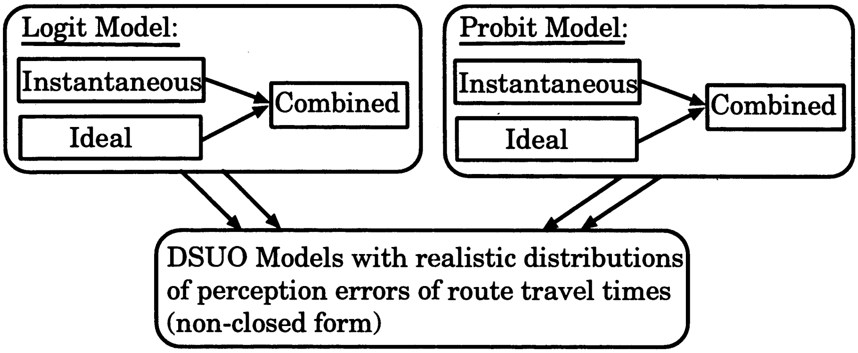



Figure 8.2: Toward Realistic SDUO Route Choice Models

# Chapter 9

# Solution Algorithms for Stochastic Dynamic User-Optimal Route Choice Models

Chapter 8 described two logit-type SDUO route choice models which are stochastic generalizations of deterministic dynamic user-optimal route choice models previously presented in Chapters 5 and 7. To solve these models for large networks, we need to develop solution algorithms avoiding route enumeration. Thus, the stochastic dynamic network algorithms in this chapter are link-based procedures that avoid route enumeration and perform dynamic assignments using only link and node variables. Some new notation is presented in Section 9.1. In Sections 9.2 and 9.3, two multiple-route dynamic route choice algorithms (DYNASTOCH) similar to Dial's efficient-route algorithm (STOCH) are suggested to solve two discrete-time *flow-independent* instantaneous and ideal SDUO route choice models. Then, the discrete formulation of the flow-dependent instantaneous SDUO route choice model is presented in Section 9.4 and a solution algorithm is presented to solve this model. In Section 9.5, the discrete formulation of ideal SDUO route choice model is presented and a solution algorithm is also proposed. In both solution algorithms for the *flow-dependent* SDUO route choice models, the DYNASTOCH algorithms are used to solve their subproblems so that explicit route enumeration can be avoided. The method of successive averages (MSA) and other methods are suggested to solve the one-dimensional search problems. Numerical examples are presented in Section 9.6 to illustrate the solution of the proposed algorithms.

## 9.1 Some New Notation

Denote the minimal free flow travel time from node $i$ to destination node $s$ as $\sigma_o^{is}$. An *efficient route* from origin $r$ to destination $s$ is redefined to include only links $a = (i, j)$ such that node $j$ is closer to destination $s$ than node $i$, i.e., $\sigma_o^{is} > \sigma_o^{js}$. In the following, all route-related constraints are defined using efficient routes. These routes need not be explicitly enumerated to solve our models. In our solution algorithms in Sections 9.2 and 9.3, we show that these routes can be implicitly generated.

To convert our continuous SDUO route choice problems into discrete formulations, the time period $[0, T]$ is subdivided into $K$ small time intervals. Each time interval is a unit of time. Our notation is summarized as follows:

| | | |
|---|---|---|
| $p, q$ | = | indexes for routes |
| $x_a(k)$ | = | number of vehicles on link $a$ at beginning of interval $k$ * |
| $u_a(k)$ | = | inflow rate into link $a$ during interval $k$ ** |
| $v_a(k)$ | = | exit flow rate from link $a$ during interval $k$ ** |
| $f^{rs}(k)$ | = | departure rate from origin $r$ toward destination $s$ during interval $k$ (given) |
| $F^{rs}(k)$ | = | cumulative number of departing vehicles from origin $r$ to destination $s$ at the beginning of interval $k$ (given) |
| $e^{rs}(k)$ | = | arrival rate at destination $s$ from origin $r$ during interval $k$ ** |
| $E^{rs}(k)$ | = | cumulative number of vehicles arriving at destination $s$ from origin $r$ at the beginning of interval $k$ * |
| $\tau_a(k)$ | = | mean actual travel time over link $a$ for flows entering link $a$ during interval $k$ |
| $\bar{\tau}_a(k)$ | = | mean estimated actual travel time over link $a$ for flows entering link $a$ during interval $k$ |
| $c_a(k)$ | = | mean instantaneous travel time on link $a$ during interval $k$ |
| $\psi_p^{rs}(k)$ | = | mean instantaneous travel time on route $p$ from $r$ to $s$ during interval $k$ |
| $\sigma^{ri}(k)$ | = | minimal mean instantaneous travel times for flows departing from origin $r$ to node $i$ during interval $k$ . |
| $\eta_p^{rs}(k)$ | = | mean actual travel time for flows departing from origin $r$ to destination $s$ over route $p$ during interval $k$ |
| $\pi^{ri}(k)$ | = | minimal mean actual travel times for flows departing from origin $r$ toward node $i$ during interval $k$ . |
| $\sigma_o^{is}$ | = | minimal free flow travel time from node $i$ to destination $s$ |
| $\delta_{ap}^{rs}$ | = | 1, if link $a$ is on route $p$ from $r$ to $s$; = 0, otherwise. |
| $\theta$ | = | route choice dispersion parameter |

* state variable
** control variable

The mean instantaneous travel time $\psi_p^{rs}(t)$ for each route $p$ from $r$ to $s$ is defined as

$$\psi_p^{rs}(k) = \sum_{a \in rsp} c_a[x_a(k), u_a(k), v_a(k)] \qquad \forall r, s, p; \tag{9.1}$$

where the summation is over all links $a$ in route $p$ from origin $r$ to destination $s$.

The minimal mean instantaneous route travel time $\sigma^{rs}(k) = \min\{\psi_p^{rs}(k)|p\}$.

Define $\eta_p^{rs}(k)$ as the mean actual travel time experienced over route $p$ by vehicles departing origin $r$ toward destination $s$ during time interval $k$. Once the mean actual link travel time $\tau_a(t)$ is determined, the mean actual route travel time $\eta_p^{rs}(k)$ is computed using the following approach. Assume route $p$ consists of nodes $(r, 1, 2, \cdots, i-1, i, \cdots, s)$. Denote $\eta_p^{ri}(k)$ as the mean actual travel time experienced over route $p$ by vehicles departing origin $r$ toward node $i$ during time interval $k$. Then, a recursive formula for mean actual route travel time $\eta_p^{rs}(k)$ is:

$$\eta_p^{ri}(k) = \eta_p^{r(i-1)}(k) + \tau_a[k + \eta_p^{r(i-1)}(k)] \qquad \forall p, r, i; i = 1, 2, \cdots, s;$$

where link $a = (i-1, i)$. The entering time on link $a$ is rounded as follows:

$$k + \eta_p^{r(i-1)}(k) \simeq i \quad \text{if} \quad i - 0.5 \leq k + \eta_p^{r(i-1)}(k) < i + 0.5,$$

where $i$ is an integer and $0 \leq i \leq K$. The impact of round-off errors on solutions will be discussed in subsequent studies. The minimal mean actual route travel time $\pi^{rs}(k) = \min\{\eta_p^{rs}(k)|p\}$.

In the following, we first discuss *flow-independent* SDUO route choice problems. In our general *flow-dependent* SDUO route choice problems, these problems function as subproblems in the solution procedure.

## 9.2 An Algorithm for the Flow-Independent Instantaneous SDUO Route Choice Model

We first discuss the *flow-independent* instantaneous SDUO route choice model and its solution algorithm. The flow-independent instantaneous SDUO route choice problem is to assign time-dependent vehicle flows between each O-D pair based on a route choice probability calculated using the fixed instantaneous route travel times for each time interval. We note that in this flow-independent problem, the instantaneous link travel times $c_a(k)$ do not depend on the link flow variables $u_a(k)$, $v_a(k)$ and $x_a(k)$.

The solution algorithm given here is similar to the algorithm proposed by Dial (1971) for static flow-independent stochastic network assignment. Our algorithm effectively implements a logit-type instantaneous route choice model at the network level. It does not assign probabilities and flows to all routes connecting each O-D pair. Instead, it is assumed that many of these routes constitute unreasonable travel choices that would not be actually considered. Consequently, our algorithm includes a preliminary phase which identifies the set of efficient routes connecting each O-D pair. The O-D departure flows are then assigned only to these routes during each time interval $k$, using the logit formula based on fixed instantaneous route travel times.

### 9.2.1 Statement of the Algorithm

The steps of this algorithm for one O-D pair $rs$ are outlined below. These steps should be repeated for each O-D pair in the network. In view of the correspondence to Dial's STOCH algorithm (1971), the following is called DYNASTOCH1 algorithm to represent a dynamic version of STOCH algorithm for the flow-independent instantaneous SDUO route choice problem.

**Step 0: Initialization.**
Compute the minimal instantaneous travel time $\sigma^{ri}(k)$ from origin $r$ to all other nodes for vehicles departing origin $r$ during time interval $k$. Calculate the likelihood for each link $a = (i,j)$ during time interval $k$:

$$L_{(i,j)}(k) = \begin{cases} \exp\{\theta\,[\sigma^{rj}(k) - \sigma^{ri}(k) - c_a(k)]\}, & \text{if } \sigma_o^{is} > \sigma_o^{js} \\ 0, & \text{otherwise} \end{cases}$$

In this expression, $c_a(k)$ is the mean instantaneous travel time on link $(i,j)$ during time interval $k$.

**Step 1: Backward Pass.**
By examining all nodes $j$ in ascending sequence with respect to $\sigma^{js}(k)$, calculate the weight for each link $a = (i,j)$ during each time interval $k$:

$$W_{(i,j)}(k) = \begin{cases} L_{(i,j)}(k), & \text{if } j = s \text{ (destination)} \\ L_{(i,j)}(k) \displaystyle\sum_{(j,m)\in A(j)} W_{(j,m)}(k), & \text{otherwise} \end{cases}$$

When the origin $r$ is reached, this step is completed.

**Step 2: Forward Pass.**
Consider all nodes $i$ in descending sequence with respect to $\sigma^{is}(k)$, starting with origin $r$. When each node $i$ is considered during each time interval $k$, compute the inflow of link $(i,j)$ during each time interval $k$ using the following formula:

$$u_{(i,j)}(k) = \begin{cases} f^{rs}(k)\dfrac{W_{(i,j)}(k)}{\sum_{(i,m)\in A(i)} W_{(i,m)}(k)}, & \text{if } i = r \text{ (origin)} \\ \left\{\displaystyle\sum_{(n,i)\in B(i)} v_{(n,i)}(k)\right\} \dfrac{W_{(i,j)}(k)}{\sum_{(i,m)\in A(i)} W_{(i,m)}(k)}, & \text{otherwise} \end{cases}$$

This step is implemented iteratively until destination $s$ is reached. Note that the sum in each denominator includes all links emanating from the upstream node of the link under consideration. The sum of the exit flow variables is taken over all links arriving at the upstream node of the link under consideration.

The flow generated by this algorithm is equivalent to a logit-based flow-independent route assignment between each O-D pair, given that only efficient routes are considered. We next present a proof for this algorithm.

## 9.2.2 Proof of the Algorithm

We now prove that the above algorithm does generate logit-based flow-independent instantaneous SDUO route choices between each O-D pair. We note that each link likelihood $L_{(i,j)}(k)$, is proportional to the logit probability that link $a = (i, j)$ is used during interval $k$ by a traveler chosen at random from among the population of trip-makers between $r$ and $s$, given that the traveler is departing from origin $r$ during interval $k$. The probability that a given route will be used is proportional to the product of all the likelihoods of the links comprising this route. The probability of using route $l$ between $r$ and $s$, $P_l^{rs}(k)$, for vehicles departing $r$ during interval $k$, is then

$$P_l^{rs}(k) = G(k) \prod_{a \in l} \left\{ L_{(i,j)}(k) \right\}^{\delta_{al}^{rs}} \tag{9.2}$$

where $G(k)$ is a proportionality constant for each time interval $k$ and the product is taken over all links in the network. The incidence variable $\delta_{al}^{rs}$ ensures that $P_l^{rs}(k)$ will include only those links in the $l$th route between $r$ and $s$. Substituting the expression for the likelihood $L_{(i,j)}(k)$ in the above equation, the choice probability of choosing a particular efficient route becomes

$$\begin{aligned} P_l^{rs}(k) &= G(k) \prod_{a \in l} \exp\left\{ \theta \left\{ \sigma^{rj}(k) - \sigma^{ri}(k) - c_a(k) \right\} \delta_{al}^{rs} \right\} && (9.3) \\ &= G(k) \exp\left\{ \theta \sum_{a \in l} \left\{ \sigma^{rj}(k) - \sigma^{ri}(k) - c_a(k) \right\} \delta_{al}^{rs} \right\} && (9.4) \\ &= G(k) \exp\left\{ \theta \left[ \sigma^{rs}(k) - \psi_l^{rs}(k) \right] \right\} && (9.5) \end{aligned}$$

The last equality results from the following summations:

$$\sum_{a \in l} \{\sigma^{rj}(k) - \sigma^{ri}(k)\} = \sigma^{rs}(k) - \sigma^{rr}(k) = \sigma^{rs}(k) \tag{9.6}$$

and

$$\sum_{a \in l} c_a(k) = \psi_l^{rs}(k) \tag{9.7}$$

Since

$$\sum_q P_q^{rs}(k) = 1.0, \tag{9.8}$$

the proportionality constant must equal

$$G(k) = \frac{1}{\sum_q \exp\left\{ \theta \left[ \sigma^{rs}(k) - \psi_q^{rs}(k) \right] \right\}} \tag{9.9}$$

Thus,

$$P_l^{rs}(k) = \frac{\exp\{\theta\,[\sigma^{rs}(k) - \psi_l^{rs}(k)]\}}{\sum\limits_q \exp\{\theta\,[\sigma^{rs}(k) - \psi_q^{rs}(k)]\}} \tag{9.10}$$

$$= \frac{\exp\{-\theta\,\psi_l^{rs}(k)]\}}{\sum\limits_q \exp\{-\theta\,\psi_q^{rs}(k)\}} \tag{9.11}$$

The above equation (9.11) depicts a logit model of route choice during each time interval $k$ among the efficient routes connecting O-D pair $rs$. Thus, the algorithm does generate a logit-type route choice probability formula using instantaneous route travel times.

The DYNASTOCH1 algorithm does not require explicit route enumeration. It does require the calculation of time-dependent minimal instantaneous travel time routes for every O-D pair in the network during each time interval $k$. In the following, we show that the forward pass of the algorithm will generate a time-dependent logit-based assignment. To do so, we only need to show that the calculated link inflows are obtained in a manner consistent with the expression in the forward pass. This is done by proving the algorithm diverts trips from each node $i$ during each interval $k$ according to appropriate conditional link probabilities. A conditional link probability during interval $k$ is the probability that a trip between $r$ and $s$ will use a particular link $a = (i, j)$ during interval $k$, given that it goes through the link's tail node $i$. This probability is stated as

$$\text{Prob}_{(i,j)|i}(k) = \frac{\text{Prob}_{(i,j),i}(k)}{\text{Prob}_i(k)} = \frac{\text{Prob}_{(i,j)}(k)}{\text{Prob}_i(k)} = \frac{\text{Prob}_{(i,j)}(k)}{\sum\limits_l \text{Prob}_{(i,l)}(k)} \tag{9.12}$$

Denote $P$ as an efficient route from $r$ to $s$. The probability of using link $a = (i, j)$ during interval $k$ is the summation of probabilities of using route $P$ during interval $k$, where link $a = (i, j)$ is on route $P$. It follows that

$$\text{Prob}_{(i,j)}(k) = \sum_{P:\ (i,j) \text{ in } P} \text{Prob}_P(k) \tag{9.13}$$

It is useful to write (9.13) in a more elaborate form, so as to facilitate cancellation of common factors in the numerator and denominator of equation (9.12). To do so, an efficient route through link $(i, j)$ can be partitioned into three sets of links:

1. $P_i = \{$all links topologically preceding link $(i, j)\}$;
2. link $(i, j) = \{(i, j)\}$;
3. $P_j = \{$all links topologically following link $(i, j)\}$.

Denote $\mathcal{P}_i$ as the set of $P_i$ and $\mathcal{P}_j$ as the set of $P_j$. Then

$$\sum_{P:\,(i,j)\text{ in }P} \text{Prob}_P(k) = h(k)\, L_{(i,j)}(k) \left\{ \sum_{P\in\mathcal{P}_i} \prod_{(m,n)} L_{(m,n)}(k) \right\} \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(k) \right\} \tag{9.14}$$

Equation (9.14) follows from the fact that all the efficient routes can be constructed by independently choosing a member from each of $\mathcal{P}_i$ and $\mathcal{P}_j$ and putting link $(i,j)$ in between. Such combinations constitute efficient routes. Substituting (9.13) and (9.14) into (9.12), it follows that

$$\begin{aligned}
&\text{Prob}_{(i,j)|i}(k) \\
&= \frac{h(k)\, L_{(i,j)}(k) \left\{ \sum_{P\in\mathcal{P}_i} \prod_{(m,n)} L_{(m,n)}(k) \right\} \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(k) \right\}}{\sum_l h(k)\, L_{(i,l)}(k) \left\{ \sum_{P\in\mathcal{P}_i} \prod_{(m,n)} L_{(m,n)}(k) \right\} \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(k) \right\}} \\
&= \frac{h(k)\, L_{(i,j)}(k) \left\{ \sum_{P\in\mathcal{P}_i} \prod_{(m,n)} L_{(m,n)}(k) \right\} \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(k) \right\}}{h(k) \left\{ \sum_{P\in\mathcal{P}_i} \prod_{(m,n)} L_{(m,n)}(k) \right\} \sum_l \left\{ L_{(i,l)}(k) \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(k) \right\} \right\}} \\
&= \frac{L_{(i,j)}(k) \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(k) \right\}}{\sum_l \left\{ L_{(i,l)}(k) \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(k) \right\} \right\}} = \frac{W_{(i,j)}(k)}{\sum_l W_{(i,l)}(k)}
\end{aligned} \tag{9.15}$$

By the link weights calculated in the Backward Pass (Step 1), the quotient in Forward Pass (Step 2) is equal to the right-hand-side of equation (9.15). Thus, we complete the proof that the diversion inflows are indeed those implied by the probability defined in equation (9.2). Therefore, our algorithm does generate a logit-type route flow assignment using instantaneous route travel times. The proof is complete. We also note that the flow over each link should satisfy the flow propagation constraints. For a detailed discussion of these constraints, we refer readers to Chapter 4.

## 9.3 An Algorithm for the Flow-Independent Ideal SDUO Route Choice Model

We now discuss the *flow-independent* ideal SDUO route choice model and its solution algorithm. The flow-independent ideal SDUO route choice problem is to assign vehicle flows between each O-D pair based on a route choice probability calculated using the given actual route travel times for each time interval. We note that the ideal SDUO route choice problem is different from the instantaneous SDUO route choice problem because it uses the actual route travel times instead of instantaneous route travel times.

The solution algorithm given here is also similar to Dial's algorithm for static flow-independent stochastic network assignment. This algorithm effectively implements a logit-type ideal dynamic route choice model at the network level. As in the instantaneous model, it assigns probabilities and flows to efficient routes connecting each O-D pair for each time interval. This algorithm also includes a preliminary phase which identifies the set of efficient routes connecting each O-D pair. The O-D departure flows are then assigned only to these routes during each interval $k$, using the logit formula based on actual route travel times. This algorithm is very similar to the algorithm presented in Section 9.2 for the flow-independent instantaneous SDUO route choice model. However, in this algorithm, all probabilities $\text{Prob}_{(i,j)|i}(t)$ for assigning inflows to link $(i,j)$ are evaluated during time interval $t$ instead of time interval $k$. Here, interval $t$ is the arrival time interval at node $i$. Thus, all derivations in DYNASTOCH1 must be revised using time interval $t$ instead of time interval $k$.

### 9.3.1 Statement of the Algorithm

The steps of this algorithm for one O-D pair $rs$ are outlined below. These steps must be repeated for each O-D pair in the network. Similar to the DYNASTOCH1 algorithm in Section 9.2, the following procedure is called the DYNASTOCH2 algorithm to represent its dynamic version for the flow-independent ideal SDUO route choice problem.

**Step 0: Initialization.**

Compute the minimal actual travel time $\pi^{js}(t)$ from node $j$ to destination $s$ for vehicles departing node $j$ during interval $t$. Calculate the likelihood for each link $(i,j)$ during each interval $t$:

$$L_{(i,j)}(t) = \begin{cases} \exp\{\theta\,[\pi^{is}(t) - \pi^{js}[t+\tau_a(t)] - \tau_a(t)]\}, & \text{if } \sigma_o^{is} > \sigma_o^{js} \\ 0, & \text{otherwise} \end{cases}$$

In this expression, $\tau_a(t)$ is the mean actual travel time on link $a = (i,j)$ during time interval $t$.

**Step 1: Backward Pass.**

By examining all nodes $j$ in ascending sequence with respect to $\pi^{js}(t)$, calculate the weight for each link $a = (i, j)$ during each time interval $t$:

$$W_{(i,j)}(t) = \begin{cases} L_{(i,j)}(t), & \text{if } j = s \text{ (destination)} \\ L_{(i,j)}(t) \sum\limits_{(j,m)\in A(j)} W_{(j,m)}[t + \tau_a(t)], & \text{otherwise} \end{cases}$$

When the origin $r$ is reached, this step is completed.

**Step 2: Forward Pass.**

Consider all nodes $i$ in descending sequence with respect to $\pi^{is}(t)$, starting with origin $r$. When each node $i$ is considered during each time interval $t$, compute the inflow of link $(i, j)$ during each interval $t$ using the following formula:

$$u_{(i,j)}(t) =$$

$$\begin{cases} f^{rs}(t) \quad \dfrac{W_{(i,j)}(t)}{\sum\limits_{(i,l)\in A(i)} W_{(i,l)}(t)}, & \text{if } i = r \text{ (origin)} \\ \left\{ \sum\limits_{(n,i)\in B(i)} v_{(n,i)}(t) \right\} \quad \dfrac{W_{(i,j)}(t)}{\sum\limits_{(i,l)\in A(i)} W_{(i,l)}(t)}, & \text{otherwise} \end{cases}$$

This step is implemented iteratively until destination $s$ is reached. Note that the sum in the denominator includes all links emanating from the upstream node of the link under consideration. The sum of the exit flow variables is taken over all links arriving at the upstream node of the link under consideration.

The flow generated by this algorithm is equivalent to a logit-based flow-independent route assignment between each O-D pair, given that only reasonable routes are considered. The proof for this algorithm is similar to that for the instantaneous SDUO route choice model. In the following, we present a proof for this algorithm.

## 9.3.2 Proof of the Algorithm

We now prove that the algorithm does generate logit-based flow-independent ideal SDUO route choices between each O-D pair. We note that each link likelihood $L_{(i,j)}(t)$, is proportional to the logit probability that link $a = (i, j)$ is used during time interval $t$ by a traveler chosen at random from among the population of trip-makers between $r$ and $s$, given that the traveler is at node $i$ during interval $t$. The probability that a given route will be used is proportional to the product of all the likelihoods of the links comprising this route. Suppose route $l$ consists of nodes $(r, 1, 2, \cdots, i-1, i, \cdots, n, s)$ and links $(1, 2, \cdots, h)$. The probability of using route $l$ between $r$ and $s$, $P_l^{rs}(k)$, for vehicles departing $r$ during interval $k$ and arriving at node $i$ during interval $t$, is then

$$P_l^{rs}(k) = G(k) \prod_{a \in l} \{L_{(i,j)}(t)\}^{\delta_{al}^{rs}} \tag{9.16}$$

where $G(k)$ is a proportionality constant for each time interval $k$ and the product is taken over all links in the network. Here, $t = k + \eta_l^{ri}(k)$. The incidence variable $\delta_{al}^{rs}$ ensures that $P_l^{rs}(k)$ will include only those links in the $l$th route between $r$ and $s$. Substituting the expression for the likelihood $L_{(i,j)}(t)$ in the above equation, the choice probability of choosing a particular efficient route becomes

$$\begin{aligned} P_l^{rs}(k) &= G(k) \prod_{a \in l} \exp\left\{\theta \left\{\pi^{is}(t) - \pi^{js}[t + \tau_a(t)] - \tau_a(t)\right\} \delta_{al}^{rs}\right\} & (9.17) \\ &= G(k) \exp\left\{\theta \sum_{a \in l} \left\{\pi^{is}(t) - \pi^{js}[t + \tau_a(t)] - \tau_a(t)\right\} \delta_{al}^{rs}\right\} & (9.18) \\ &= G(k) \exp\left\{\theta \left[\pi^{rs}(k) - \eta_l^{rs}(k)\right]\right\} & (9.19) \end{aligned}$$

The last equality results from the following summations:

$$\begin{aligned} \sum_{a \in l} \{\pi^{is}(t) - \pi^{js}[t + \tau_a(t)]\} &= \pi^{rs}(k) - \pi^{1s}[k + \eta_l^{r1}(k)] \\ &+ \pi^{1s}[k + \eta_l^{r1}(k)] - \pi^{2s}[k + \eta_l^{r2}(k)] \\ &+ \vdots \\ &+ \pi^{ns}[k + \eta_l^{rn}(k)] - \pi^{ss}[k + \eta_l^{rs}(k)] \\ &= \pi^{rs}(k) - \pi^{ss}[k + \eta_l^{rs}(k)] \\ &= \pi^{rs}(k) \end{aligned} \tag{9.20}$$

and

$$\sum_{a \in l} \tau_a(t) = \tau_1(k) + \tau_2[k + \eta_l^{r1}(k)] + \cdots + \tau_h[k + \eta_l^{rn}(k)] = \eta_l^{rs}(k) \tag{9.21}$$

Since

$$\sum_q P_q^{rs}(k) = 1.0, \tag{9.22}$$

the proportionality constant must equal

$$G(k) = \frac{1}{\sum_q \exp\left\{\theta \left[\pi^{rs}(k) - \eta_q^{rs}(k)\right]\right\}} \tag{9.23}$$

Thus,

$$P_l^{rs}(k) = \frac{\exp\{\theta\,[\pi^{rs}(k) - \eta_l^{rs}(k)]\}}{\sum_q \exp\{\theta\,[\pi^{rs}(k) - \eta_q^{rs}(k)]\}} \tag{9.24}$$

$$= \frac{\exp\{-\theta\,\eta_l^{rs}(k)]\}}{\sum_q \exp\{-\theta\,\eta_q^{rs}(k)\}} \tag{9.25}$$

Equation (9.25) depicts a logit model of route choice during each interval $k$ among the efficient routes connecting O-D pair $rs$. This algorithm does generate a route choice probability using actual route travel times.

The DYNASTOCH2 algorithm does not require route enumeration. It does require the calculation of time-dependent minimal actual travel time routes for every O-D pair in the network during each time interval $k$. In the following, we show that the forward pass of the algorithm will generate a time-dependent logit-based route flow assignment. To do so, we only need to show that the calculated link inflows are obtained in a manner consistent with the expression in the forward pass. This is done by proving the algorithm diverts trips from each node $i$ during each time interval $t$ (arrival time interval at node $i$) according to appropriate conditional link probabilities. A conditional link probability during interval $t$ is the probability that a trip between $r$ and $s$ will use a particular link $a = (i,j)$ during interval $t$, given it goes through the link's tail node $i$ during interval $t$. This probability is stated as

$$\text{Prob}_{(i,j)|i}(t) = \frac{\text{Prob}_{(i,j),i}(t)}{\text{Prob}_i(t)} = \frac{\text{Prob}_{(i,j)}(t)}{\text{Prob}_i(t)} = \frac{\text{Prob}_{(i,j)}(t)}{\sum_j \text{Prob}_{(i,j)}(t)} \tag{9.26}$$

The probability of using link $a = (i,j)$ during interval $t$ is the summation of probabilities of departure trips using route $P$ at origin $r$ during interval $k$, where link $a = (i,j)$ is on route $P$, given that the traveler reaches node $i$ during interval $t = k + \eta_P^{ri}(k)$. It follows that

$$\text{Prob}_{(i,j)}(t) = \sum_{P:\ (i,j)\text{ in }P} \text{Prob}_P(t) \tag{9.27}$$

It is more useful to write (9.27) in a more elaborate form, so as to facilitate cancellation of common factors in the numerator and denominator of equation (9.26). To do so, an efficient route through link (i,j) can be partitioned into three sets of links:

1. $P_i = \{$all links topologically preceding link $(i,j)\}$;
2. link $(i,j) = \{(i,j)\}$;
3. $P_j = \{$all links topologically following link $(i,j)\}$.

Denote $\mathcal{P}_i$ as the set of $P_i$ and $\mathcal{P}_j$ as the set of $P_j$. Then

$$\sum_{P:\,(i,j)\text{ in }P} \text{Prob}_P(t) = h(t)\, L_{(i,j)}(t) \left\{ \sum_{P\in\mathcal{P}_i} \prod_{(m,n)} L_{(m,n)}(t') \right\} \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(t'') \right\} \quad (9.28)$$

Note that $t' + \eta_P^{mi}(t') = t$ where $t' < t$, and $t + \eta_P^{im}(t) = t''$ where $t'' > t$. Equation (9.28) follows from the fact that all efficient routes can be constructed by independently choosing a member from each of $\mathcal{P}_i$ and $\mathcal{P}_j$ and putting link $(i,j)$ in between, given that flows arrive at node $i$ during interval $t$. Such combinations constitute efficient routes. Substituting (9.27) and (9.28) into (9.26), it follows that

$$\begin{aligned}
&\text{Prob}_{(i,j)|i}(t) \\
&= \frac{h(t)\, L_{(i,j)}(t) \left\{ \sum_{P\in\mathcal{P}_i} \prod_{(m,n)} L_{(m,n)}(t') \right\} \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(t'') \right\}}{\sum_l h(t)\, L_{(i,l)}(t) \left\{ \sum_{P\in\mathcal{P}_i} \prod_{(m,n)} L_{(m,n)}(t') \right\} \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(t'') \right\}} \\
&= \frac{h(t)\, L_{(i,j)}(t) \left\{ \sum_{P\in\mathcal{P}_i} \prod_{(m,n)} L_{(m,n)}(t') \right\} \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(t'') \right\}}{h(t) \left\{ \sum_{P\in\mathcal{P}_i} \prod_{(m,n)} L_{(m,n)}(t') \right\} \sum_l \left\{ L_{(i,l)}(t) \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(t'') \right\} \right\}} \\
&= \frac{L_{(i,j)}(t) \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(t'') \right\}}{\sum_l \left\{ L_{(i,l)}(t) \left\{ \sum_{P\in\mathcal{P}_j} \prod_{(m,n)} L_{(m,n)}(t'') \right\} \right\}} = \frac{W_{(i,j)}(t)}{\sum_{(i,l)} W_{(i,l)}(t)}
\end{aligned} \quad (9.29)$$

where the numerator of (9.29) is the definition of $W_{(i,j)}(t)$. By the link weights calculated in the Backward Pass (Step 1), the quotient in Forward Pass (Step 2) is equal to the right-hand-side of equation (9.29). Thus, we complete the proof that the diversion inflows are indeed those implied by the probabilities defined in equation (9.16). We also note the flow over each link should satisfy the flow propagation constraints.

# 9.4 An Algorithm for the Instantaneous SDUO Route Choice Model

## 9.4.1 A Discrete Time Instantaneous Model

To convert our instantaneous SDUO route choice problem into an NLP, the time period $[0, T]$ is subdivided into $K$ small time intervals. (The time intervals are not necessarily equal.) To simplify the formulation, we modify the estimated mean actual travel time on each link in the following way so that each estimated mean travel time is equal to a multiple of the time interval.

$$\bar{\tau}_a(k) = i \quad \text{if} \quad i - 0.5 \leq \bar{\tau}_a(k) < i + 0.5,$$

where $i$ is an integer and $0 \leq i \leq K$.

The optimal control program presented in Chapter 8 can then be reformulated as a discrete time NLP as follows:

$$\min_{u,v,x,e,E} \quad \sum_{k=1}^{K} \sum_{a} \Bigg\{ \int_0^{u_a(k)} g_{1a}[x_a(k), \omega] d\omega + \int_0^{v_a(k)} g_{2a}[x_a(k), \omega] d\omega$$
$$+ \frac{1}{\theta} [\sum_{rs} \sum_{p} \sum_{a} u_{ap}^{rs}(k) \ln u_{ap}^{rs}(k)$$
$$- \sum_{rs} \sum_{p} \sum_{a \notin B(s)} v_{ap}^{rs}(k) \ln v_{ap}^{rs}(k)] \Bigg\} \tag{9.30}$$

s.t.

Relationship between state and control variables:

$$x_{ap}^{rs}(k+1) = x_{ap}^{rs}(k) + u_{ap}^{rs}(k) - v_{ap}^{rs}(k) \qquad \forall a, p, r, s; k = 1, \cdots, K; \tag{9.31}$$

$$E^{rs}(k+1) = E^{rs}(k) + \sum_{a \in B(s)} \sum_{p} v_{ap}^{rs}(k) \qquad \forall r, s; k = 1, \cdots, K; \tag{9.32}$$

Flow conservation constraints:

$$\sum_{a \in A(r)} \sum_{p} u_{ap}^{rs}(k) = f^{rs}(k) \qquad \forall r, s; k = 1, \cdots, K; \tag{9.33}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(k) - \sum_{a \in A(j)} u_{ap}^{rs}(k) = 0 \quad \forall j, p, r, s; j \neq r, s; k = 1, \cdots, K; \tag{9.34}$$

Flow propagation constraints:

$$x_{ap}^{rs}(k) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[k + \bar{\tau}_a(k)] - x_{bp}^{rs}(k)\} + \{E_p^{rs}[k + \bar{\tau}_a(k)] - E_p^{rs}(k)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; k = 1, \cdots, K+1; \quad (9.35)$$

Definitional constraints:

$$\sum_p u_{ap}^{rs}(k) = u_a^{rs}(k), \qquad \sum_{rs} u_a^{rs}(k) = u_a(k), \qquad \forall a; k = 1, \cdots, K; \quad (9.36)$$

$$\sum_p v_{ap}^{rs}(k) = v_a^{rs}(k), \qquad \sum_{rs} v_a^{rs}(k) = v_a(k), \qquad \forall a; k = 1, \cdots, K; \quad (9.37)$$

$$\sum_p x_{ap}^{rs}(k) = x_a^{rs}(k), \qquad \sum_{rs} x_a^{rs}(k) = x_a(k), \qquad \forall a; k = 1, \cdots, K+1; \quad (9.38)$$

$$\sum_p E_p^{rs}(k+1) = E^{rs}(k+1), \qquad \forall r, s; k = 1, \cdots, K; \quad (9.39)$$

Nonnegativity conditions:

$$x_{ap}^{rs}(k+1) \geq 0, \quad u_{ap}^{rs}(k) \geq 0, \quad v_{ap}^{rs}(k) \geq 0, \quad \forall a, p, r, s; k = 1, \cdots, K; \quad (9.40)$$

$$E_p^{rs}(k+1) \geq 0, \qquad \forall p, r, s; k = 1, \cdots, K; \quad (9.41)$$

Boundary conditions:

$$E_p^{rs}(1) = 0 \quad \forall p, r, s; \qquad x_{ap}^{rs}(1) = 0, \qquad \forall a, p, r, s. \quad (9.42)$$

### 9.4.2 Solution Algorithm

Since the objective function involves route flow variables $u_{ap}^{rs}(k)$ and $v_{ap}^{rs}(k)$, the objective function cannot be computed using link flow variables after solving the subproblem. Thus, as for stochastic *static* route choice models, the method of successive averages (MSA) (Powell and Sheffi, 1982) is suggested to solve the one-dimensional search problem. Other one-dimensional search methods suggested by Chen and Alfa (1991) for stochastic static route choice models can also be used.

As discussed in Chapter 8, the first-order necessary conditions of the continuous time model fit a logit-type instantaneous route choice model. This flow-independent subproblem can be efficiently solved using the DYNASTOCH1 algorithm discussed in Section 9.1 without route enumeration. We note that the DYNASTOCH1 algorithm is much different from Dial's STOCH algorithm in that DYNASTOCH1 has to be implemented on an expanded time-space network presented in Chapter 6 and the flow propagation constraints ensure the flow progression over links. In the following algorithm, the inner iteration involves direction finding and flow variable updates; and the outer iteration involves updating estimates of actual link travel time $\bar{\tau}_a(k)$ in the flow propagation constraints. The flowchart of the algorithm for solving our instantaneous SDUO route choice model is shown in Figure 9.1 and the algorithm is summarized as follows:

**Step 0: Initialization.**

Find an initial feasible solution $\{x_a^{(1)}(k)\}$, $\{u_a^{(1)}(k)\}$, $\{v_a^{(1)}(k)\}$, $\{E^{(1)}(k)\}$. Set the outer iteration counter m = 1.

**Step 1: Diagonalization. (Outer Iteration.)**

Find a new estimate of the actual link travel time $\bar{\tau}_a^{(m)}(k)$ for the flow propagation constraint and solve the flow-dependent instantaneous SDUO route choice program as follows. Set the inner iteration counter n = 1.

[Step 1.1]: *Update.* Calculate the link travel time functions using the temporarily fixed link flow variables.

[Step 1.2]: *Direction Finding.* Implement the DYNASTOCH1 algorithm for the logit-based dynamic route flows based on the temporarily fixed instantaneous link travel time and find an auxiliary set of link flow variables.

[Step 1.3]: *Move.* Use the step size $\alpha^{(n)} = 1/(n+1)$ generated by MSA to find a new solution by combining current solution $\{u_a(k)\}$, $\{v_a(k)\}$, $\{x_a(k)\}$, $\{E^{rs}(k)\}$ and previous solution $\{p_a(k)\}$, $\{q_a(k)\}$, $\{y_a(k)\}$, $\{\bar{E}^{rs}(k)\}$ as follows.

$$u_a^{(n+1)}(k) = u_a^{(n)}(k) + [p_a^{(n)}(k) - u_a^{(n)}(k)]/(n+1)$$

$$v_a^{(n+1)}(k) = v_a^{(n)}(k) + [q_a^{(n)}(k) - v_a^{(n)}(k)]/(n+1)$$

$$x_a^{(n+1)}(k) = x_a^{(n)}(k) + [y_a^{(n)}(k) - x_a^{(n)}(k)]/(n+1)$$

$$E_a^{(n+1)}(k) = E_a^{(n)}(k) + [\bar{E}_a^{(n)}(k) - E_a^{(n)}(k)]/(n+1)$$

[Step 1.4]: *Convergence Test for Inner Iterations.* If n equals a prespecified number, go to step 2; otherwise, set $n = n+1$ and go to step 1.1.

**Step 2: Convergence Test for Outer Iterations.**

If $\bar{\tau}_a^{(m)}(k) \simeq \bar{\tau}_a^{(m+1)}(k)$, stop. The current solution, $\{u_a(k)\}$, $\{v_a(k)\}$, $\{x_a(k)\}$, $\{E^{rs}(k)\}$, is in a near instantaneous SDUO state; otherwise, set m=m+1 and go to step 1.

According to Powell and Sheffi (1982), the inner iteration procedure will converge. Since the convergence of outer iteration (diagonalization) is also robust (Florian and Spiess, 1982), we expect that our algorithm will converge to our desired instantaneous SDUO route choice solutions. In order to speed up convergence, an incremental assignment technique is suggested for finding a good starting solution before the diagonalization procedure. Since the subproblem can be decomposed by each artificial origin-destination pair, this problem is also a good candidate for solution with parallel computing techniques.

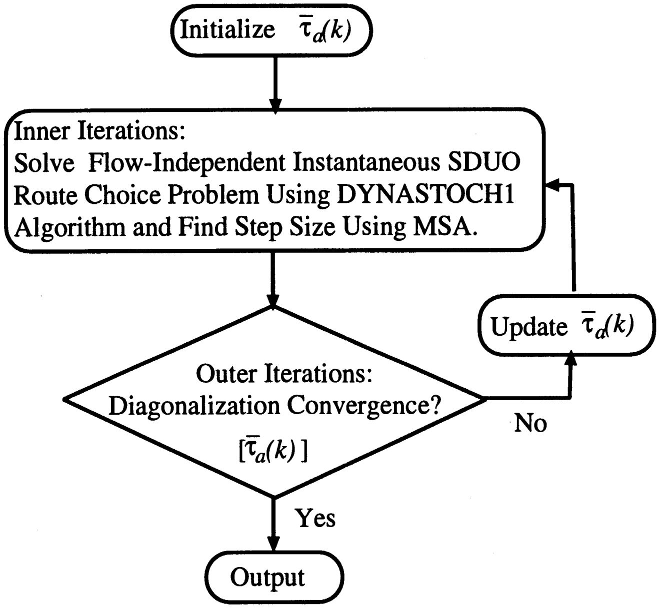



Figure 9.1: Flowchart of the Instantaneous Solution Algorithm

## 9.5 An Algorithm for the Ideal SDUO Route Choice Model

### 9.5.1 A Discrete Time Ideal Model

As before, we modify the estimated mean actual travel time on each link in the following way so that each estimated mean travel time is equal to a multiple of the time increment.

$$\tau_a(k) = i \quad \text{if} \quad i - 0.5 \leq \tau_a(k) < i + 0.5,$$

where $i$ is an integer and $0 \leq i \leq K$. We note that the above approximation is for flow propagation constraints only. It does not apply to the computation of route travel times.

The optimal control program presented in Chapter 8 can then be reformulated as a discrete time NLP as follows:

$$\min_{u,v,x,e,E,f,F;\pi} \quad \sum_{k=1}^{K}\sum_{a}\left\{\int_0^{u_a(k)} g_{1a}[x_a(k),\omega]d\omega + \int_0^{v_a(k)} g_{2a}[x_a(k),\omega]d\omega\right\} \tag{9.43}$$

s.t.

Relationship between state and control variables:

$$x_{ap}^{rs}(k+1) = x_{ap}^{rs}(k) + u_{ap}^{rs}(k) - v_{ap}^{rs}(k) \qquad \forall a, p, r, s; k = 1, \cdots, K; \tag{9.44}$$

$$F_p^{rs}(k+1) = F_p^{rs}(k) + f_p^{rs}(k) \qquad \forall r, s, p; k = 1, \cdots, K; \tag{9.45}$$

$$E_p^{rs}(k+1) = E_p^{rs}(k) + e_p^{rs}(k) \qquad \forall r, s, p; k = 1, \cdots, K; \tag{9.46}$$

Flow conservation constraints:

$$f_p^{rs}(k) = \sum_{a \in A(r)} \delta_{ap}^{rs} \, u_{ap}^{rs}(k) \qquad \forall p, r, s; k = 1, \cdots, K; \tag{9.47}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(k) - \sum_{a \in A(j)} u_{ap}^{rs}(k) = 0 \qquad \forall j, p, r, s; j \neq r, s; k = 1, \cdots, K; \tag{9.48}$$

$$e_p^{rs}(k) = \sum_{a \in B(s)} \delta_{ap}^{rs} \, v_{ap}^{rs}(k) \qquad \forall p, r, s \neq r; k = 1, \cdots, K; \tag{9.49}$$

Logit route flow constraints:

$$f_p^{rs}(k) = f^{rs}(k) \frac{\exp[-\theta \eta_p^{rs}(k)]}{\sum_q \exp[-\theta \eta_q^{rs}(k)]} \qquad \forall r, s, p; k = 1, \cdots, K; \tag{9.50}$$

Constraints for mean actual route travel times:

$$F_p^{rs}(k) = E_p^{rs}[k + \eta_p^{rs}(k)] \qquad \forall r, s, p; k = 1, \cdots, K; \tag{9.51}$$

Flow propagation constraints:

$$x_{ap}^{rs}(k) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[k + \tau_a(k)] - x_{bp}^{rs}(k)\} + \{E_p^{rs}[k + \tau_a(k)] - E_p^{rs}(k)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; k = 1, \cdots, K+1; \tag{9.52}$$

Definitional constraints:

$$\sum_p u_{ap}^{rs}(k) = u_a^{rs}(k), \qquad \sum_{rs} u_a^{rs}(k) = u_a(k), \qquad \forall a; k = 1, \cdots, K; \tag{9.53}$$

$$\sum_p v_{ap}^{rs}(k) = v_a^{rs}(k), \qquad \sum_{rs} v_a^{rs}(k) = v_a(k), \qquad \forall a; k = 1, \cdots, K; \tag{9.54}$$

$$\sum_p x_{ap}^{rs}(k) = x_a^{rs}(k), \qquad \sum_{rs} x_a^{rs}(k) = x_a(k), \qquad \forall a; k = 1, \cdots, K+1; \tag{9.55}$$

$$\sum_p e_p^{rs}(k) = e^{rs}(k), \quad \sum_p E_p^{rs}(k) = E^{rs}(k), \quad \forall r, s; k = 1, \cdots, K+1; \tag{9.56}$$

$$\sum_p f_p^{rs}(k) = f^{rs}(k), \quad \sum_p F_p^{rs}(k) = F^{rs}(k), \quad \forall r, s; k = 1, \cdots, K+1; \quad (9.57)$$

Nonnegativity conditions:

$$x_{ap}^{rs}(k+1) \geq 0, \quad u_{ap}^{rs}(k) \geq 0, \quad v_{ap}^{rs}(k) \geq 0, \qquad \forall a, p, r, s; k = 1, \cdots, K; \quad (9.58)$$

$$e_p^{rs}(k) \geq 0, \quad f_p^{rs}(k) \geq 0, \quad E_p^{rs}(k+1) \geq 0, \quad F_p^{rs}(k+1) \geq 0, \quad \forall p, r, s; k = 1, \cdots, K; \quad (9.59)$$

Boundary conditions:

$$E_p^{rs}(1) = 0, \qquad F_p^{rs}(1) = 0, \qquad \forall p, r, s; \qquad x_{ap}^{rs}(1) = 0, \qquad \forall a, p, r, s. \quad (9.60)$$

Since the objective function can be evaluated using link flow variables $u_a(k)$, $v_a(k)$ and $x_a(k)$, the objective function can be computed after solving the subproblem. Thus, traditional one-dimensional search methods such as the bisection method can be used.

The nonlinear route flow constraints (9.51) may not hold strictly as equalities because of cumulative round-off errors of link flow variables over routes after time discretization. In the inner iteration of the following algorithm, the logit-type assignment constraints (9.50) and nonlinear route flow constraints (9.51) are automatically satisfied by implementing the DYNASTOCH2 algorithm. Since the flow propagation constraints are temporarily fixed in each outer iteration (diagonalization), the remaining constraints are the logit-type assignment constraints (9.50), route flow constraints (9.51), flow conservation constraints and nonnegativity. Thus, the DYNASTOCH2 algorithm can be used to solve the subproblem within each outer iteration. In the outer iterations, the estimates of actual link travel times in the flow propagation constraints are updated iteratively.

### 9.5.2 Solution Algorithm

The flowchart of a heuristic algorithm for solving our ideal SDUO route choice model is shown in Figure 9.2; the algorithm is summarized as follows.

**Step 0: Initialization.**

Find an initial feasible solution $\{x_a^{(1)}(k)\}$, $\{u_a^{(1)}(k)\}$, $\{v_a^{(1)}(k)\}$, $\{E^{(1)}(k)\}$. Set the outer iteration counter m = 1.

**Step 1: Diagonalization. (Outer Iteration)**

Find a new estimate of actual link travel time $\tau_a^{(m)}(k)$ for the flow propagation constraints and solve the flow-dependent ideal SDUO route choice program as follows. Set the inner iteration counter n = 1.

[Step 1.1]: ***Update.*** Calculate the link travel time functions using the temporarily fixed link flow variables.

[Step 1.2]: *Direction Finding.* Implement the DYNASTOCH2 algorithm for the logit-based dynamic route flows based on the temporarily fixed actual link travel time and find an auxiliary set of link flow variables.

[Step 1.3]: *Line Search.* Find the optimal step size $\alpha^{(n)}$ that solves the one dimensional search problem using the bisection method.

[Step 1.4]: *Move.* Find a new solution by combining current solution $\{u_a(k)\}$, $\{v_a(k)\}$, $\{x_a(k)\}$, $\{E^{rs}(k)\}$ and previous solution $\{p_a(k)\}$, $\{q_a(k)\}$, $\{y_a(k)\}$, $\{\bar{E}^{rs}(k)\}$.

[Step 1.5]: *Convergence Test for Inner Iterations.* If n equals a prespecified number, go to step 2; otherwise, set $n = n + 1$ and go to step 1.1.

**Step 2: Convergence Test for Outer Iterations.**
If $\tau_a^{(m)}(k) \simeq \tau_a^{(m+1)}(k)$, stop. The current solution, $\{u_a(k)\}$, $\{v_a(k)\}$, $\{x_a(k)\}$, $\{E^{rs}(k)\}$, is in a near ideal SDUO state; otherwise, set $m = m+1$ and go to step 1.

In order to speed up convergence, the incremental assignment technique is also suggested for finding a good starting solution before the diagonalization procedure.

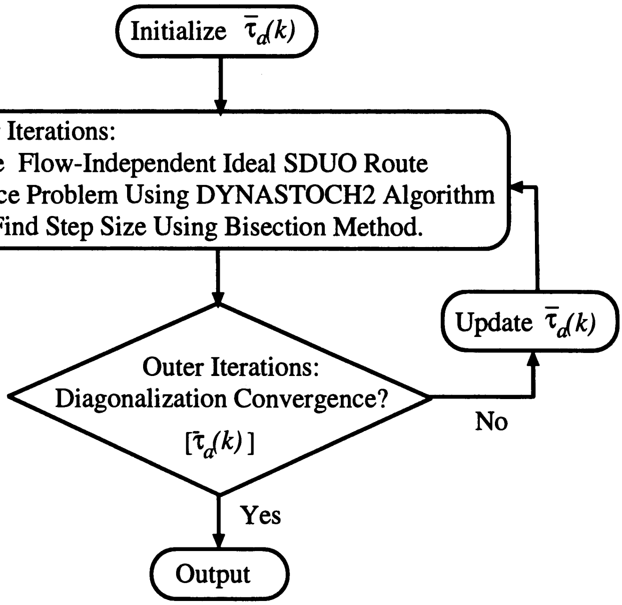



Figure 9.2: Flowchart of the Ideal Solution Algorithm

## 9.6 Numerical Examples

We illustrate the solutions of both SDUO route choice models with the 4-link, 4-node test network shown in Figure 9.3. Time period $[0, T]$ is subdivided into $K = 8$ small time intervals. The algorithms were coded in FORTRAN and solved on a IBM 3090-300J.

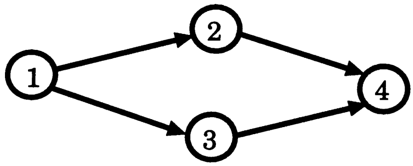



Figure 9.3: Test Network

The following link travel time functions were used in the computations:

$$c_a(k) = \tau_a(k) = g_{1a}(k) + g_{2a}(k)$$

$$g_{1a}(k) = \beta_{1a} + \beta_{2a}[u_a(k)]^2 + \beta_{3a}[x_a(k)]^2$$

$$g_{2a}(k) = \beta_{4a} + \beta_{5a}[v_a(k)]^2 + \beta_{6a}[x_a(k)]^2$$

where time interval $k = 1, 2, \cdots, 8$. The same trip table for both models is given in Table 9.1. The parameters for each link travel time function for the two models are given in Tables 9.2 and 9.3, respectively. The route choice dispersion parameter $\theta$ is given as 1 for both models. The optimal link flows and the corresponding optimal link travel times for the two models are given in Tables 9.4 and 9.6, respectively. The optimal route travel times for the two models are given in Tables 9.5 and 9.7, respectively. In this discrete time example, $x_a(k)$ represents vehicles on the link at the beginning of interval $k$; $u_a(k)$ and $v_a(k)$ represent inflow and exit flow during interval $k$.

Table 9.1: Required Flows from Origin 1 to Destination 4

| Time Interval $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Flow/interval | 40. | 0. | 0. | 0. | 0. | 0. | 0. | 0. |

In this example, note from Table 9.5 and 9.7 that the mean instantaneous travel times and mean future travel times on routes 1–2–4 and 1–3–4 are not equal during each interval. These travel times are a property of the logit-type route flows in each interval.

Table 9.2: Parameters of Link Travel Time Functions for Instantaneous Model

| link $a$ | $\beta_{1a}$ | $\beta_{2a}$ | $\beta_{3a}$ | $\beta_{4a}$ | $\beta_{5a}$ | $\beta_{6a}$ |
|---|---|---|---|---|---|---|
| 1—2 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 1—3 | 2. | 0.006 | 0. | 0. | 0.030 | 0.004 |
| 2—4 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 3—4 | 1. | 0.006 | 0. | 0. | 0.030 | 0.004 |

Table 9.3: Parameters of Link Travel Time Functions for Ideal Model

| link $a$ | $\beta_{1a}$ | $\beta_{2a}$ | $\beta_{3a}$ | $\beta_{4a}$ | $\beta_{5a}$ | $\beta_{6a}$ |
|---|---|---|---|---|---|---|
| 1—2 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 1—3 | 2. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 2—4 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 3—4 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |

## 9.7 Notes

Stochastic route choice models and solution algorithms have been studied extensively under the assumption of static traffic conditions. Dial (1971) presented the STOCH method to perform a logit-based, flow-independent stochastic traffic assignment. Daganzo and Sheffi (1977) presented a probit-based stochastic user-optimal route choice model. Subsequently, a Monte Carlo simulation approach to solving this problem were presented by Sheffi and Powell (1982). Fisk (1980) proposed a stochastic user-optimal (SUO) route choice model based on logit-type route flow method. The method of successive averages (MSA) was suggested to solve this model (Sheffi, 1985). Recently, two improved algorithms were proposed by Chen and Alfa (1991) for solving the logit-type SUO route choice model. Their algorithms use information in the objective function to speed up the convergence of the one-dimensional search.

In this chapter, we have suggested algorithms for solving two stochastic dynamic route choice problems. The solutions of the models result in dispersed dynamic route choice governed by a logit distribution incorporating both mean instantaneous route travel times and mean actual route travel times. Two DYNASTOCH algorithms were proposed to solve the subproblems of the two SDUO route choice models. The DYNASTOCH algorithms are implemented using link flow variables so that explicit route enumeration can be avoided. The algorithms are implemented over an expanded time-space network which enables our previously developed solution techniques for dynamic network models to be fully used. Future research also includes developing other efficient solution algorithms and calibration of the route choice dispersion parameters for

Table 9.4: Optimal Link Flows and Travel Times for Instantaneous Model

| Interval $k$ | Link $a$ | Vehicles $x_a(k+1)$ | Inflow $u_a(k)$ | Exit Flow $v_a(k)$ | Vehicles $x_a(k)$ | Travel Time $c_a(k)\&\tau_a(k)$ |
|---|---|---|---|---|---|---|
| 1 | 1 — 2 | 28.8 | 28.8 | 0.0 | 0.0 | 1.83 |
| 2 | 1 — 2 | 25.1 | 0.0 | 3.6 | 10.8 | 2.85 |
| 3 | 1 — 2 | 14.3 | 0.0 | 10.8 | 25.1 | 4.02 |
| 4 | 1 — 2 | 0.0 | 0.0 | 14.3 | 14.3 | 4.49 |
| 1 | 1 — 3 | 11.2 | 11.2 | 0.0 | 0.0 | 2.76 |
| 2 | 1 — 3 | 11.2 | 0.0 | 0.0 | 11.2 | 2.50 |
| 3 | 1 — 3 | 2.3 | 0.0 | 8.9 | 2.3 | 2.18 |
| 4 | 1 — 3 | 0.0 | 0.0 | 2.3 | 0.0 | 2.00 |
| 1 | 2 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 2 — 4 | 3.6 | 3.6 | 0.0 | 0.0 | 1.01 |
| 3 | 2 — 4 | 10.8 | 10.8 | 3.6 | 3.6 | 1.34 |
| 4 | 2 — 4 | 21.5 | 14.3 | 3.6 | 10.8 | 1.64 |
| 5 | 2 — 4 | 14.3 | 0.0 | 7.2 | 21.5 | 2.70 |
| 6 | 2 — 4 | 14.3 | 0.0 | 0.0 | 14.3 | 1.41 |
| 7 | 2 — 4 | 5.9 | 0.0 | 8.4 | 14.3 | 2.47 |
| 8 | 2 — 4 | 0.0 | 0.0 | 5.9 | 5.9 | 1.60 |
| 1 | 3 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 3 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 3 | 3 — 4 | 8.9 | 8.9 | 0.0 | 0.0 | 1.48 |
| 4 | 3 — 4 | 8.5 | 2.3 | 2.7 | 8.9 | 1.57 |
| 5 | 3 — 4 | 2.3 | 0.0 | 6.2 | 8.5 | 2.46 |
| 6 | 3 — 4 | 2.1 | 0.0 | 0.2 | 2.3 | 1.02 |
| 7 | 3 — 4 | 0.0 | 0.0 | 2.1 | 2.1 | 1.15 |

Table 9.5: Instantaneous Route Travel Times

| Interval | Mean Instantaneous Route Travel Times | |
|---|---|---|
| $k$ | Route 1-2-4 | Route 1-3-4 |
| 1 | 2.83 | 3.76 |
| 2 | 3.86 | 3.50 |
| 3 | 5.36 | 3.66 |
| 4 | 6.13 | 3.57 |
| 5 | 3.70 | 4.46 |
| 6 | 2.41 | 3.02 |
| 7 | 3.47 | 3.15 |
| 8 | 2.60 | 3.0 |

Table 9.6: Optimal Link Flows and Travel Times for Ideal Model

| Interval $k$ | Link $a$ | Vehicles $x_a(k+1)$ | Inflow $u_a(k)$ | Exit Flow $v_a(k)$ | Vehicles $x_a(k)$ | Travel Time $c_a(k)\&\tau_a(k)$ |
|---|---|---|---|---|---|---|
| 1 | 1 — 2 | 24.5 | 24.5 | 0.0 | 0.0 | 1.60 |
| 2 | 1 — 2 | 21.5 | 0.0 | 3.0 | 24.5 | 2.33 |
| 3 | 1 — 2 | 5.5 | 0.0 | 16.0 | 21.5 | 5.77 |
| 4 | 1 — 2 | 0.0 | 0.0 | 5.5 | 5.5 | 1.52 |
| 1 | 1 — 3 | 15.5 | 15.5 | 0.0 | 0.0 | 2.24 |
| 2 | 1 — 3 | 15.5 | 0.0 | 0.0 | 15.5 | 2.48 |
| 3 | 1 — 3 | 0.0 | 0.0 | 15.5 | 15.5 | 6.08 |
| 4 | 1 — 3 | 0.0 | 0.0 | 0.0 | 0.0 | 2.00 |
| 1 | 2 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 2 — 4 | 3.0 | 3.0 | 0.0 | 0.0 | 1.01 |
| 3 | 2 — 4 | 16.0 | 16.0 | 3.0 | 3.0 | 1.41 |
| 4 | 2 — 4 | 18.6 | 5.5 | 2.9 | 16.0 | 1.67 |
| 5 | 2 — 4 | 5.5 | 0.0 | 13.1 | 18.6 | 4.26 |
| 6 | 2 — 4 | 5.5 | 0.0 | 0.0 | 5.5 | 1.06 |
| 7 | 2 — 4 | 0.0 | 0.0 | 5.5 | 5.5 | 1.52 |
| 1 | 3 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 2 | 3 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.00 |
| 3 | 3 — 4 | 15.5 | 15.5 | 0.0 | 0.0 | 1.24 |
| 4 | 3 — 4 | 12.6 | 0.0 | 2.9 | 15.5 | 1.60 |
| 5 | 3 — 4 | 0.0 | 0.0 | 12.6 | 12.6 | 3.70 |

Table 9.7: Ideal Route Travel Times

| Interval | Mean Actual Route Travel Times | |
|---|---|---|
| $k$ | Route 1-2-4 | Route 1-3-4 |
| 1 | 3.01 | 3.48 |
| 2 | 6.59 | 4.08 |

logit SDUO route choice models. Using realistic link travel time functions, we will implement our models and solution algorithms on larger networks.

We are investigating other SDUO route choice models based on different types of distributions of travel time perception errors. A straightforward extension is the dynamic generalization of probit route choice models. However, more realistic distributions of travel time perception errors than the normal distribution should be considered, such as the Gamma distribution. This issue provides a challenge to both mathematical modeling and estimation of parameters of the distribution of perception errors from data.

# Chapter 10

# Combined Departure Time/ Route Choice Models

A dynamic route guidance system seeks to improve the utilization of transportation network capacity and reduce travel times, congestion and the effect of incidents. Provided with early detection of incidents and congestion, users of the system will be able to choose alternative routes, if there is excess capacity in the network, or shift their departure times to avoid congestion when no road capacity is available.

Journey-to-work travelers have especially important requirements for avoiding congested routes in order to arrive at work on time. Each departure time choice is based on minimal origin-destination travel times at each possible departure time. Of course, any change in departure times will alter the traffic flow patterns in the network so that route and departure time decisions of other travelers will be affected.

The choice of departure time has been addressed by several researchers, including Abkowitz (1981) and Hendrickson and Plank (1984), who developed work trip scheduling models. De Palma et al (1983) and Ben-Akiva et al (1984) modeled departure time choice over a simple network with one bottleneck using the general continuous logit model. Mahmassani and Herman (1984) used a traffic flow model to derive the equilibrium joint departure time and route choice pattern over a parallel route network. Mahmassani and Chang (1987) further developed the concept of equilibrium departure time choice and presented the boundedly-rational user-equilibrium concept under which all drivers in the system are satisfied with their current travel choices, and thus feel no need to improve their outcome by changing decisions. More recently, several departure time choice models have been proposed by various researchers using different approaches on dynamic traffic networks. Janson (1993) formulated a dynamic user-equilibrium route choice model in which O-D flows have variable departure times and scheduled arrival times. Friesz et al (1993) presented a joint departure time and route choice model using the variational inequality approach. Ghali and Smith (1993) also considered this problem using microscopic representation of vehicle streams.

In this chapter, we present a dynamic, user-optimal departure time and route choice model for a general network with multiple origin-destination pairs. We model this choice problem by specifying that a given number of travelers are ready for departure between each origin-destination pair at time 0. However, their departure times may be delayed to reduce their overall travel costs. This model extends our initial DUO route choice model in one important respect: both departure time and route over a road network must be chosen. Our model is formulated as a bilevel optimal control problem. The lower-level model represents the DUO departure time choice problem, and the upper-level model represents the DUO route choice problem.

Additional network constraints are presented in Section 10.1 and the bilevel model is formulated in Section 10.2. In Section 10.3, the equivalence of its optimality conditions with the desired DUO departure time/route choice conditions is demonstrated. The properties of the model are also discussed. In Section 10.4, we suggest a heuristic algorithm for solving the bilevel program and then give a numerical example to illustrate that total travel time can be decreased by choosing appropriate departure times.

## 10.1 Additional Network Constraints

We consider the following joint departure time and route choice situation: a given number of vehicles are scheduled to depart from each origin $r$ to each destination $s$ at an initial time 0. Denote the *cumulative* number of departing vehicles from origin $r$ to destination $s$ from time 0 to $t$ as the state variable $F^{rs}(t)$. In this problem, the total number of departing vehicles $F^{rs}(T)$ for each O-D pair $(r, s)$ is assumed to be given. Also, $F_p^{rs}(t)$ denotes the cumulative number of departing vehicles from origin $r$ toward destination $s$ along route $p$ by time $t$.

In addition, denote the instantaneous departure rate from origin node $r$ toward destination node $s$ at time $t$ as $f^{rs}(t)$, which is a function of time; $f_p^{rs}(t)$ denotes the departure rate on route $p$ and $f_p^{rs}(t)$ and $f^{rs}(t)$ are control variables to be determined according to the actual travel time between the origin and the destination. Then, we have an additional state equation for each origin $r$

$$\frac{dF_p^{rs}(t)}{dt} = f_p^{rs}(t) \qquad \forall p, r \neq s, s. \tag{10.1}$$

Also, at initial time $t = 0$,

$$F_p^{rs}(0) = 0, \qquad \forall p, r, s. \tag{10.2}$$

Assume that there are $P$ routes from origin $r$ to destination $s$ (these can be generated as needed). Denote the indicator parameters $\delta_{ap}^{rs}$ as

$$\delta_{ap}^{rs} = \begin{cases} 1 & \text{if link } a \text{ is on route } p \text{ between O-D pair } (r,s) \\ 0 & \text{otherwise.} \end{cases}$$

Flow conservation at origin node $r$ relates the departure rates ($f^{rs}(t)$ and $f_p^{rs}(t)$) to the flow entering each link emanating from the origin. These flow conservation equations for origin $r$ can be expressed as

$$f_p^{rs}(t) = \sum_{a \in A(r)} \delta_{ap}^{rs} u_{ap}^{rs}(t) \qquad \forall p, r, s; r \neq s; \tag{10.3}$$

$$\sum_p f_p^{rs}(t) = f^{rs}(t) \qquad \forall r, s; r \neq s. \tag{10.4}$$

Denote the *cumulative* number of vehicles arriving at destination $s$ from origin $r$ by time $t$ as the state variable $E^{rs}(t)$; $E_p^{rs}(t)$ denotes the cumulative number of vehicles arriving at destination $s$ from origin $r$ along route $p$ by time $t$. Denote the instantaneous flows *arriving at* destination node $s$ from origin node $r$ at time $t$ as $e^{rs}(t)$, which is also a control variable. The control variable $e_p^{rs}(t)$ denotes the arrival rate on route $p$. Thus, we have an additional state equation for each destination $s$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r, s \neq r. \tag{10.5}$$

At the initial time $t = 0$,

$$E_p^{rs}(0) = 0 \qquad \forall p, r, s. \tag{10.6}$$

These variables must be nonnegative at all times:

$$E_p^{rs}(t) \geq 0, \quad F_p^{rs}(t) \geq 0, \quad e_p^{rs}(t) \geq 0, \quad f_p^{rs}(t) \geq 0, \quad \forall p, r, s. \tag{10.7}$$

Flow conservation at destination node $s$ relates the arriving flow ($e^{rs}(t)$ and $e_p^{rs}(t)$) to the flow exiting each link leading to destination $s$ at time $t$. Thus, the flow conservation equations for destination $s$ can be expressed as

$$e_p^{rs}(t) = \sum_{a \in B(s)} \delta_{ap}^{rs} v_{ap}^{rs}(t) \qquad \forall p, r, s; s \neq r; \tag{10.8}$$

$$\sum_p e_p^{rs}(t) = e^{rs}(t) \qquad \forall r, s; s \neq r. \tag{10.9}$$

## 10.2 Formulation of the Bilevel Program

A number of vehicles are ready to depart at the initial time 0, but these drivers may prefer to delay their departure times in order to reduce their driving time. Drivers are assumed to make their departure time choices so as to minimize their individual disutility functions defined on travel time and pre-trip delay. The criteria for choosing each departure time consist of:

1. the waiting time before departure;
2. the actual travel time between the origin and destination;
3. a bonus for early arrival or a penalty for late arrival.

Of course, the change of departure flow rates will affect the traffic on the network so that the travel times for other travelers could change.

In reality, drivers' choices of departure time and route are interrelated decisions. Given a desired arrival time, say at the workplace, choice of departure time depends on the driver's estimate of en route travel time. Likewise, choice of route depends on the travel times of alternative routes, which also may vary by time of day. In the formulation presented here, these choices are represented as a bilevel optimal control problem, which is equivalent to a dynamic leader-follower game (Cruz, 1978).

The dynamic *route choice* problem is formulated as a single optimal control problem. In the equivalent dynamic game, this formulation corresponds to a single *controller* allocating fixed departure flows at each time $t$ to user-optimal routes, given the departure frequencies. We define this controller to be the leader of the game.

For each origin-destination (O-D) pair, a *departure time coordinator* determines the departing flows at each time $t$. These departure coordinators are defined as the followers of the game, and are represented by $n(n-1)$ O-D-specific optimal control problems to which the user-optimal travel time at time $t$ is exogenous. Since these problems are independent by O-D pair, $n(n-1)$ separate problems can be used to represent all O-D pairs. See Figure 10.1.

| | |
|---|---|
| Upper Level: Route Choice Controller: | Allocate given departure flows to routes such that all used routes have equal travel times. |

where the departure flows are given by:

| | |
|---|---|
| Lower Level: Departure Time Coordinator for O-D Pair $rs$: | Allocate departure flows from $r$ to $s$, given O-D travel times, to minimize a weighted sum of waiting and travel times and arrival penalty. |

Figure 10.1: Bilevel Choice Program Formulation

Our formulation assumes that the route choice controller, who is responsible for allocating all O-D flows to routes, knows the objective of each departure coordinator. Through this knowledge, the route choice coordinator is able to achieve a lower value of his/her objective function than if the departure time objectives were unknown. In contrast, the departure time coordinators know only the O-D travel times at time $t$ provided to them. Knowledge of the route choice objective is not needed for their coordination task.

It is worth noting that the opposite formulation can also be examined – departure time choice at the upper level, represented by a single controller, and route choice at the lower level. However, our model which chooses equilibrium routes subject to the requirement that departure times be optimal for all travelers is equally plausible. Furthermore, our bilevel optimal control problem is much more tractable.

Next, we formulate the lower level problem as $n(n-1)$ optimal control problems representing the departure time coordinators. Then, the upper level optimal control problem representing the route choice controller is defined in Section 10.2.2. Finally, the bilevel problem is presented. We demonstrate in Section 10.3 that the solution of our model satisfies the desired departure time and route choice conditions.

## 10.2.1 Lower Level Problem: Departure Time Choice

We first consider the lower-level problem of departure time choice. A disutility function $\mathcal{U}^{rs}(t)$ based on departure times is defined for travelers departing from origin $r$ to destination $s$ at time $t$. This disutility function represents a weighted sum of:

1. waiting time at the origin node;
2. driving time during the trip;
3. a bonus for early arrival or a penalty for late arrival.

Denote $\pi^{rs}(t)$ as the minimal travel time experienced by vehicles departing from origin $r$ to destination $s$ at time $t$. That is, $\pi^{rs}(t)$ is a functional of all link flow variables at time $t$, i.e., $\pi^{rs}(t) = \pi^{rs}[u(\omega), v(\omega), x(\omega), \omega]$ where $\omega \geq t$. This functional is neither a state variable nor a control variable, and it is not fixed; moreover, it is not available in closed form. Nevertheless, it can be evaluated when $u(\omega)$, $v(\omega)$ and $x(\omega)$ are temporarily fixed.

We define one unit of disutility to equal one unit of in-vehicle driving time, and one unit of waiting time prior to departure to be equivalent to $\alpha$ units of disutility ($\alpha \leq 1$). Since all travelers are able to depart at time 0, $\alpha t$ is the disutility due to waiting. Sometimes, $\alpha$ can become negative so that waiting time at the origin is a utility instead of disutility. In other words, drivers prefer to stay at home and regard waiting at home as a utility. Furthermore, we assume there is a desired arrival time interval $[t^*_{rs} - \Delta_{rs}, t^*_{rs} + \Delta_{rs}]$ for travelers at each destination $s$, where $t^*_{rs}$ is the center of the required arrival time interval (e.g. work start time) associated with travelers departing from origin $r$ toward destination $s$. $\Delta_{rs}$ represents the arrival time flexibility at destination $s$ for travelers departing from origin $r$ toward destination $s$.

We also define the disutility for early or late arrival as follows

$\mathcal{V}^{rs}[t, \pi^{rs}(t); t^*_{rs}] =$

$$\begin{cases} \gamma_1[t + \pi^{rs}(t) - t^*_{rs} + \Delta^*_{rs}]^2 & \text{if } t + \pi^{rs}(t) < t^*_{rs} - \Delta^*_{rs} \quad \text{(early arrival)} \\ 0 & \text{if } |t + \pi^{rs}(t) - t^*_{rs}| \leq \Delta^*_{rs} \\ \gamma_2[t + \pi^{rs}(t) - t^*_{rs} - \Delta^*_{rs}]^2 & \text{if } t + \pi^{rs}(t) > t^*_{rs} + \Delta^*_{rs} \quad \text{(late arrival)} \end{cases}$$

where $t$ is the departure time of travelers and $\gamma_1, \gamma_2$ are parameters ($\gamma_1 \leq 0, \gamma_2 \gg \alpha$). This arrival time disutility function is shown in Figure 10.2. Thus, the disutility function for the joint departure time and route choice problem is constructed as

$$\mathcal{U}^{rs}(t) = \alpha t + \pi^{rs}(t) + \mathcal{V}^{rs}[t, \pi^{rs}(t); t^*_{rs}] \qquad \forall r, s, \tag{10.10}$$

where $t$ is the departure time of travelers. In some situations where arrival time is more important, the impact of waiting time on the disutility function is not significant. Thus, the term $\alpha t$ can be dropped for these situations. On the other hand, when the arrival time is not important, the disutility term $\mathcal{V}^{rs}(\cdot)$ (due to early or late arrival) can be dropped. However, the disutility $\alpha t$ becomes important to determine the departure time and has to be kept in the disutility function $\mathcal{U}^{rs}(t)$.

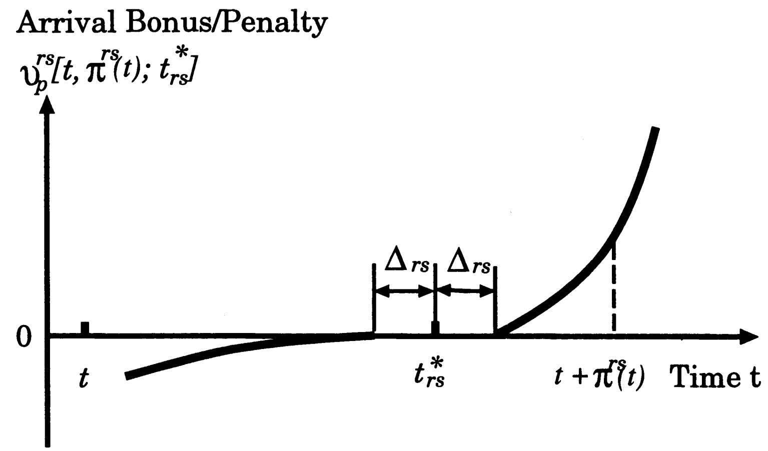



Figure 10.2: Bonus/Penalty for Early/Late Arrival

The dynamic user-optimal departure time choice conditions can then be defined as:

$$\mathcal{U}^{rs}(t) = \mathcal{U}^{rs}_{min} \qquad \text{if } f^{rs}(t) > 0 \qquad \forall r, s; \tag{10.11}$$

$$\mathcal{U}^{rs}(t) \geq \mathcal{U}^{rs}_{min} \qquad \text{if } f^{rs}(t) = 0 \qquad \forall r, s; \tag{10.12}$$

where $\mathcal{U}^{rs}_{min}$ is the minimal $rs$ disutility. Additional boundary conditions are the following.

$$F^{rs}(0) = 0 \qquad \forall r, s. \tag{10.13}$$

Also, by definition of $f^{rs}(t)$ (the O-D departure rate) and $F^{rs}(t)$ (the cumulative departure rate), it follows that

$$\int_0^T f^{rs}(t)dt = F^{rs}(T) \qquad \text{is given} \qquad \forall r, s; \tag{10.14}$$

or

$$\frac{dF^{rs}(t)}{dt} = f^{rs}(t) \qquad \forall r, s. \tag{10.15}$$

In Section 10.2.3, we state the Lower Level Problem whose solution yields conditions (10.11)-(10.12), given the functionals $\pi^{rs}(t)$.

### 10.2.2 Upper Level Problem: Route Choice

Next we discuss the upper-level problem of route choice. The dynamic user-optimal route choice problem is to find the dynamic trajectories of link states and inflow and exit flow control variables, given the network, the link travel time functions and the time-dependent O-D departure rate requirements. The O-D departure rates are specified by the lower level problem and are therefore exogenous to the upper level problem. In this joint DUO departure time/route choice problem, the minimal instantaneous O-D travel time is not a good basis for adjustment of departure times. Thus, we use the ideal DUO route choice model developed in Chapter 7 as our upper-level problem. The formulation for the route choice upper-level problem is summarized in the next bilevel model.

### 10.2.3 Bilevel Program Formulation

Using optimal control theory, a bilevel optimization program of the dynamic, user-optimal departure time and route choice problem is formulated as follows.

**Upper Level: Ideal DUO Route Choice**

$$\min_{u,v,x,e,E,f_p^{rs},F_p^{rs};\pi} \quad \int_0^T \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t), \omega]d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t), \omega]d\omega \right\} dt \tag{10.16}$$

s.t.

Relationships between state and control variables:

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \tag{10.17}$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall r, s, p; \tag{10.18}$$

$$\frac{dF_p^{rs}(t)}{dt} = f_p^{rs}(t) \qquad \forall r, s, p; \tag{10.19}$$

Flow conservation constraints:

$$f_p^{rs}(t) = \sum_{a \in A(r)} \delta_{ap}^{rs} u_{ap}^{rs}(t) \qquad \forall p, r, s; \tag{10.20}$$

$$e_p^{rs}(t) = \sum_{a \in B(s)} \delta_{ap}^{rs} v_{ap}^{rs}(t) \qquad \forall p, r, s; \tag{10.21}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \tag{10.22}$$

Constraints equilibrating *actual* route travel times:

$$F^{rs}(t) = E^{rs}[t + \pi^{rs}(t)] \qquad \forall r, s; \tag{10.23}$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[t + \tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall r, s, p, j; a \in B(j); j \neq r; \tag{10.24}$$

Definitional constraints:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \tag{10.25}$$

$$\sum_{p} x_{ap}^{rs}(t) = x_a^{rs}(t), \quad \sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \quad \sum_{rs} x_a^{rs}(t) = x_a(t), \quad \forall a; \tag{10.26}$$

$$\sum_{p} E_p^{rs}(t) = E^{rs}(t), \qquad \sum_{p} F_p^{rs}(t) = F^{rs}(t), \qquad \forall r, s; \tag{10.27}$$

$$\sum_{p} f_p^{rs}(t) = f^{rs}(t), \qquad \sum_{p} e_p^{rs}(t) = e^{rs}(t), \qquad \forall r, s; \tag{10.28}$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \qquad u_{ap}^{rs}(t) \geq 0, \qquad v_{ap}^{rs}(t) \geq 0 \qquad \forall a, p, r, s; \tag{10.29}$$

$$e_p^{rs}(t) \geq 0, \quad f_p^{rs}(t) \geq 0, \quad E_p^{rs}(t) \geq 0, \quad F_p^{rs}(t) \geq 0 \qquad \forall p, r, s; \tag{10.30}$$

Boundary conditions:

$$E_p^{rs}(0) = 0, \qquad F_p^{rs}(0) = 0 \qquad \forall p, r, s; \qquad x_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s. \tag{10.31}$$

where $f^{rs}(t)$ and $F^{rs}(t)$ solve

**Lower Level: DUO Departure Time Choice**

$$\min_{f^{rs},F^{rs}} \int_0^T \sum_{rs} \left\{ \int_0^{f^{rs}(t)} \{\alpha t + \pi^{rs}(t) + \mathcal{V}^{rs}[t, \pi^{rs}(t); t^*_{rs}]\} \, d\omega \right\} dt \quad (10.32)$$

s.t.

$$\frac{dF^{rs}(t)}{dt} = f^{rs}(t) \qquad \forall r; s; \quad (10.33)$$

$$F^{rs}(0) = 0 \qquad F^{rs}(T) \quad \text{given} \qquad \forall r, s; \quad (10.34)$$

$$f^{rs}(t) \geq 0, \qquad F^{rs}(t) \geq 0 \qquad \forall r, s. \quad (10.35)$$

In lower-level model (10.32)-(10.35), we only have state equations, boundary conditions and nonnegativity conditions. The control variables are $f^{rs}(t)$, and the state variables are $F^{rs}(t)$, which represent total departures over *all routes.*

In upper-level model (10.16)-(10.31), *route-specific* departure variables $f_p^{rs}(t)$ and $F_p^{rs}(t)$ must be determined. The upper-level objective function is similar to the objective function of the well-known static user-optimal (UO) model. We note that other objective functions can also be used since constraints (10.23) enforce the ideal DUO route choice. The first three constraints (10.17)-(10.19) are state equations for each link $a$ and for cumulative effects at origins and destinations. Equations (10.20)-(10.22) are flow conservation constraints at each node including origins and destinations. Equation (10.23) is constraint which equilibrates flows based on *actual* route travel times. Other constraints include flow propagation constraints, definitional constraints, nonnegativity, and boundary conditions.

In summary, in the upper-level program the control variables are $u_{ap}^{rs}(t)$, $v_{ap}^{rs}(t)$, $e_p^{rs}(t)$, and $f_p^{rs}(t)$; the state variables are $x_{ap}^{rs}(t)$, $E_p^{rs}(t)$, and $F_p^{rs}(t)$; the functionals are $\pi^{rs}(t)$, which must be determined in a diagonalization fashion as discussed in Chapter 7. Note that the upper-level problem alone is an ideal DUO route choice model for the case of fixed departure times, because the route flow constraints guarantee the ideal DUO route choice conditions without regard to whether the O-D flows $f^{rs}(t)$ are fixed exogenously or decided by a lower-level departure time choice problem.

# 10.3 Optimality Conditions and Equivalence Analysis

## 10.3.1 Optimality Conditions

We first derive the optimality conditions for the lower-level model (10.32)-(10.35). The Hamiltonian for the lower-level model is

$$\mathcal{H}_1 = \sum_{rs} \int_0^{f^{rs}(t)} [\alpha t + \pi^{rs}(t) + \mathcal{V}^{rs}[t, \pi^{rs}(t); t^*_{rs}]] \, d\omega + \sum_{rs} \mu^{rs}(t) f^{rs}(t)$$

where $\mu^{rs}(t)$ is the Lagrange multiplier associated with each origin-destination pair's state equation (10.33). For each $rs$, the first order necessary conditions of the lower-level program (10.32)-(10.35) include

$$\frac{\partial \mathcal{H}_1}{\partial f^{rs}(t)} = \alpha t + \pi^{rs}(t) + \mathcal{V}^{rs}[t, \pi^{rs}(t); t^*_{rs}] + \mu^{rs}(t) \geq 0, \qquad \forall r, s; \qquad (10.36)$$

$$f^{rs}(t) \frac{\partial \mathcal{H}_1}{\partial f^{rs}(t)} = 0 \qquad \forall r, s; \qquad (10.37)$$

$$\frac{d\mu^{rs}(t)}{dt} = -\frac{\partial \mathcal{H}_1}{\partial F^{rs}(t)} = 0 \qquad \forall r, s; \qquad (10.38)$$

$$f^{rs}(t) \geq 0, \qquad \forall r, s. \qquad (10.39)$$

An alternative representation of bilevel program (10.16)-(10.35) can be given by converting it into a standard optimization model. As suggested by Cruz (1978) and Bard (1984), this can be achieved by appending the optimality conditions (10.36)-(10.39) of the departure time choice model (lower problem) to the constraint set of the route choice model (upper problem). The solution to the resulting single level model would also be a solution to the original bilevel departure time/route choice problem. The *equivalent single level program* is reformulated as:

Min (10.16)

s.t. (10.17)-(10.31) (Upper level model constraints)

(10.33)-(10.35) (Lower level model constraints)

(10.36)-(10.39) (Lower level model optimality conditions)

We do this only to analyze the optimality conditions of the bilevel program. From an algorithmic point of view, the model would still be solved as a bilevel program.

We can construct the Hamiltonian for this single level program and derive the corresponding first order necessary conditions for each link $a$, origin node $r$ and destination node $s$. For our analysis of optimality, we need only one part of the constraints of the single level program: constraints (10.36)-(10.39) and (10.23). Other constraints and first order necessary conditions are not used for our analysis of the DUO state. We now show that constraints (10.36)-(10.39) for this single level program are identical to the DUO departure time choice. Also, we note that constraints (10.23) for this single level program directly guarantee the ideal DUO route choice conditions.

## 10.3.2 DUO Equivalence Analysis

Constraints (10.36)-(10.39) for the equivalent single level program are also the optimality conditions of the lower level departure time choice model. In the following, we prove that these conditions are equivalent to DUO departure time choice conditions so that the equivalent single level program generates departure flows satisfying the DUO departure time choice conditions.

The costate equation (10.38) can be integrated as

$$\mu^{rs}(t) = A \qquad \forall r, s; \tag{10.40}$$

where $A$ is an integral constant. This equation applies to any time $t \in [0, T]$. Thus, from equations (10.36)-(10.39), we obtain the following equations.

$$f^{rs}(t)\left\{\alpha t + \pi^{rs}(t) + \mathcal{V}^{rs}[t, \pi^{rs}(t); t^*_{rs}] + A\right\} = 0, \qquad \forall r, s; \tag{10.41}$$

$$\alpha t + \pi^{rs}(t) + \mathcal{V}^{rs}[t, \pi^{rs}(t); t^*_{rs}] \geq -A, \qquad \forall r, s; \tag{10.42}$$

$$f^{rs}(t) \geq 0, \qquad \forall r, s. \tag{10.43}$$

Note that the left hand side of (10.42) consists of: 1) disutility due to the waiting time; 2) minimal actual O-D travel time; and 3) a bonus for early arrival or a penalty for late arrival. The above conditions (10.41)-(10.43) hold for each O-D pair $(r, s)$ in the network. For any O-D pair $(r, s)$, if there are vehicles departing at time $t$, then $f^{rs}(t)$ will be positive, so the quantities in braces in equation (10.41) will be zero, i.e., equation (10.42) will hold as an equality. (Since the total disutility in braces of equation (10.41) is positive by definition, A is clearly negative.) Thus, travelers departing at time $t$ have disutility equal to $-A$. Inequality (10.42) states that at optimality, this $rs$ disutility is less than or equal to the disutility for departures at any time $t$. Therefore, the disutility for departures at time $t$ equals the minimal disutility for origin-destination $(r, s)$ at any time $t$. For any time $t$, if there are no vehicles departing origin $r$, then the departure rate $f^{rs}(t)$ equals zero, so that (10.42) may hold as a strict inequality. Thus, the disutility $\alpha t + \pi^{rs}(t) + \mathcal{V}^{rs}[t, \pi^{rs}(t); t^*_{rs}]$ at any time $t$ will not be less than the minimal disutility $|A|$. The above interpretation implies that the optimality conditions of the lower-level program are consistent with the DUO departure time choices.

Since the optimality conditions for the lower level departure time choice model are one part of the constraints for the equivalent single level program, the above analysis results also apply to this problem. Thus, the equivalent single level program (and thus the bilevel optimal control program) will generate O-D departure flows which satisfy the DUO departure time choice conditions.

Note that constraints (10.23) still apply to the above equivalent single level program. These constraints guarantee that the bilevel optimal control program generates traffic flows satisfying the ideal DUO route choice conditions, given any O-D departure flows determined by the revised constraint set of the single level program.

We have shown that the set of departure flows and link flows that solves the equivalent single level program satisfy both the DUO departure time choice conditions and the ideal DUO route choice conditions. Therefore, the solution to the original bilevel departure time/route choice program satisfies both the DUO departure time choice conditions and the ideal DUO route choice conditions.

## 10.4 Solution Algorithm and An Example

As noted by many researchers, the bilevel nonlinear program is generally nonconvex; therefore, it is very difficult to find its global optimal solution. In the following, we present a heuristic solution algorithm which solves a discretized version of our bilevel optimal control problem. A numerical example is also presented to illustrate that total disutility of travel can be decreased by choosing appropriate departure times.

To convert our optimal control problem into a nonlinear programming problem (NLP), time period $[0, T]$ is subdivided into $K$ small time intervals. Time interval $k$ is denoted as $k = [k, k+1]$. (These time intervals are not necessarily equal.) Then, the optimal control program can be reformulated as a discrete time NLP.

### 10.4.1 Solution Algorithm

In the resulting discrete time problem, $x_a(k)$ represents vehicles on the link at the beginning of interval $k$; $u_a(k)$ and $v_a(k)$ represent inflow and exit flow during interval $k$. Let $\tau_a(k)$ denote the travel time for vehicles entering link $a$ at the beginning of interval $k = [k, k+1]$, and let $\pi^{rs}(k)$ be the average minimal $r-s$ travel time for vehicles departing origin $r$ during interval $k$. Let $f^{rs}(k)$ denote the O-D departure flow during interval $k$.

We use the diagonalization technique to solve our bilevel NLP. In this procedure, the actual travel times over each link $a$, $\tau_a(k)$, are temporarily fixed and are updated iteratively. At each iteration, since each $\tau_a(k)$ is temporarily fixed, the minimal O-D travel time functional $\pi^{rs}(k)$ can be computed and is also temporarily fixed.

By discretizing the time period, the upper level route choice model becomes a discrete time NLP, and the lower level departure time choice model becomes a discrete time linear programming problem (LP). In our heuristic algorithm, at each iteration, the lower level departure time choice model is solved first to obtain the O-D departure flows $f^{rs}(k)$. The upper level route choice model is solved by the Frank-Wolfe technique (Frank and Wolfe, 1956) with penalty functions for the nonlinear constraints (10.23). We note that in the route choice problem, since constraints (10.23) are put in the objective function as penalty terms, only flow conservation and flow propagation equations remain so that the Frank-Wolfe technique can be used to solve the modified program.

After solving the route choice problem for fixed $\tau_a(k)$, the link travel times corresponding to the solution obtained for $x_a(k)$, $u_a(k)$ and $v_a(k)$ are compared to the functions $\tau_a(k)$. If the link travel times corresponding to the solution are different from $\tau_a(k)$ and the penalty term does not approach zero, the $\tau_a(k)$ are reset to these travel times and the process is repeated. Given the robust nature of the diagonalization technique, we expect that the solution will converge to the DUO solution. The flowchart of the solution procedure is shown in Figure 10.3.

The LP *departure time* subproblem is rather simple to solve. The main difficulty of solving the bilevel program is solving the NLP route choice subproblem efficiently.

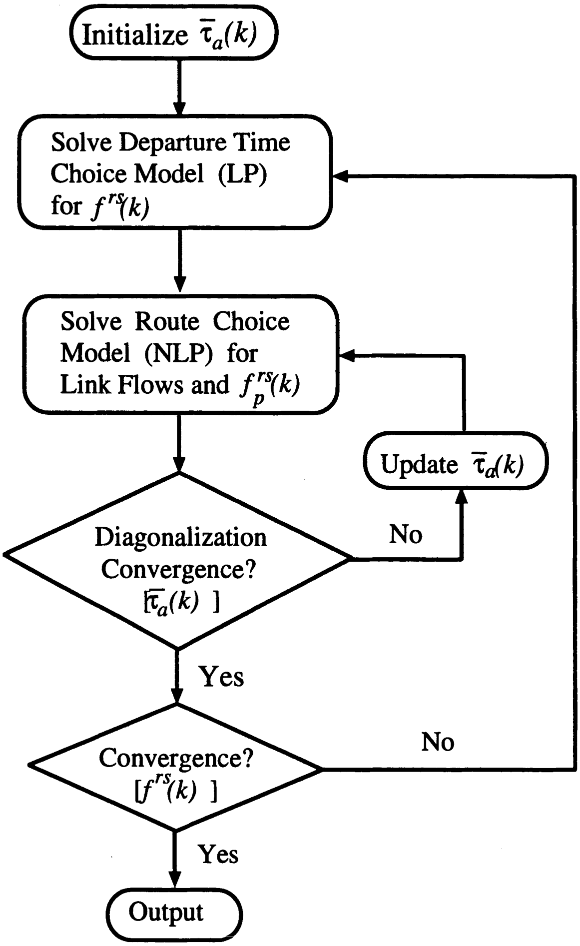



Figure 10.3: Flowchart of the Solution Algorithm

## 10.4.2 Numerical Example

We illustrate the solution of our bilevel choice model with the 4-link, 4-node test network shown in Figure 10.4. The assignment time period $[0, T]$ is subdivided into $K = 8$ small time intervals. The algorithm was coded in FORTRAN and solved on a IBM 3090-300J. As proposed in Chapter 5, the following link travel time functions were used in the computations:

$$\tau_a(k) = c_a(k) = g_{1a}(k) + g_{2a}(k)$$
$$g_{1a}(k) = \beta_{1a} + \beta_{2a}[u_a(k)]^2 + \beta_{3a}[x_a(k)]^2$$
$$g_{2a}(k) = \beta_{4a} + \beta_{5a}[v_a(k)]^2 + \beta_{6a}[x_a(k)]^2$$

where the time interval $k = 1, 2, \cdots, 8$. The same function is used to represent both the instantaneous and actual link travel time functions in order to simplify the presentation. The parameters for each link travel time function are given in Table 10.1.

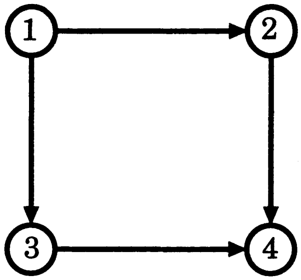



Figure 10.4: Test Network

In this example, we compare the travel times and disutilities from origin 1 to destination 4 under two different specified departure flow patterns. Since the network is symmetric, constraints (10.23), which equilibrate the actual route travel times, are satisfied automatically. Thus, it is not necessary to use penalty functions to enforce these constraints in this example, although with a general network it would be. The initial and improved departure flow patterns are given in Table 10.2; the user optimal link flows and the corresponding link travel times for the initial departure flow pattern are shown in Table 10.3. The optimal link flows and optimal link travel times for the improved departure flow pattern are given in 10.4.

Table 10.1: Parameters of Link Travel Time Functions

| link $a$ | $\beta_{1a}$ | $\beta_{2a}$ | $\beta_{3a}$ | $\beta_{4a}$ | $\beta_{5a}$ | $\beta_{6a}$ |
|---|---|---|---|---|---|---|
| 1—2 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 1—3 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 2—4 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |
| 3—4 | 1. | 0.001 | 0. | 0. | 0.015 | 0.002 |

We first discuss the flow propagation in the network, using link 2–4 in Table 10.4 as an example (see Figure 10.5).

Table 10.2: Departure Flow Patterns from Origin 1 to Destination 4

Initial Pattern

| Interval $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| O-D Flow | 60. | 0. | 0. | 0. | 0. | 0. | 0. | 0. |

Improved Pattern

| Interval $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| O-D Flow | 30. | 0. | 30. | 0. | 0. | 0. | 0. | 0. |

- Interval 2: Vehicles from link 1–2 enter link 2–4 during interval 2: $u_{24}(2) = 6.5$. Since the travel time on link 2–4, $\tau_{24}(2)$, is 1.0 interval for the first vehicles entering at the beginning of interval 2, these vehicles exit link 2–4 during interval 3: $v_{24}(3) = 6.5$.

- Interval 3: There are $u_{24}(3) = 8.5$ vehicles entering link 2–4 during interval 3. Since travel time $\tau_{24}(3)$ is 1.8 intervals for the first vehicles entering at the beginning of interval 3, only $v_{24}(4) = 7.4$ of the 8.5 vehicles exit during interval 4, and the remaining 1.1 vehicles exit during interval 5.

- Interval 4: The travel time $\tau_{24}(4)$ is 2.0 intervals for the first vehicles entering at the beginning of interval 4. There are $u_{24}(4) = 6.2$ additional vehicles entering during interval 4. The remaining 1.1 vehicles already on the link (see the paragraph above) exit during interval 5, allowing 5.5 vehicles also to exit, for a total of 6.6 exiting vehicles; $6.2 - 5.5 = 0.7$ vehicles of the 6.2 vehicles entering during interval 4 exit during interval 6.

- Interval 5: $u_{24}(5) = 8.8$ vehicles enter during interval 5. Those vehicles begin to exit link 2–4 during interval 6 (6.2 vehicles) and finish exiting during interval 7 (2.6 vehicles).

We now consider the improvement of travel times and disutilities from origin 1 to destination 4 as a result of the changed departure flow pattern. Assume that parameter $\alpha$ for waiting time equals 0.5; there is no arrival penalty in this example.

With the *initial* departure flow pattern shown in Table 10.2, the O-D travel time for the first vehicle departing at the beginning of interval 1 is 3 time intervals. (In Table 10.3, vehicles start to exit links 2–4 and 3–4 during interval 3; the O-D travel times are rounded integers.) Also, the O-D travel time for the last vehicle departing at the end of interval 1 is 6 time intervals. (In Table 10.3, vehicles finish exiting link 2–4 and 3–4 during interval 7). The *average* O-D travel time for these vehicles is $(3 + 6)/2 = 4.5$ time intervals. Since the average waiting time is zero for the departure flow pattern shown in Table 10.2, total disutility averages $(\alpha \cdot 0 + 4.5) \cdot 60 = 270$ units.

By way of comparison, with the *improved* departure flow pattern shown in Table 10.2, the O-D travel time for the first vehicle departing at the beginning of interval 1 is also 3.0 intervals. (Vehicles start to exit links 2–4 and 3–4 during interval 3; see footnote * in Figure 10.5.) Also the O-D travel time for the last vehicle departing at the end of interval 1 is 4.0 intervals; see footnote * in Figure 10.5. The O-D travel time for the first vehicle departing at the beginning of interval 3 is 3.0 intervals. Also, the O-D travel time for the last vehicle departing at the end of interval 3 is $7.0 - 3.0 = 4.0$ intervals; see footnote ** in Figure 10.5. The average O-D travel time for vehicles departing during interval 1 is $(3 + 4)/2 = 3.5$ time intervals, and the waiting time is 0. The average O-D travel time for vehicles departing during interval 3 is $(3 + 4)/2 = 3.5$ intervals, and the waiting time is 2 intervals. Total disutility averages $[(\alpha \cdot 0 + 3.5) \cdot 30 + (\alpha \cdot 2 + 3.5) \cdot 30] = 240$ units. (Recall that $\alpha = 0.5$.) Thus, by delaying some departures, total disutility due to waiting and traveling is reduced.

## 10.5 Notes

In contrast to departure time choice, arrival time choice may be more significant for drivers. One example is the home-to-work trip in which a latest arrival time needs to be strictly guaranteed. Our model uses desired arrival time to determine driver departure times. The penalty for late arrival must be very large. On the other hand, the disutility due to waiting at the origin is not important. In Chapter 11, we explored how the desired arrival time can be specified for different groups of travelers in a joint mode/departure time/route choice problem. In future research, we will also address work-to-home trips whose arrival time requirements seem to be more elastic since drivers are more concerned with avoiding congestion. The utility due to waiting at the origin is important because drivers prefer to stay at workplaces until traffic congestion decreases. Consequently, the arrival penalty is not significant.

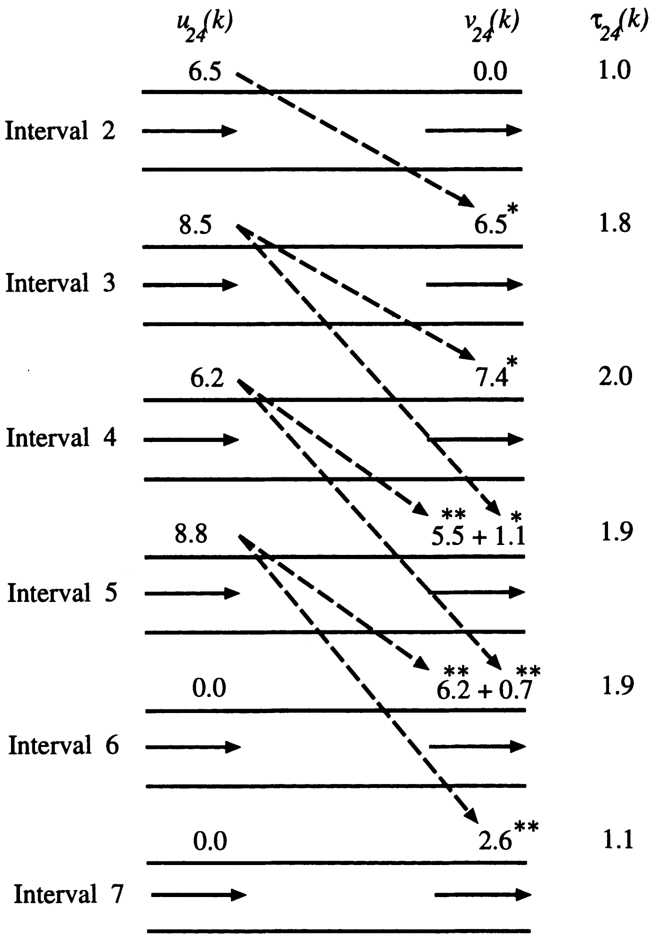
The bilevel program in this chapter is formulated as a hierarchical optimal control model. We note that this program can also be regarded as nonhierarchical so that the analysis of the optimality conditions is simpler. In other words, the upper-level route choice program and the lower-level departure time choice program have no leader-follower relationship in this case and can be solved simultaneously.

Table 10.3: Optimal Link Flows and Travel Times for Initial O-D Flows

| Interval $k$ | Link $a$ | Vehicles $x_a(k+1)$ | Inflow $u_a(k)$ | Exit Flow $v_a(k)$ | Vehicles $x_a(k)$ | Travel Time $c_a(k)$ & $\tau_a(k)$ |
|---|---|---|---|---|---|---|
| 1 | 1 — 2 | 30.0 | 30.0 | 0.0 | 0.0 | 1.9 |
| 2 | 1 — 2 | 21.7 | 0.0 | 8.3 | 30.0 | 3.8 |
| 3 | 1 — 2 | 21.7 | 0.0 | 0.0 | 21.7 | 1.9 |
| 4 | 1 — 2 | 12.4 | 0.0 | 9.2 | 21.7 | 3.2 |
| 5 | 1 — 2 | 0.0 | 0.0 | 12.4 | 12.4 | 3.6 |
| 1 | 1 — 3 | 30.0 | 30.0 | 0.0 | 0.0 | 1.9 |
| 2 | 1 — 3 | 21.7 | 0.0 | 8.3 | 30.0 | 3.8 |
| 3 | 1 — 3 | 21.7 | 0.0 | 0.0 | 21.7 | 1.9 |
| 4 | 1 — 3 | 12.4 | 0.0 | 9.2 | 21.7 | 3.2 |
| 5 | 1 — 3 | 0.0 | 0.0 | 12.4 | 12.4 | 3.6 |
| 1 | 2 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 2 | 2 — 4 | 8.3 | 8.3 | 0.0 | 0.0 | 1.1 |
| 3 | 2 — 4 | 0.0 | 0.0 | 8.3 | 8.3 | 2.2 |
| 4 | 2 — 4 | 9.2 | 9.2 | 0.0 | 0.0 | 1.1 |
| 5 | 2 — 4 | 13.4 | 12.4 | 8.3 | 9.2 | 2.4 |
| 6 | 2 — 4 | 3.8 | 0.0 | 9.6 | 13.4 | 2.7 |
| 7 | 2 — 4 | 0.0 | 0.0 | 3.8 | 3.8 | 1.2 |
| 1 | 3 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 2 | 3 — 4 | 8.3 | 8.3 | 0.0 | 0.0 | 1.1 |
| 3 | 3 — 4 | 0.0 | 0.0 | 8.3 | 8.3 | 2.2 |
| 4 | 3 — 4 | 9.2 | 9.2 | 0.0 | 0.0 | 1.1 |
| 5 | 3 — 4 | 13.4 | 12.4 | 8.3 | 9.2 | 2.4 |
| 6 | 3 — 4 | 3.8 | 0.0 | 9.6 | 13.4 | 2.7 |
| 7 | 3 — 4 | 0.0 | 0.0 | 3.8 | 3.8 | 1.2 |

Table 10.4: Optimal Link Flows and Travel Times for Improved O-D Flows

| Interval $k$ | Link $a$ | Vehicles $x_a(k+1)$ | Inflow $u_a(k)$ | Exit Flow $v_a(k)$ | Vehicles $x_a(k)$ | Travel Time $c_a(k)$ & $\tau_a(k)$ |
|---|---|---|---|---|---|---|
| 1 | 1 — 2 | 15.0 | 15.0 | 0.0 | 0.0 | 1.2 |
| 2 | 1 — 2 | 8.5 | 0.0 | 6.5 | 15.0 | 2.1 |
| 3 | 1 — 2 | 15.0 | 15.0 | 8.5 | 8.5 | 2.5 |
| 4 | 1 — 2 | 8.8 | 0.0 | 6.2 | 15.0 | 2.0 |
| 5 | 1 — 2 | 0.0 | 0.0 | 8.8 | 8.8 | 2.3 |
| 1 | 1 — 3 | 15.0 | 15.0 | 0.0 | 0.0 | 1.2 |
| 2 | 1 — 3 | 8.5 | 0.0 | 6.5 | 15.0 | 2.1 |
| 3 | 1 — 3 | 15.0 | 15.0 | 8.5 | 8.5 | 2.5 |
| 4 | 1 — 3 | 8.8 | 0.0 | 6.2 | 15.0 | 2.0 |
| 5 | 1 — 3 | 0.0 | 0.0 | 8.8 | 8.8 | 2.3 |
| 1 | 2 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 2 | 2 — 4 | 6.5 | 6.5 | 0.0 | 0.0 | 1.0 |
| 3 | 2 — 4 | 8.5 | 8.5 | 6.5 | 6.5 | 1.8 |
| 4 | 2 — 4 | 7.4 | 6.2 | 7.4 | 8.5 | 2.0 |
| 5 | 2 — 4 | 9.5 | 8.8 | 6.6 | 7.4 | 1.9 |
| 6 | 2 — 4 | 2.6 | 0.0 | 6.9 | 9.5 | 1.9 |
| 7 | 2 — 4 | 0.0 | 0.0 | 2.6 | 2.6 | 1.1 |
| 1 | 3 — 4 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| 2 | 3 — 4 | 6.5 | 6.5 | 0.0 | 0.0 | 1.0 |
| 3 | 3 — 4 | 8.5 | 8.5 | 6.5 | 6.5 | 1.8 |
| 4 | 3 — 4 | 7.4 | 6.2 | 7.4 | 8.5 | 2.0 |
| 5 | 3 — 4 | 9.5 | 8.8 | 6.6 | 7.4 | 1.9 |
| 6 | 3 — 4 | 2.6 | 0.0 | 6.9 | 9.5 | 1.9 |
| 7 | 3 — 4 | 0.0 | 0.0 | 2.6 | 2.6 | 1.1 |




* These vehicles departed the origin during interval 1.
** These vehicles departed the origin during interval 3.

Figure 10.5: Flow Propagation on Link 2–4

# Chapter 11

# Combined Departure Time/ Mode/Route Choice Models

We consider the efficient operation of an integrated transportation system within an IVHS environment. A dynamic route guidance system would improve utilization of the overall capacity of the transportation system so as to reduce travel times, congestion and incidents. By providing early detection of incidents and congestion in the transportation network, the route guidance system would redistribute traffic among the available modes and routes when there is excess capacity in some parts of the road network or shift the departure times of travelers to avoid peak-hour congestion when no additional road capacity is available. Furthermore, the route guidance system would provide travelers with accurate, current information on both transit and road networks so that some motorists could make their own time-cost tradeoffs and shift to transit, if appropriate.

In this chapter, we address the dynamic mode/departure time/route choice problem with multiple stratifications of users, each with a different propensity to use transit or high occupancy vehicles. This stratification could be by income level, automobile ownership, or vehicle occupancy regulation. As in conventional planning models, socio-economic factors can be considered in mode choice models in time-dependent circumstances. The advantage of formally considering multiple classes of users, who value their time and convenience differently, is to model travelers' mode choices more accurately.

Furthermore, different people have different propensities to use different travel modes. For example, senior citizens may prefer local streets to freeways; therefore, freeways may constitute a specific mode for this particular group of travelers. In principle, by considering groups or classes of travelers with specified time-cost tradeoffs in each group, it is possible to predict mode choice deterministically.

A shift of travelers from cars to transit or from low-occupancy cars to high-occupancy cars may significantly decrease road congestion and increase the efficiency of the overall transportation system. Moreover, in the road network, journey-to-work trips have especially important requirements for avoiding congested routes so as to arrive at work on time. Since each departure time choice is based on prevailing origin-destination travel times, any change in departure times will alter the traffic flow patterns in the network so that route choice decisions of other travelers will be modified. Therefore, in order to achieve this balanced allocation to various departure times and different modes, an integrated model including all elements (mode, departure time and route choice) should be constructed.

There have been extensive studies in mode choice analysis. Wilson (1969) studied the trip distribution, modal split and trip assignment problem using entropy maximizing methods. Florian and Nguyen (1978) presented a combined trip distribution, modal split and trip assignment model. Route choice in their model is based on the user-optimal principle, and the mode choice is given by a logit model. Boyce (1978) also discussed the equilibrium solutions for combined location, mode choice and trip assignment models.

Studies into multiple groups of travelers in travel choice models were begun by Dafermos (1972). She presented traffic assignment models in a multiclass-user transportation network. Later, LeBlanc and Abdulaal (1982) presented combined mode split/assignment and distribution/mode split/assignment models with multiple groups of travelers.

In this chapter, we present a dynamic user-optimal (DUO) mode, departure time and route choice model for a transportation network with multiple origin-destination pairs. The model developed in this chapter extends the joint departure time/route choice model in Chapter 10 to the case in which the combined mode, departure time and route choice should be considered with multiple classes of travelers. We model this choice problem by specifying that a given number of travelers are ready for departure between each origin-destination pair at the beginning of each of several short time periods. However, motorists may shift to transit or delay their departure times to reduce their overall travel costs. The model extends our previous dynamic user-optimal departure time and route choice model in two important respects: 1) alternative mode choices are available; and 2) travelers are stratified into different groups according to travelers' socio-economic characteristics.

The model is formulated as a two-stage simultaneous (non-hierarchical) optimization program. The first-stage problem represents dynamic logit-type modal choice. The second-stage problem represents a hierarchical leader-follower problem which solves the DUO departure time and route choice problem for motorists.

The problem is described in the next section. The formulation of the two-stage model is described in Sections 11.2, 11.3, and 11.4. In Section 11.5, the equivalence of the optimality conditions of the two-stage program with the DUO mode/departure time/route choice conditions is demonstrated. Finally, the properties of the model are discussed.

## 11.1 Two-Stage Travel Choice Model

A multiple origin-destination transportation network is considered. For simplicity, the transportation network is defined to consist of a transit network and a road network. Consider a fixed time period $[0, T]$. The length of the time period is sufficient to allow all travelers in peak period to complete their trips. We consider the following mode, departure time, and route choice situation with $(K+1)$ points $T_1, T_2, \cdots, T_{K+1}$ on the time horizon for the fixed time period $[0, T]$. These points divide the time period $[0, T]$ into $K$ intervals, where $T_1 = 0$ and $T_{K+1} = T$. Any interval $k$ is denoted as $[T_k, T_{k+1}]$, $k = 1, 2, \cdots, K$, and these intervals may or may not be equal in length. The length of each interval would typically be 15 to 30 minutes for non-peak periods. For the peak periods, this interval should be set to be consistent with the time headway of transit operations, such as 5 or 10 minutes. We also assume no departure time choice option for transit users, implying that headways are uniform over the time period $[0, T]$.

For each O-D pair $rs$, the group of travelers departing during period $k$ can be further stratified into $K$ smaller sub-groups according to the socio-economic characteristics of each traveler. There are several approaches to stratify travelers. The typical one for mode choice problem is to classify travelers based on income and age (see Table 11.1). There are 9 combinations in this approach. Other approaches of stratification are discussed in the multi-group route choice problems in Chapter 12.

Table 11.1: Stratification of Travelers Based on Income and Age

| | Degree of Change | | |
|---|---|---|---|
| Income | High | Middle | Low |
| Age | Old | Middle | Young |

For any interval $k$, travelers in group $m$ are ready to depart by transit or car at an initial time $T_k$. Thus, based on the total disutilities of using transit and using car, travelers are split into 2 groups. For each O-D pair $rs$ and each traveler group $m$, the total number of travelers departing during interval $k$ is denoted as $R_m^{rs}(k)$ and is given exogenously. Similarly, let $Q_m^{rs}(k)$ and $G_m^{rs}(k)$ denote the total numbers of travelers of group $m$ departing by transit and automobile, respectively, from origin $r$ to destination $s$ during interval $k$. It follows that

$$G_m^{rs}(k) + Q_m^{rs}(k) = R_m^{rs}(k) \qquad \text{given} \qquad \forall m, r, s, k. \tag{11.1}$$

Moreover, motorists may prefer to delay their departure times within period $[T_k, T_{k+1}]$ in order to reduce their driving times. Note that motorists departing during interval $k$ can only shift their departure times within interval $k$. In future studies, the above assumption can be relaxed so that motorists departing during interval $k$ can shift their departure times within intervals $l \geq k$.

Drivers are assumed to make their departure time choices so as to minimize their individual disutility functions defined on travel time and pre-trip delay. The criteria for choosing each departure time consist of:

1. the waiting time before departure;
2. the actual travel time between the origin and destination;
3. a bonus for early arrival or a penalty for late arrival.

Of course, a change of departure flow rate will change the traffic on the road network so that the actual origin-destination travel times will change.

Travelers' choices of mode, departure time and route are interrelated decisions. Given a desired arrival time, say at the workplace, choices of mode and departure time depend on the traveler's estimate of en route travel time on each mode. Thus, choice of mode depends on the travel disutilities of alternative modes, which also may vary by time of day. Because there is a partition of the decision variables between two ordered stages: mode choice and departure time/route choice for motorists, we use a two-stage simultaneous optimization programming formulation in which mode choices are first-stage decision variables and departure time/route choices for motorists are second-stage decision variables. See Figure 11.1.

Furthermore, in the road network, choice of route depends on the travel times of alternative routes, which also may vary by time of day. In the second-stage optimization program (Figure 11.1), these choices are represented as a hierarchical bilevel optimal control program, which is equivalent to a dynamic leader-follower game (Cruz, 1978). The dynamic route choice problem for motorists is formulated as a single optimal control problem. In the equivalent dynamic game, this formulation corresponds to a single *controller* allocating fixed departure flows at each time $t$ to user-optimal routes, given the departure frequencies. We define this controller to be the leader of the game.

For each O-D pair, a *departure time coordinator* for motorists determines the departing flows at each time $t$. These departure coordinators are defined as the followers of the game, and are represented by $n(n-1)$ O-D-specific optimal control problems to which the user-optimal travel time at time $t$ is exogenous. Since these problems are independent by O-D pair, $n(n-1)$ separate problems for each interval $k$ can be used to represent all O-D pairs.

The hierarchical bilevel program for motorists assumes that the route choice controller, who is responsible for allocating all O-D flows to routes for each time interval $k$, knows the objective of each departure coordinator. Through this knowledge, the route choice coordinator is able to achieve a lower value of his/her objective function than if the departure time objectives were unknown. In contrast, the departure time coordinators know only the O-D

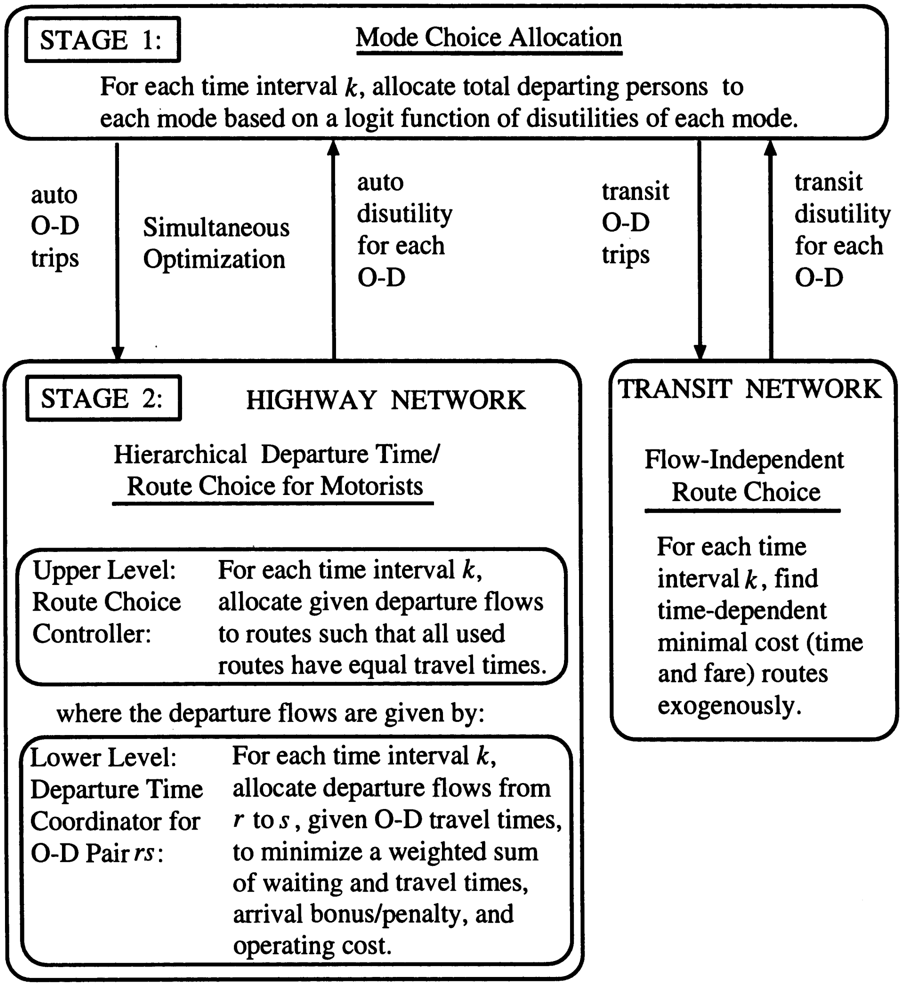



Figure 11.1: Two-Stage Simultaneous Travel Choice Model

travel times at time $t$ provided to them. Knowledge of the route choice objective is not needed for their coordination task.

In the following the lower level problem for motorists representing the departure time coordinators is formulated as $n(n-1)$ optimal control problems. Then, the upper level optimal control problem for motorists is defined to represent the route choice controller. Finally, the overall two-stage simultaneous travel choice program is presented in Figure 11.1.

## 11.2 First Stage: Mode Choice Problem

We consider the mode-choice problem for travelers in group $m$ during any interval $[T_k, T_{k+1}]$. For simplicity, it is assumed that the automobile occupancy factor is 1; however, an occupancy factor could be used to convert the person flow to vehicle flow since both automobile and transit flows are expressed in terms of persons per unit of time. We require the modal share for automobile and transit be given by a binary logit function for each time interval $k$:

$$P(\text{auto}) = \frac{\exp(-\theta_m \mu_m^{rs}(k))}{\exp(-\theta_m \nu_m^{rs}(k)) + \exp(-\theta_m \mu_m^{rs}(k))} \tag{11.2}$$

$$P(\text{transit}) = \frac{\exp(-\theta_m \nu_m^{rs}(k))}{\exp(-\theta_m \nu_m^{rs}(k)) + \exp(-\theta_m \mu_m^{rs}(k))} \tag{11.3}$$

where $\mu_m^{rs}(k)$ and $\nu_m^{rs}(k)$ are the minimal total disutilities by auto and by transit, respectively, for travelers in group $m$ departing during time interval $k$ from $r$ to $s$, and $\theta_m$ is a positive parameter which needs to be calibrated. Using the above function, we can calculate $G_m^{rs}(k)$, which is the total number of departing motorists in group $m$ during interval $k$ and is also required for the second-stage departure time/route choice program for motorists.

We formulate the mode choice problem as a discrete-time nonlinear programming problem (NLP), which is a dynamic extension of many previous models in the static environment. Our model is presented in Section 11.4. The minimal total disutility by car, $\mu_m^{rs}(k)$, will be determined in the second-stage departure time/route choice problem for motorists.

We first discuss the route choice problem for the transit network. Since the transit network has fixed fares and travel times, minimal cost routes can be determined exogenously in order to compute $\nu_m^{rs}(k)$, the minimal total disutility by transit. The transit network is assumed to consist of a set of access links, transfer links and transit line segments. As stated in Florian and Nguyen (1978), a transit route is composed of a number of segments. We associate a time-dependent travel cost with each route of the transit network during interval $k$ as follows:

1. a walking (or driving time) and a waiting time with access links to/from stations;

2. a walking time and a waiting time with a transfer link;

3. an in-vehicle time with a line segment;

4. origin-destination fare.

For any interval $k$, it is assumed that the link travel cost is independent of the transit link flows. Thus, the time-dependent travel cost $C_p(k)$ over route $p$ for travelers departing during interval $k$ can be easily computed. We suppose that a dynamic user-optimal state occurs on the road network, which will be discussed in Section 11.3, where we discuss departure time choice and route choice. Similarly, we suppose that a dynamic user-optimal state occurs on the transit network. That is

$$C_p(k) = \nu_m^{rs}(k)/\eta_m \qquad \text{if } Q_{mp}^{rs}(k) > 0 \tag{11.4}$$

$$C_p(k) \geq \nu_m^{rs}(k)/\eta_m \qquad \text{if } Q_{mp}^{rs}(k) = 0 \tag{11.5}$$

where $\eta_m$ is the disutility scaling parameter associated with the social-economic characteristics of group $m$ travelers. Since $C_p(k)$ is fixed, $\nu_m^{rs}(k)/\eta_m$ is the minimal travel cost from $r$ to $s$ for travelers departing during interval $k$. Thus, the route choice problem in a transit network is simply a time-dependent minimal route cost problem for each interval $k$.

## 11.3 Second Stage: Departure Time/Route Choice for Motorists

Next we consider the second-stage problem of departure time/route choice for the road network. The second-stage problem can be formulated as an optimization program or a variational inequality. In this chapter, we consider how to formulate a bilevel optimal control program for departure time/route choice. The variational inequality model is presented in Chapter 14.

### 11.3.1 Lower-Level: Departure Time Choice for Motorists

A disutility function $\mathcal{U}_m^{rs}(t)$ based on departure times is defined for group $m$ drivers departing from origin $r$ to destination $s$ at time $t$. This function represents a weighted sum of total elapsed time during the journey, including the waiting time at the origin node, the driving time during the trip, any bonus/penalty for early/late arrival, and auto operating cost. We focus the following discussion on the disutility for any group $m$ motorists. We define one unit of in-vehicle driving time to equal $\gamma_m$ units of disutility, and one unit of waiting time prior to departure to be equivalent to $\alpha_m$ units of disutility ($\alpha_m \leq 1$).

For each group of travelers departing during interval $k$, there is a required arrival time interval $[t_{rs}^*(k) - \Delta_{rs}(k), t_{rs}^*(k) + \Delta_{rs}(k)]$ at each destination $s$, where $t_{rs}^*(k)$ is the center of the required arrival time interval associated with travelers departing from origin $r$ during interval $k$ toward destination $s$.

$\Delta_{rs}(k)$ represents the arrival time flexibility at destination $s$ for travelers departing from origin $r$ during interval $k$ toward destination $s$. We also define the disutility for early or late arrival as follows:

$$\mathcal{V}_m^{rs}[t, \pi^{rs}(t); t_{rs}^*(k)] =$$

$$\begin{cases} \beta_{1i}[t + \pi^{rs}(t) - t_{rs}^*(k) + \Delta_{rs}^*(k)]^2 & \text{if } t + \pi^{rs}(t) < t_{rs}^*(k) - \Delta_{rs}^*(k) \\ 0 & \text{if } |t + \pi^{rs}(t) - t_{rs}^*(k)| \leq \Delta_{rs}^*(k) \\ \beta_{2i}[t + \pi^{rs}(t) - t_{rs}^*(k) - \Delta_{rs}^*(k)]^2 & \text{if } t + \pi^{rs}(t) > t_{rs}^*(k) + \Delta_{rs}^*(k) \end{cases}$$

where time $t$ is the departure time of travelers and $\beta_{1i}, \beta_{2i}$ are parameters ($\beta_{1i} \leq 0, \beta_{2i} \gg \alpha$). This relationship is shown in Figure 11.2. This disutility of late arrival is continuous and differentiable with respect to $t$ and $\pi^{rs}(t)$.

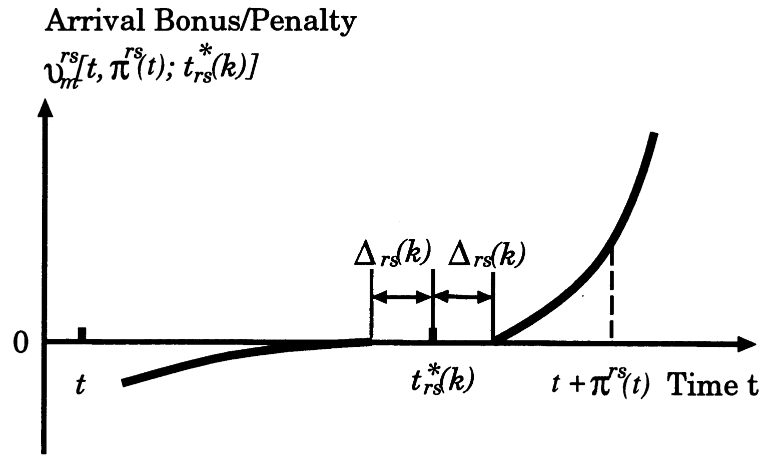



Figure 11.2: Arrival Time Disutility

Let $C^{rs}$ denote the auto operating cost from origin $r$ to destination $s$; for simplicity, we assume this cost is fixed for each O-D pair. We define one unit of operating cost to equal $\xi_m$ units of disutility. Even though the operating cost is not directly associated with the departure time choice of motorists, it is an important decision element when choosing between transit and car and is used in the mode choice model. Thus, the disutility function for group $m$ motorists departing during interval $k$ is constructed as

$$\mathcal{U}_m^{rs}(t; k) = \alpha_m(t - T_k) + \gamma_m \pi^{rs}(t) + \mathcal{V}_m^{rs}[t, \pi^{rs}(t); t_{rs}^*(k)] + \xi_m C^{rs}$$

$$\forall m, r, s, k, t \in (T_k, T_{k+1}) \qquad (11.6)$$

where time $t$ is the departure time of travelers of group $m$ departing during interval $k$ and $T_k$ is the starting time of interval $k$. Note that since travelers of group $m$ desire to depart at time $T_k$, $\alpha_m(t - T_k)$ is the disutility of waiting.

For O-D pair $rs$ and any time $t$, denote $f_m^{rs}(t)$ and $F_m^{rs}(t)$ as the departure rate and the cumulative number of departures of group $m$ motorists, respectively. Then for motorists, we construct the dynamic user-optimal departure time choice conditions as

$$\mathcal{U}_m^{rs}(t;k) = \mu_m^{rs}(k) \qquad \text{if } f_m^{rs}(t) > 0 \qquad \forall m, r, s, k, t \in (T_k, T_{k+1}); \tag{11.7}$$

$$\mathcal{U}_m^{rs}(t;k) \geq \mu_m^{rs}(k) \qquad \text{if } f_m^{rs}(t) = 0 \qquad \forall m, r, s, k, t \in (T_k, T_{k+1}); \tag{11.8}$$

where $\mu_m^{rs}(k)$ is the minimal $rs$ disutility for group $m$ motorists departing during interval $k$. By definition of $f_m^{rs}(t)$ (the O-D departure rate) and $F_m^{rs}(t)$ (the cumulative number of departing vehicles), it follows that

$$\int_0^t f_m^{rs}(t)dt = F_m^{rs}(t) \qquad \forall m, r, s, k, t \in (T_k, T_{k+1}); \tag{11.9}$$

or

$$\frac{dF_m^{rs}(t)}{dt} = f_m^{rs}(t) \qquad \forall m, r, s, k, t \in (T_k, T_{k+1}); \tag{11.10}$$

where $F_m^{rs}(t)$ and $f_m^{rs}(t)$ are state and control variables, respectively. Also, we have corner point or boundary conditions as follows:

$$F_m^{rs}(T_{k+1}) = \sum_{j=1}^{k} G_m^{rs}(j) \qquad \forall m, r, s, k; \tag{11.11}$$

$$F_m^{rs}(T_1) = F_m^{rs}(0) = 0 \qquad \forall m, r, s; \tag{11.12}$$

where the total number of departing group $m$ motorists during interval $j$ $(1 \leq j \leq k)$, $G_m^{rs}(j)$, is given by the first-stage mode choice program. When solving the second-stage departure time/route choice program for motorists, the lower-level departure time choice model decides the instantaneous departure rate $f_m^{rs}(t)$ for each group $m$ and each O-D pair $rs$. Then, the departure rates sum up as follows:

$$f^{rs}(t) = \sum_m f_m^{rs}(t) \qquad \forall r, s$$

where $f^{rs}(t)$ is the input for the upper-level route choice model for motorists.

## 11.3.2 Upper-Level: Route Choice for Motorists

In this section, we discuss the upper-level problem of route choice in the hierarchical departure time/route choice program for the road network. We assume that all motorists in all groups choose minimal travel time routes. The stratification of travelers is used for mode and departure time choices. Thus, the dynamic route choice problem is the same as in Chapter 7. The stratification of travelers for route choice problem is addressed in Chapters 12-13.

The dynamic route choice problem is to find the dynamic trajectories of link states and inflow and exit flow control variables, given the road network, the link travel time functions and the time-dependent O-D flow requirements. The O-D requirements are specified by the lower-level departure time choice problem for motorists and are therefore temporarily fixed in the upper-level route choice problem for motorists. The formulation of the problem is based on the underlying choice criterion that each traveler uses the route that minimizes his/her actual travel time when departing from the origin to his/her destination.

As formulated in Chapter 7, the minimal actual O-D travel time $\pi^{rs}(t)$ for motorists departing at time $t$ from origin $r$ to destination $s$ can be determined from the following equation

$$F^{rs}(t) = E^{rs}[t + \pi^{rs}(t)] \qquad \forall r, s \tag{11.13}$$

where $F^{rs}(t)$ is given by the lower-level departure time choice model and $E^{rs}[t+\pi^{rs}(t)]$ is computed from flow conservation equations for destinations. This constraint is also used to guarantee that the motorists departing at the same time $t$ from origin $r$ to destination $s$ should arrive at the destination $s$ at the same time $[t + \pi^{rs}(t)]$. The detailed analysis is given in Chapter 7. Note that in the above equation, the actual travel time $\pi^{rs}(t)$ may be greater than any interval $[T_k, T_{k+1}]$. Thus, we implicitly assume that $\pi^{rs}(t)$ is differentiable over the time period $[0, T]$.

## 11.4 Formulation of the Two-Stage Travel Choice Model

Using optimal control and nonlinear programming theory, a two-stage simultaneous optimization program of the dynamic user-optimal mode, departure time and route choice model is formulated as follows.

**FIRST-STAGE: Mode Choice**

$$\begin{aligned} \min_{G,Q} \quad & \sum_{rs}\sum_{k}\sum_{m}\left\{G_m^{rs}(k)\left[\frac{1}{\theta_m}\ln G_m^{rs}(k) + \mu_m^{rs}(k)\right]\right. \\ + \quad & \left. Q_m^{rs}(k)\left[\frac{1}{\theta_m}\ln Q_m^{rs}(k) + \nu_m^{rs}(k)\right]\right\} \end{aligned} \tag{11.14}$$

s.t.

$$G_m^{rs}(k) + Q_m^{rs}(k) = R_m^{rs}(k) \quad \text{given} \quad \forall m, r, s, k; \tag{11.15}$$

$$G_m^{rs}(k) \geq 0, \qquad Q_m^{rs}(k) \geq 0 \quad \forall m, r, s, k; \tag{11.16}$$

where the minimal disutility $\mu_m^{rs}(k)$ for motorists solves

**SECOND-STAGE: Departure Time/Route Choice for Motorists**

**Upper-Level: Route Choice for Motorists**

$$\min_{u,v,x,e,E,f_p^{rs},F_p^{rs};\pi} \quad \int_0^T \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t),\omega]d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t),\omega]d\omega \right\} dt \tag{11.17}$$

s.t.

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a,p,r,s; \tag{11.18}$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p,r,s; \tag{11.19}$$

$$\frac{dF_p^{rs}(t)}{dt} = f_p^{rs}(t) \qquad \forall p,r,s; \tag{11.20}$$

$$f_p^{rs}(t) = \sum_{a\in A(r)} \delta_{ap}^{rs} u_{ap}^{rs}(t) \qquad \forall p,r,s; \tag{11.21}$$

$$e_p^{rs}(t) = \sum_{a\in B(s)} \delta_{ap}^{rs} v_{ap}^{rs}(t) \qquad \forall p,r,s; \tag{11.22}$$

$$\sum_{a\in B(j)} v_{ap}^{rs}(t) = \sum_{a\in A(j)} u_{ap}^{rs}(t) \qquad \forall j,p,r,s; j\neq r,s; \tag{11.23}$$

$$F^{rs}(t) = E^{rs}[t+\pi^{rs}(t)] \qquad \forall r,s; \tag{11.24}$$

$$x_{ap}^{rs}(t) = \sum_{b\in\tilde{p}} \{x_{bp}^{rs}[t+\tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t+\tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall r,s,p,j; a\in B(j); j\neq r; \tag{11.25}$$

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \tag{11.26}$$

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \qquad \sum_{rs} x_a^{rs}(t) = x_a(t), \qquad \forall a; \tag{11.27}$$

$$\sum_p E_p^{rs}(t) = E^{rs}(t), \qquad \sum_p F_p^{rs}(t) = F^{rs}(t), \qquad \forall r,s; \tag{11.28}$$

$$\sum_p f_p^{rs}(t) = f^{rs}(t), \qquad \sum_p e_p^{rs}(t) = e^{rs}(t), \qquad \forall r,s; \tag{11.29}$$

$$x_{ap}^{rs}(t)\geq 0, \quad u_{ap}^{rs}(t)\geq 0, \quad v_{ap}^{rs}(t)\geq 0 \quad \forall a,p,r,s; \tag{11.30}$$

$$e_p^{rs}(t)\geq 0, \quad f_p^{rs}(t)\geq 0, \quad E_p^{rs}(t)\geq 0, \quad F_p^{rs}(t)\geq 0 \qquad \forall p,r,s; \tag{11.31}$$

$$E_p^{rs}(0)=0, \quad F_p^{rs}(0)=0 \quad \forall p,r,s; \quad x_{ap}^{rs}(0)=0, \quad \forall a,p,r,s; \tag{11.32}$$

where the O-D departure $f^{rs}(t)$ and $F^{rs}(t)$ for motorists solve

**Lower-Level: Departure Time Choice for Motorists**

$$\min_{f^{rs}, f_m^{rs}, F_m^{rs}} \sum_{k=1}^{K} \int_{T_k}^{T_{k+1}} \sum_{rs} \sum_{m} \left\{ \int_0^{f_m^{rs}(t)} U_m^{rs}(t; k) d\omega \right\} dt \tag{11.33}$$

s.t.

$$\frac{dF_m^{rs}(t)}{dt} = f_m^{rs}(t) \qquad \forall m, r, s, k, t \in (T_k, T_{k+1}); \tag{11.34}$$

$$F_m^{rs}(T_{k+1}) = \sum_{j=1}^{k} G_m^{rs}(j) \qquad \forall m, r, s, k; \tag{11.35}$$

$$F_m^{rs}(T_1) = F_m^{rs}(0) = 0 \qquad \forall m, r, s; \tag{11.36}$$

$$f^{rs}(t) = \sum_{m} f_m^{rs}(t) \qquad \forall r, s, t \in (T_k, T_{k+1}); \tag{11.37}$$

$$f_m^{rs}(t) \geq 0, \qquad F_m^{rs}(t) \geq 0 \qquad \forall m, r, s, k, t \in (T_k, T_{k+1}); \tag{11.38}$$

where $G_m^{rs}(k)$ is given by first-stage mode choice program.

The first-stage program (11.14)-(11.16) is a discrete time NLP program. Two terms in the objective function are conventional entropy functions of trip flows by transit and by auto, respectively, for each interval $k$. The only constraints are flow conservation (11.15) and nonnegativity (11.16). The decision variables are $G_m^{rs}(k)$ and $Q_m^{rs}(k)$, which represent the total numbers of travelers by car and by transit, respectively. The disutility function $\mu_m^{rs}(k)$ for motorists is determined by the second-stage departure time choice program. The disutility function $\nu_m^{rs}(k)$ for transit travelers is calculated exogenously.

The second-stage program (11.17)-(11.38) is a hierarchical bilevel optimal control program for motorists. The upper-level model (11.17)-(11.32) for motorists is the ideal DUO route choice model presented in Chapter 7. In this model, the route specific departure variables $f_p^{rs}(t)$ and $F_p^{rs}(t)$ must be found. The two terms of the objective function are similar to the objective function of the well-known static user-optimal (UO) model. The first three constraints (11.18)-(11.20) are state equations for each link $a$ and cumulative effects at origins and destinations. Equations (11.21)-(11.23) are flow conservation constraints at each node including origins and destinations. Equation (11.24) equilibrates the actual route travel times. The other constraints include flow propagation, definitional, nonnegativity, and boundary conditions. In summary, in the upper-level model the control variables are $u_{ap}^{rs}(t)$, $v_{ap}^{rs}(t)$, $e_p^{rs}(t)$, and $f_p^{rs}(t)$; the state variables are $x_{ap}^{rs}(t)$, $E_p^{rs}(t)$, and $F_p^{rs}(t)$; and the functionals are $\pi^{rs}(t)$. The inputs for this upper-level model include the instantaneous O-D flow rate $f^{rs}(t)$, which is a decision variable in the lower-level departure time choice model.

In the lower-level departure time choice model (11.33)-(11.38) for motorists, we have only state equations (11.34), boundary conditions (11.35) and (11.36), flow conservation equations (11.37) and nonnegativity conditions

(11.38). The control variables are $f_m^{rs}(t)$ and $f^{rs}(t)$, and the state variables are $F_m^{rs}(t)$, which represent cumulative group $m$ motorist departures from origin $r$ to destination $s$. The inputs for this model include $G_m^{rs}(k)$ and $\pi^{rs}(t)$. The total number of motorists $G_m^{rs}(k)$ in group $m$ during interval $k$ is determined by the first-stage mode choice program. The actual O-D travel time function $\pi^{rs}(t)$ for motorists is determined by the upper-level route choice model for motorists. As shown in the optimality conditions, the minimal disutility $\mu_m^{rs}(k)$ for motorists can be computed after solving this bilevel model. Note that $\mu_m^{rs}(k)$ is an input to the first-stage program.

The relationship of decision variables in the overall two-stage simultaneous optimization program is shown in Figure 11.3. The lower-level departure time choice model provides a set of corner point (boundary) conditions at intervals $T_1, T_2, \cdots, T_{K+1}$ for the bilevel departure time/route choice program for motorists. Also, in the two-stage simultaneous optimization program, the first-stage mode choice program provides boundary conditions for a cluster of hierarchical departure time/route choice programs for each interval $k$ in the second-stage program. We prove in the next section that the optimal solution to the two-stage simultaneous optimization program satisfies the required DUO mode/departure time/route choice conditions.

## 11.5 Optimality Conditions

Since the two-stage travel choice programs (11.14)-(11.16) and (11.17)-(11.38) are solved simultaneously, we discuss the optimality conditions of each program separately.

### 11.5.1 Optimality Conditions for First-Stage Program

The Lagrangian for the first-stage mode choice program (11.14)-(11.16) is

$$\begin{aligned}\mathcal{L} &= \sum_{rs}\sum_{k}\sum_{m}\left\{G_m^{rs}(k)\left[\frac{1}{\theta_m}\ln G_m^{rs}(k) + \mu_m^{rs}(k)\right]\right.\\ &+ \left. Q_m^{rs}(k)\left[\frac{1}{\theta_m}\ln Q_m^{rs}(k) + \nu_m^{rs}(k)\right]\right\}\\ &+ \sum_{rs}\sum_{k}\sum_{m}\eta_m^{rs}(k)[G_m^{rs}(k) + Q_m^{rs}(k) - R_m^{rs}(k)]\end{aligned}$$

where $\eta_m^{rs}(k)$ is the Lagrange multiplier associated with group $m$ travelers for each O-D pair $rs$ and interval $k$. Then if $G_m^{rs}(k) > 0$ and $Q_m^{rs}(k) > 0$, one part of the Karush-Kuhn-Tucker conditions are the following:

$$\frac{1}{\theta_m}[\ln G_m^{rs}(k) + 1] + \mu_m^{rs}(k) + \eta_m^{rs}(k) = 0, \qquad \forall m, r, s, k, \tag{11.39}$$

$$\frac{1}{\theta_m}[\ln Q_m^{rs}(k) + 1] + \nu_m^{rs}(k) + \eta_m^{rs}(k) = 0, \qquad \forall m, r, s, k. \tag{11.40}$$

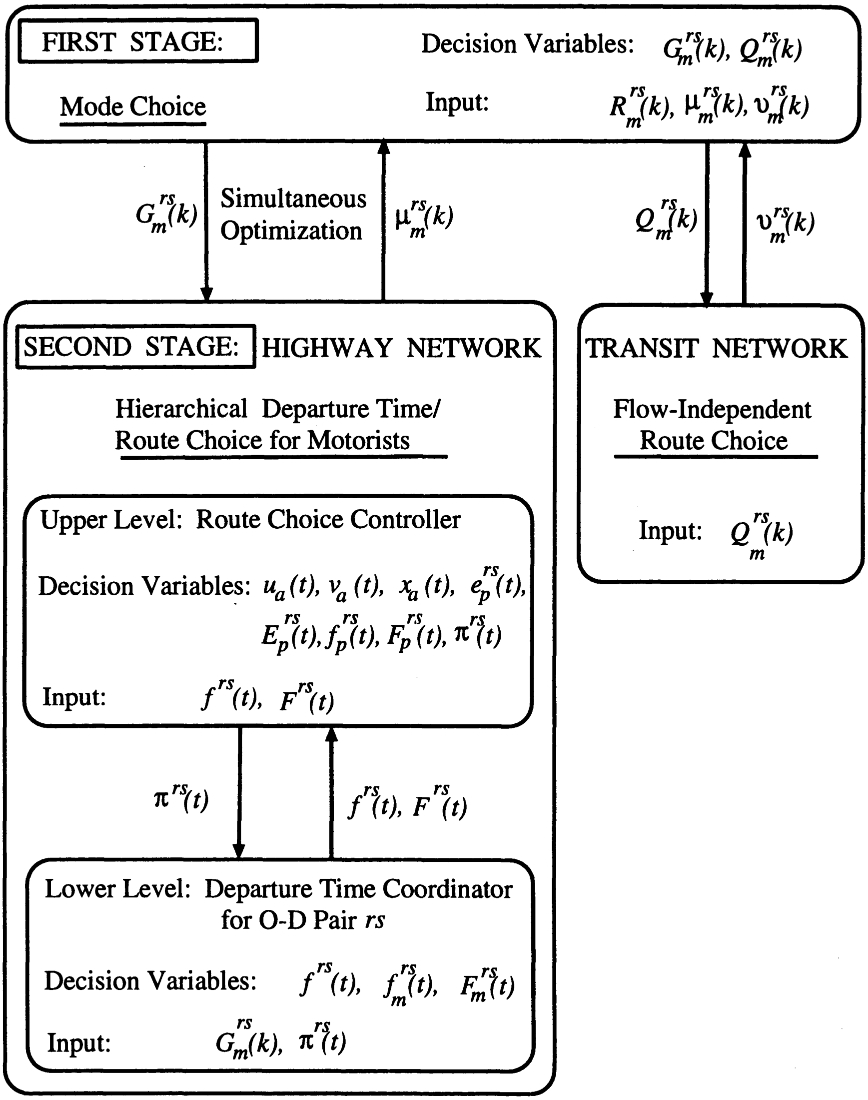



Figure 11.3: Decision Variables in the Two-Stage Travel Choice Program

We single out a particular origin-destination pair $rs$. Thus, for each group $m$ travelers during interval $k$, the O-D flows for each mode are given by the expression

$$G_m^{rs}(k) = \exp(-\theta_m \eta_m^{rs}(k) - 1) \cdot \exp(-\theta_m \mu_m^{rs}(k)), \qquad \forall m, r, s, k, \qquad (11.41)$$

$$Q_m^{rs}(k) = \exp(-\theta_m \eta_m^{rs}(k) - 1) \cdot \exp(-\theta_m \nu_m^{rs}(k)), \qquad \forall m, r, s, k. \qquad (11.42)$$

Substituting the above two equations into equation (11.15), we obtain

$$\exp(-\theta_m \eta_m^{rs}(k) - 1) = \frac{R_m^{rs}(k)}{\exp(-\theta_m \mu_m^{rs}(k)) + \exp(-\theta_m \nu_m^{rs}(k))},$$

$$\forall m, r, s, k. \qquad (11.43)$$

Substituting the above equation into equation (11.41), we obtain the modal share expression for automobile users

$$G_m^{rs}(k) = R_m^{rs}(k) \frac{\exp(-\theta_m \mu_m^{rs}(k))}{\exp(-\theta_m \mu_m^{rs}(k)) + \exp(-\theta_m \nu_m^{rs}(k))}, \qquad \forall m, r, s, k. \qquad (11.44)$$

Equation (11.44) is the required logit-type mode choice condition for interval $k$.

## 11.5.2 Optimality Conditions for Second-Stage Program

Since the lower-level model (11.33)-(11.38) for motorists is a cluster of OCP programs for each interval $[T_k, T_{k+1}]$, $k = 1, 2, \cdots, K$, we first derive the optimality conditions for an arbitrary interval $[T_k, T_{k+1}]$. The Hamiltonian of the lower-level model for each interval $k$ is

$$\mathcal{H}_1 = \sum_{rs} \sum_m \int_0^{f_m^{rs}(t)} \mathcal{U}_m^{rs}(t; k) d\omega + \sum_{rs} \sum_m \phi_m^{rs}(t)\, f_m^{rs}(t)$$

where $\phi_m^{rs}(t)$ is the Lagrange multiplier associated with group $m$ travelers for each O-D pair $rs$. For each $rs$, the first order necessary conditions of the lower-level model (11.33)-(11.38) include

$$\frac{\partial \mathcal{H}_1}{\partial f_m^{rs}(t)} = \mathcal{U}_m^{rs}(t; k) + \phi_m^{rs}(t) \geq 0, \qquad \forall m, r, s, k, t \in [T_k, T_{k+1}], \qquad (11.45)$$

$$\text{and} \qquad f_m^{rs}(t) \frac{\partial \mathcal{H}_1}{\partial f_m^{rs}(t)} = 0 \qquad \forall m, r, s, k, t \in [T_k, T_{k+1}]; \qquad (11.46)$$

$$\frac{d\phi_m^{rs}(t)}{dt} = -\frac{\partial \mathcal{H}_1}{\partial F_m^{rs}(t)} = 0 \qquad \forall m, r, s, k, t \in [T_k, T_{k+1}]; \qquad (11.47)$$

$$f_m^{rs}(t) \geq 0, \qquad \forall m, r, s, k, t \in [T_k, T_{k+1}]. \qquad (11.48)$$

An alternative representation of bilevel program (11.17)-(11.38) for motorists can be given by converting it into a standard optimization program. As suggested by Cruz (1978) and Bard (1984), this can be achieved by appending the optimality conditions of the departure time choice model (lower problem) to the constraint set of the route choice model (upper problem). The solution to the resulting single level program would also be a solution to the original bilevel departure time/route choice program. Then the *equivalent single level program* for motorists is reformulated as

| | |
|---|---|
| Min | (11.17) |
| s.t. | (11.18)-(11.32) (Upper level model constraints) |
| | (11.34)-(11.38) (Lower level model constraints) |
| | (11.45)-(11.48) (Lower level model optimality conditions) |

This conversion is used only to analyze the optimality conditions of the bilevel program for motorists. From an algorithmic point of view, the model is still solved as a bilevel program.

Next, we discuss the optimality conditions of the lower level departure time choice model. The costate equation (11.47) can be integrated as

$$\phi_m^{rs}(t) = A_m^{rs}(k) \qquad \forall m, r, s, k; \tag{11.49}$$

where $A_m^{rs}(k)$ is an integral constant for interval $k$, and this equation applies for any time $t \in [T_k, T_{k+1}]$. Denote

$$\mu_m^{rs}(k) = -A_m^{rs}(k)$$

which is an input to the first-stage mode choice model. By definition, we have

$$\phi_m^{rs}(t) = -\mu_m^{rs}(k) \qquad \forall m, r, s, k. \tag{11.50}$$

Substituting equation (11.50) into equations (11.45)-(11.46), we obtain the following equations.

$$f_m^{rs}(t)\left\{\mathcal{U}_m^{rs}(t;k) - \mu_m^{rs}(k)\right\} = 0, \qquad \forall m, r, s, k, t \in [T_k, T_{k+1}]; \tag{11.51}$$

$$\mathcal{U}_m^{rs}(t;k) \geq \mu_m^{rs}(k), \qquad \forall m, r, s, k, t \in [T_k, T_{k+1}]; \tag{11.52}$$

$$f_m^{rs}(t) \geq 0, \qquad \forall m, r, s, k, t \in [T_k, T_{k+1}]. \tag{11.53}$$

The above conditions (11.51)-(11.53) hold for any group $m$ of travelers and each O-D pair $rs$ during any interval $k$. In the following, we discuss the above equations (11.51)-(11.53) for each specific group $m$ of travelers and each O-D pair $(r, s)$ during any interval $k$. For any O-D pair $(r, s)$, if there are group $m$ vehicles departing at time $t \in [T_k, T_{k+1}]$, then $f_m^{rs}(t)$ will be positive, so the quantities in braces in equation (11.51) will be zero, i.e., equation (11.52) will hold as an equality. Thus, the drivers which depart at time $t \in [T_k, T_{k+1}]$ have disutility equal to $\mu_m^{rs}(k)$. Equation (11.52) states that at the optimal solution, the $rs$ disutility $\mu_m^{rs}(k)$ is less than or equal to the disutility for departures at any time $t \in [T_k, T_{k+1}]$. Therefore, the disutility for departing at time $t \in [T_k, T_{k+1}]$ equals the minimal disutility for origin-destination $(r, s)$ at any time $t \in [T_k, T_{k+1}]$. For any time $t \in [T_k, T_{k+1}]$, if there are no vehicles in group $m$ departing origin $r$, then the departure rate $f_m^{rs}(t)$ equals zero, so that (11.52) may hold as a strict inequality. Thus, the disutility $\mathcal{U}_m^{rs}(t; k)$ at any time $t \in [T_k, T_{k+1}]$ will not be less than the minimal disutility $\mu_m^{rs}(k)$. The above interpretation implies that the optimality conditions of the lower-level model are consistent with the DUO departure time choices for motorists.

Since the optimality conditions for the lower level departure time choice model are one part of the constraints for the equivalent single level model for motorists, the above results also apply to this problem. Thus, the equivalent single level model for motorists will generate O-D departure flows which satisfy the DUO departure time choice conditions. Note that constraints (11.24) still apply to the above equivalent single level model for motorists. Those two constraints guarantee that the bilevel optimal control model for motorists generates traffic flows satisfying the ideal DUO route choice conditions, given any O-D departure flows determined by the revised constraint set of the single level model for motorists.

We have shown that the set of departure flows and link flows that solves the equivalent single level model for motorists satisfy both the DUO departure time choice conditions and the ideal DUO route choice conditions. Therefore, the solution to the original second-stage departure time/route choice model satisfies both the DUO departure time choice conditions and the ideal DUO route choice conditions.

In summary, the optimality conditions of the two-stage programs (11.14)-(11.38) state the DUO mode, departure time and route choice properties.

## 11.6 Notes

We have presented a combined model in which dynamic user-optimal mode, departure time and route choice occurs. It is recognized that different travelers perceive the time-cost tradeoff differently, and thus distinct groups of travelers are included in the model. The advantage over sequential dynamic models is that the time-dependent interaction among the mode choice, departure time choice and route choice is inherently recognized.

The most likely application of the model which we outlined is in real-time ATIS in conjunction with APTS. Since time-dependent road pricing policies can be easily adopted in our model, an extended combined model incorporating time-dependent road pricing (especially congestion pricing) will find important application in an ATMS. The calibration of the mode choice model requires the determination of the parameter $\theta_m$ which plays a key role in determining the modal shares on the road and transit networks. In contrast to the static mode choice model, the disutilities for both motorists and transit users are time-dependent. Thus, the estimation procedure for $\theta_m$ is more complicated and remains a major task in the future.

The model presented in this chapter can be extended to include more than two modes. In particular, bus lanes and HOV (High Occupancy Vehicle) lanes can be specified as a special mode in road networks. The modal choice model discussed in the chapter assumes that no interactions exist between the transit links and the auto links. However, this assumption is not generally applicable to bus transit. Buses move with road traffic and experience the same time-dependent congestion and delays as automobiles. The interaction of the two interdependent modes can be studied using the multi-group variational inequality models for route choice problems in Chapters 12 and 13.

Generally, a two-stage programming model is very hard to solve. For this combined model, since the mode choice program will result in a logit type mode choice function for transit users and motorists, the main difficulty lies in how to solve the departure time choice and route choice programs efficiently.

# Chapter 12

# Variational Inequality Models of Instantaneous Dynamic User-Optimal Route Choice Problems

In this chapter, we present several variational inequality (VI) models for instantaneous dynamic user-optimal route choice problems for a network with multiple origin-destination pairs. In Section 12.1, a route-time-based VI model is first proposed. The equivalence of the VI model with the route-time-based instantaneous DUO route choice conditions is demonstrated. In order to generalize this route-based model, travelers are stratified into several groups and a multi-group route-cost-based VI model is developed in Section 12.2.

Since explicit route enumeration is needed to solve these route-based VI models, we also formulate two types of link-based VI models. In Section 12.3, a link-time-based VI model is proposed and the equivalence of the VI model with the travel-time-based instantaneous DUO route choice conditions is demonstrated. The multi-group link-cost-based VI model is presented in Section 12.4. In Section 12.5, we discuss the relationships between VI models and optimization models. As an example, we demonstrate that the link-time-based VI model presented in Section 12.3 can be reduced to an optimal control model similar to those presented in Chapter 5. Thus, the diagonalization algorithm and the F-W technique in Chapter 6 can be used to solve this link-time-based VI model. In the solution, we note that explicit route enumeration is unnecessary.

## 12.1 A Route-Time-Based VI Model of Instantaneous Route Choice

### 12.1.1 Route-Time-Based Conditions

Recall from Chapter 4 that the instantaneous travel time $c_a(t)$ over link $a$ is assumed to be dependent on the number of vehicles $x_a(t)$, the inflow $u_a(t)$ and the exit flow $v_a(t)$ on link $a$ at time $t$. This instantaneous link time is the travel time that would be incurred if traffic conditions on the link remain unchanged while traversing the link. We assume the instantaneous travel time $c_a(t)$ on link $a$ is the sum of two components: 1) an instantaneous flow-dependent running time $g_{1a}[x_a(t), u_a(t)]$ over link $a$; and 2) an instantaneous queuing delay $g_{2a}[x_a(t), v_a(t)]$. It follows that

$$c_a(t) = g_{1a}[x_a(t), u_a(t)] + g_{2a}[x_a(t), v_a(t)]. \tag{12.1}$$

The two components $g_{1a}[x_a(t), u_a(t)]$ and $g_{2a}[x_a(t), v_a(t)]$ of the time-dependent link travel time function $c_a[x_a(t), u_a(t), v_a(t)]$ are assumed to be nonnegative and differentiable with respect to $x_a(t), u_a(t)$ and $x_a(t), v_a(t)$, respectively.

Consider the flow which originates at node $r$ at time $t$ and is destined for node $s$. There is a set of routes $\{p\}$ between O-D pair $rs$. Define the instantaneous travel time function $\psi_p^{rs}(t)$ for each route $p$ between $rs$ as

$$\psi_p^{rs}(t) = \sum_{a \in rsp} c_a[x_a(t), u_a(t), v_a(t)] \qquad \forall r, s, p; \tag{12.2}$$

where the summation is over all links $a$ in route $p$ from origin $r$ to destination $s$. Denote $f_p^{rs}(t)$ as the route inflow from origin $r$ to destination $s$ over route $p$ at time $t$. We define route $p$ from origin $r$ to destination $s$ as being used at time $t$ if $f_p^{rs}(t) > 0$. Then, we recall the definition of the route-time-based instantaneous dynamic user-optimal (DUO) state as follows.

> **Route-Time-Based Instantaneous DUO State:** *If, for each O-D pair at each instant of time, the instantaneous travel times for all routes that are being used equal the minimal instantaneous route travel time, the dynamic traffic flow over the network is in a route-time-based instantaneous dynamic user-optimal state.*

We now write the route-time-based instantaneous DUO route choice conditions which are equivalent to the above definition. Denote $\sigma^{rs}(t)$ as the minimal instantaneous route travel time from origin $r$ to destination $s$ at time $t$. The equivalent instantaneous DUO route choice conditions can be stated as follows:

$$\psi_p^{rs^*}(t) - \sigma^{rs^*}(t) \geq 0 \qquad \forall p, r, s; \tag{12.3}$$

$$f_p^{rs^*}(t) \left[\psi_p^{rs^*}(t) - \sigma^{rs^*}(t)\right] = 0 \qquad \forall p, r, s; \tag{12.4}$$

$$f_p^{rs}(t) \geq 0 \qquad \forall p, r, s. \tag{12.5}$$

Equation (12.4) states that if there is a positive route inflow $f_p^{rs^*}(t) > 0$, the instantaneous route travel time $\psi_p^{rs^*}(t)$ must equal the minimal instantaneous route travel time $\sigma^{rs^*}(t)$. Otherwise, the instantaneous route travel time $\psi_p^{rs^*}(t)$ may be greater than or equal to the minimal instantaneous route travel time $\sigma^{rs^*}(t)$.

The above route-time-based definition of an instantaneous DUO state and its corresponding route choice conditions are defined for each O-D pair only. They are not defined for each decision node-destination pair as in Chapter 5 (which is equivalent to the link-time-based instantaneous DUO route choice in Section 12.3). Thus, rerouting strategies are not provided for travelers at intermediate decision nodes or intersections.

### 12.1.2 Dynamic Network Constraints

The constraint set for the route-time-based VI model is summarized as follows.

Relationship between state and control variables:

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \tag{12.6}$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r; s \neq r; \tag{12.7}$$

Flow conservation constraints:

$$f^{rs}(t) = \sum_{a \in A(r)} \sum_p u_{ap}^{rs}(t) \qquad \forall r, s; \tag{12.8}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \tag{12.9}$$

$$\sum_{a \in B(s)} \sum_p v_{ap}^{rs}(t) = e^{rs}(t) \qquad \forall r, s; s \neq r; \tag{12.10}$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[t + \tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; \tag{12.11}$$

Definitional constraints:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \tag{12.12}$$

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \qquad \forall a; \tag{12.13}$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \quad u_{ap}^{rs}(t) \geq 0, \quad v_{ap}^{rs}(t) \geq 0 \quad \forall a, p, r, s; \tag{12.14}$$

$$e_p^{rs}(t) \geq 0, \qquad E_p^{rs}(t) \geq 0, \qquad \forall p, r, s; \tag{12.15}$$

Boundary conditions:

$$E_p^{rs}(0) = 0, \qquad \forall p, r, s; \tag{12.16}$$

$$x_a^{rs}(0) = 0, \qquad \forall a, r, s. \tag{12.17}$$

We note that other constraints, such as FIFO constraints, capacity constraints and oversaturation constraints, can be added to this VI model. To simplify the analysis, we ignore them here.

### 12.1.3 The Route-Time-Based VI Model

The equivalent variational inequality formulation of the route-time-based instantaneous DUO route choice conditions (12.3)-(12.5) may be stated as follows.

*Theorem 12.1.* The dynamic traffic flow pattern satisfying the network constraint set (12.6)-(12.17) is in a route-time-based instantaneous DUO route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \sum_{rs} \sum_p \psi_p^{rs^*}(t) \left[ f_p^{rs}(t) - f_p^{rs^*}(t) \right] dt \geq 0 \tag{12.18}$$

*Proof of Necessity.*

We need to prove that the route-time-based instantaneous DUO route choice conditions (12.3)-(12.5) imply variational inequality (12.18). For any route $p$, a feasible inflow at time $t$ is

$$f_p^{rs}(t) \geq 0. \tag{12.19}$$

Multiplying instantaneous DUO route choice condition (12.3) by the above equation, we have

$$f_p^{rs}(t) \, [\psi_p^{rs^*}(t) - \sigma^{rs^*}(t)] \geq 0 \qquad \forall p, r, s. \tag{12.20}$$

We subtract equation (12.4) from equation (12.20) and obtain

$$\left[ f_p^{rs}(t) - f_p^{rs^*}(t) \right] \left[ \psi_p^{rs^*}(t) - \sigma^{rs^*}(t) \right] \geq 0 \qquad \forall p, r, s. \tag{12.21}$$

Summing equation (12.21) for all routes $p$ and all O-D pairs $rs$, it follows that

$$\begin{aligned}
& \sum_{rs}\sum_{p}\left[f_p^{rs}(t) - f_p^{rs^*}(t)\right]\left[\psi_p^{rs^*}(t) - \sigma^{rs^*}(t)\right] \\
= \ & \sum_{rs}\sum_{p}\left[f_p^{rs}(t) - f_p^{rs^*}(t)\right]\psi_p^{rs^*}(t) - \sum_{rs}\sigma^{rs^*}(t)\sum_{p}\left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] \\
= \ & \sum_{rs}\sum_{p}\left[f_p^{rs}(t) - f_p^{rs^*}(t)\right]\psi_p^{rs^*}(t) \geq 0 \qquad (12.22)
\end{aligned}$$

where the flow conversation equation

$$\sum_{p} f_p^{rs}(t) = \sum_{p} f_p^{rs^*}(t) = f^{rs}(t)$$

holds for each O-D $rs$ at each time $t$. Integrating the above equation from 0 to $T$, we obtain variational inequality (12.18).

*Proof of Sufficiency.*

We need to prove that any solution $f_p^{rs^*}(t)$ to variational inequality (12.18) satisfies the route-time-based instantaneous DUO route choice conditions (12.3)-(12.5). We know that the first and third instantaneous DUO route choice conditions (12.3) and (12.5) hold by definition. Thus, we need to prove that the second instantaneous DUO route choice condition (12.4) also holds.

Assume that the second instantaneous DUO route choice condition (12.4) does not hold only for a route $q$ for O-D pair $kn$ during time interval $[t_1-\delta, t_1+\delta] \in [0,T]$, i.e.,

$$f_q^{kn^*}(t) > 0 \quad \text{and} \quad \psi_q^{kn^*}(t) - \sigma^{kn^*}(t) > 0 \quad \forall\, t \in [t_1-\delta, t_1+\delta] \qquad (12.23)$$

Since the second instantaneous DUO route choice condition (12.4) holds for all other routes other than route $q$ for O-D pair $kn$ at any time $t$ and for O-D pair $kn$ at any time $t \notin [t_1-\delta, t_1+\delta]$, it follows that

$$\begin{aligned}
& \int_0^T \sum_{rs}\sum_{p} f_p^{rs^*}(t)\,[\psi_p^{rs^*}(t) - \sigma^{rs^*}(t)]\,dt \\
= \ & \int_{t_1-\delta}^{t_1+\delta} f_q^{kn^*}(t)\,[\psi_q^{kn^*}(t) - \sigma^{kn^*}(t)]\,dt > 0 \qquad (12.24)
\end{aligned}$$

We note that all other terms in the above equation vanish because of instantaneous DUO route choice condition (12.4).

For each O-D pair $rs$, we can always find one minimal instantaneous travel time route $l$ for vehicles departing origin $r$ at time $t$, where route $l$ was evaluated under the optimal route inflow pattern $\{f_p^{rs^*}(t)\}$. For this route $l$, the first instantaneous DUO route choice condition (12.3) becomes an equality by definition. It follows that

$$\psi_l^{rs^*}(t) - \sigma^{rs^*}(t) = 0 \qquad \forall l, r, s. \tag{12.25}$$

Next, we need to find a set of feasible route inflows $f_p^{rs}(t)$ so that the following equations

$$f_p^{rs}(t) \ \left[\psi_p^{rs^*}(t) - \sigma^{rs^*}(t)\right] = 0 \qquad \forall p, r, s \tag{12.26}$$

always hold. We consider departure flows $f^{rs}(t)$ for all O-D pairs at each time $t$. For each O-D pair $rs$ at each time $t$, we assign O-D departure flow $f^{rs}(t)$ to the minimal travel time route $l$, which was evaluated under the optimal route inflow pattern $\{f_p^{rs^*}(t)\}$. This will generate a set of feasible route inflow patterns $\{f_p^{rs}(t)\}$ which always satisfies equation (12.26) because flows are not assigned to routes with non-minimal travel times which were evaluated under the optimal route inflow pattern $\{f_p^{rs^*}(t)\}$. Summing equations (12.26) for all routes $p$ and all O-D pairs $rs$, it follows that

$$\sum_{rs}\sum_{p} f_p^{rs}(t) \ \left[\psi_p^{rs^*}(t) - \sigma^{rs^*}(t)\right] = 0 \tag{12.27}$$

Integrating the above equation for time period $[0, T]$, we have

$$\int_0^T \sum_{rs}\sum_{p} f_p^{rs}(t) \ \left[\psi_p^{rs^*}(t) - \sigma^{rs^*}(t)\right] \ dt \ = 0 \tag{12.28}$$

We subtract equation (12.24) from equation (12.28) and obtain

$$\begin{aligned}
& \int_0^T \sum_{rs}\sum_{p} \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] \ \left[\psi_p^{rs^*}(t) - \sigma^{rs^*}(t)\right] \ dt \\
= & \int_0^T \left\{\sum_{rs}\sum_{p} \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] \ \psi_p^{rs^*}(t) \right. \\
- & \left. \sum_{rs} \sigma^{rs^*}(t) \sum_{p} \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] \right\} \ dt \\
= & \int_0^T \sum_{rs}\sum_{p} \psi_p^{rs^*}(t) \ \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] \ dt \ < \ 0
\end{aligned} \tag{12.29}$$

where the flow conversation equation

$$\sum_{p} f_p^{rs}(t) = \sum_{p} f_p^{rs^*}(t) = f^{rs}(t)$$

holds for each O-D $rs$ at each time $t$ so that the second term of the equation vanishes. The above equation contradicts variational inequality (12.18). Therefore, any optimal solution $\{f_p^{rs^*}(t)\}$ to variational inequality (12.18) satisfies the second instantaneous DUO route choice condition (12.4). Since we proved the necessity and sufficiency of the equivalence of variational inequality (12.18) to route-time-based instantaneous DUO route choice conditions (12.3)-(12.5), the proof is complete.

## 12.2 A Multi-Group Route-Cost-Based VI Model of Instantaneous Route Choice

In this section, we define an instantaneous dynamic user-optimal (DUO) model based on travel costs or disutilities instead of travel times. To be consistent with Chapter 11, we still stratify travelers into $M$ groups for each O-D pair according to the socio-economic characteristics of each traveler. When $M = 1$, the following definition and VI model reduces to a single group model, but one which is different from the above VI model based on instantaneous travel times.

For multi-group route choice problems, there are several approaches of stratifying travelers into groups. The first approach is to classify travelers based on income and age (see Table 12.2). There are 9 combinations in this approach. The second approach is to classify travelers based on route diversion willingness (see Table 12.2). There are 3 combinations in this approach.

The third approach is to classify travelers based on driving behavior (see Table 12.3), which was proposed by Codelli et al (1993). There are 3 combinations in this approach. The above stratifications can also be combined into more detailed classifications. We leave the subsequent analysis of those combinations for empirical studies.

Table 12.1: Stratification of Travelers Based on Income and Age

| | Degree of Change | | |
|---|---|---|---|
| Income | High | Middle | Low |
| Age | Old | Middle | Young |

Table 12.2: Stratification of Travelers Based on Route Diversion Willingness

| | | |
|---|---|---|
| One Route | Few Alternative Routes (2-3) | En Route Diversion |

Table 12.3: Stratification of Travelers Based on Driving Behavior

| Cautious Driver | Rushed Driver | Ruthless Driver |
|---|---|---|

## 12.2.1 Multi-Group Route-Cost-Based Conditions

The instantaneous route disutility function is dependent on the instantaneous route travel time, fuel consumption enroute, automobile operating cost, etc. Denote $\tilde{c}_{ma}(t)$ as the instantaneous disutility function for travelers of group $m$ entering link $a$ at time $t$. It follows that

$$\tilde{c}_{ma}(t) = \alpha_{ma} + \beta_{ma}\, c_a(t) \qquad \forall m, a \tag{12.30}$$

where $\alpha_{ma}$ is a fixed instantaneous disutility parameter for group $m$ travelers on link $a$ and $\beta_{ma}$ is a parameter to transform instantaneous link travel time $c_a(t)$ into the disutility of group $m$ travelers. Denote $\tilde{\psi}^{rs}_{mp}(t)$ as the instantaneous route travel disutility for group $m$ travelers from origin $r$ to destination $s$ at time $t$, and $\tilde{\sigma}^{rs}_m(t)$ as the minimal instantaneous route travel disutility for group $m$ travelers from origin $r$ to destination $s$ at time $t$. The instantaneous route travel disutility for all allowable routes is computed as follows:

$$\tilde{\psi}^{rs}_{mp}(t) = \sum_{a \in rsp} \tilde{c}_{ma}(t) \qquad \forall m, p, r, s \tag{12.31}$$

where the summation is over all links $a$ on route $p$. The minimal instantaneous route travel disutility for each O-D pair $(r, s)$ is

$$\tilde{\sigma}^{rs}_m(t) = \min_p \tilde{\psi}^{rs}_{mp}(t) \qquad \forall m, r, s \tag{12.32}$$

We then present the definition of the multi-group route-cost-based instantaneous DUO state as follows.

**Multi-Group Route-Cost-Based Instantaneous DUO State:** *If, for each group $m$ and each O-D pair at each instant of time, the instantaneous travel* disutilities *for all routes that are being used equal the minimal instantaneous route travel* disutility, *the dynamic traffic flow over the network is in a multi-group route-cost-based instantaneous dynamic user-optimal state.*

Denote $f^{rs}_{mp}(t)$ as the route inflow of group $m$ from origin $r$ to destination $s$ over route $p$ at time $t$. The equivalent route-based multi-group instantaneous DUO route choice conditions can be summarized as follows:

$$\tilde{\psi}^{rs*}_{mp}(t) - \tilde{\sigma}^{rs*}_m(t) \geq 0 \qquad \forall m, p, r, s; \tag{12.33}$$

$$f_{mp}^{rs*}(t)\left[\tilde{\psi}_{mp}^{rs*}(t) - \tilde{\sigma}_{m}^{rs*}(t)\right] = 0 \qquad \forall m, p, r, s; \tag{12.34}$$

$$f_{mp}^{rs}(t) \geq 0 \qquad \forall m, p, r, s. \tag{12.35}$$

The above definition of the multi-group instantaneous DUO state and the corresponding route-based route choice conditions are defined for each origin-destination pair. They are not defined for each decision node-destination pair as Chapter 5 (which is similar to the link-based definitions in the next sections). Thus, this definition does not provide any rerouting strategies for travelers at any intermediate intersection.

## 12.2.2 Dynamic Network Constraints

The dynamic network constraints are written for each group of travelers, designated by index $m$ The constraint set for this problem is summarized as follows.

Relationship between state and control variables:

$$\frac{dx_{map}^{rs}}{dt} = u_{map}^{rs}(t) - v_{map}^{rs}(t) \qquad \forall m, a, p, r, s; \tag{12.36}$$

$$\frac{dE_{mp}^{rs}(t)}{dt} = e_{mp}^{rs}(t) \qquad \forall m, p, r; s \neq r; \tag{12.37}$$

Flow conservation constraints:

$$f_{m}^{rs}(t) = \sum_{a\in A(r)} \sum_{p} u_{map}^{rs}(t) \qquad \forall m, r, s; \tag{12.38}$$

$$\sum_{a\in B(j)} v_{map}^{rs}(t) = \sum_{a\in A(j)} u_{map}^{rs}(t) \qquad \forall j, m, p, r, s; j \neq r, s; \tag{12.39}$$

$$\sum_{a\in B(s)} \sum_{p} v_{map}^{rs}(t) = e_{m}^{rs}(t) \qquad \forall m, r, s; s \neq r; \tag{12.40}$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b\in\tilde{p}} \{x_{bp}^{rs}[t + \tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_{p}^{rs}[t + \tau_a(t)] - E_{p}^{rs}(t)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; \tag{12.41}$$

Definitional constraints:

$$\sum_{mrsp} u_{map}^{rs}(t) = u_a(t), \qquad \sum_{mrsp} v_{map}^{rs}(t) = v_a(t), \qquad \forall a; \tag{12.42}$$

$$\sum_{mrsp} x_{map}^{rs}(t) = x_a(t), \qquad \forall a; \tag{12.43}$$

Nonnegativity conditions:

$$x_{map}^{rs}(t) \geq 0, \quad u_{map}^{rs}(t) \geq 0, \quad v_{map}^{rs}(t) \geq 0 \quad \forall m, a, p, r, s; \tag{12.44}$$

$$e_{mp}^{rs}(t) \geq 0, \quad E_{mp}^{rs}(t) \geq 0, \quad \forall m, p, r, s; \tag{12.45}$$

Boundary conditions:

$$E_{mp}^{rs}(0) = 0, \qquad \forall m, p, r, s; \tag{12.46}$$

$$x_{map}^{rs}(0) = 0, \qquad \forall m, a, p, r, s. \tag{12.47}$$

### 12.2.3 The Multi-Group Route-Cost-Based VI Model

Then, the equivalent variational inequality formulation of multi-group route-cost-based instantaneous DUO route choice conditions (12.33)-(12.35) may be stated as follows.

*Theorem 12.2.* The dynamic traffic flow pattern satisfying network constraint set (12.36)-(12.47) is in a multi-group route-cost-based instantaneous DUO route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \sum_{rs} \sum_{mp} \tilde{\psi}_{mp}^{rs*}(t) \left[ f_{mp}^{rs}(t) - f_{mp}^{rs*}(t) \right] dt \geq 0 \tag{12.48}$$

The proofs of necessity and sufficiency for variational inequality (12.48) follow in the same manner as in Section 12.1.3 for the single group route-time-based case.

## 12.3 A Link-Time-Based VI Model of Instantaneous Route Choice

Both VI models presented above are route-based; their solution requires explicit route enumeration. Although the route-based model is intuitive in terms of understanding, route enumeration is a great burden if the network is large, which is termed the curse of dimensionality in optimal control theory. Figure 12.1 shows a $5 \times 5$ one-way square grid network with $N = 25$ nodes and $L = 36$ links. The total number of routes from node 1 to node 25 is 64.

Table 12.4 illustrates the increase of links and routes with the increase of nodes in such a grid network. Basically, the number of links increases linearly with the increase of nodes. However, the number of routes increases exponentially with the increase of nodes. For example, when there are only 100 nodes, the number of routes is well over 20,000. When there are 400 nodes, the number of routes is over $10^9$. We note that the routes in these one-way grid networks

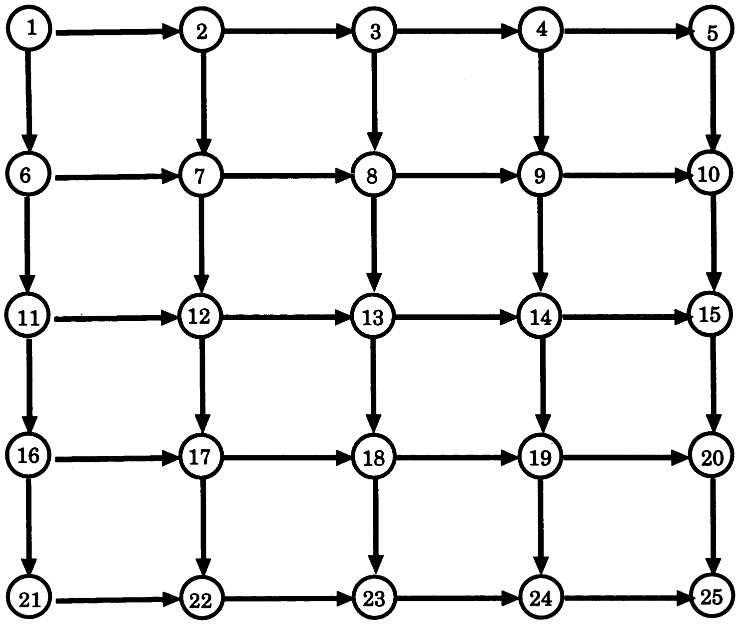



Figure 12.1: Example Grid Network

are efficient routes in the sense of the definition of Dial (1971); i.e., any link on the route takes the vehicles further away from the origin and closer to the destination. From Table 12.4, we can conclude that explicit route enumeration is infeasible for large networks.

Table 12.4: Number of Nodes, Links and Routes

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Number of Nodes $N$ | 4 | 9 | 16 | 25 | 36 | 49 | 64 | 100 | 400 |
| Number of Links $L$ | 4 | 12 | 24 | 40 | 60 | 84 | 120 | 180 | 760 |
| Number of Routes | 2 | 6 | 20 | 64 | 202 | 660 | 2212 | $> 20,000$ | $> 10^9$ |

To distinguish the link-based model from the route-based model, we can focus on the variational inequality instead of their constraints because both constraints are route-based. If the variational inequality can be formulated using link-based variables instead of route-based variables, a carefully designed solution algorithm for such a variational inequality will not require explicit route enumeration. We will demonstrate this point later. This is also true for optimization problems. If the objective function of an optimization program can be formulated using link-based variables instead of route-based variables, a carefully designed solution algorithm will not need explicit route enumeration. Recall that in the mathematical programming formulation for the static user-optimal route choice model, the objective function can be formulated using either link-based variables or route-based variables. However, the link-based objective function is preferred because for such a link-based objective function, the Frank-Wolfe algorithm does not need explicit route enumeration. This property is also true for the optimal control models in this book.

Because dynamic traffic flow does not have a constant flow rate during propagation over links and routes, route-based VI models can not be transformed into link-based VI models. Thus, it is very difficult to develop a solution algorithm for a route-based VI without explicit route enumeration. This issue is the most critical constraint for applying route-based VI to realistic transportation networks. Therefore, we propose a link-based VI which overcomes this problem. In addition to this contribution, we note that our formulation approach is different from others (Smith, 1993; Friesz et al, 1993).

The set of dynamic network constraints for the link-time-based VI model is identical to constraint set (12.6)-(12.17) of the route-time-based VI model in Section 12.1. The basic difference between the two models is that the variational inequality of the link-time-based VI model is formulated using link-based flow variables instead of route-based variables as in the route-time-based VI model.

## 12.3.1 Link-Time-Based Conditions

In contrast to the previous sections, we now introduce a new set of instantaneous DUO route choice conditions based on link and node variables, instead of route-based variables. In this problem, link $a$ is defined as *being used* at time $t$ if $u_a(t) > 0$. A route $p$ from decision node $i$ to destination $s$ is defined as *being used* at time $t$ if $u^s_{ap}(t) > 0$ for the first link $a \in p$. In the following, we define an instantaneous DUO route choice state based on link and node variables.

> **Link-Time-Based Instantaneous DUO State:** *If, for any departure flow from each decision node to each destination node at each instant of time, the instantaneous travel times equal the minimal instantaneous route travel time, the dynamic traffic flow over the network is in a link-time-based instantaneous dynamic user-optimal state.*

The above definition of an instantaneous DUO state is identical to that in Chapter 5. Note that the route-time-based instantaneous DUO state defined in Section 12.1 is a subset of the link-time-based instantaneous state because the decision node includes the origin so that each O-D pair is one decision node-destination pair.

Define $\sigma^{is^*}(t)$ as the minimal instantaneous travel time for vehicles departing from node $i$ to destination $s$ at time $t$, the asterisk denoting that the travel time is computed using link-time-based instantaneous DUO traffic flows.

For link $a = (i, j)$, the minimal instantaneous travel time $\sigma^{is^*}(t)$ from node $i$ to destination $s$ should be equal to or less than the minimal instantaneous travel time $\sigma^{js^*}(t)$ from node $j$ to destination $s$ plus the instantaneous link travel time $c_a^*(t)$ at time instant $t$. It follows that

$$\sigma^{js^*}(t) + c_a^*(t) \geq \sigma^{is^*}(t) \qquad \forall a = (i, j), s. \tag{12.49}$$

If any departure flow from node $i$ to destination $s$ enters link $a$ at time $t$, or $u_a^s(t) > 0$, then the link-time-based instantaneous DUO route choice conditions require that link $a$ is on the minimal instantaneous travel time route. In other words, the instantaneous minimal travel time $\sigma^{is^*}(t)$ for vehicles departing node $i$ toward destination $s$ at time $t$ should equal the minimal instantaneous travel time $\sigma^{js^*}(t)$ from node $j$ to destination $s$ plus the instantaneous link travel time $c_a^*(t)$ at time instant $t$. It follows that

$$\sigma^{js^*}(t) + c_a^*(t) = \sigma^{is^*}(t), \qquad \text{if } u_a^{s^*}(t) > 0 \qquad \forall a = (i, j), s. \tag{12.50}$$

The above equations are also equivalent to the following:

$$\left[\sigma^{js^*}(t) + c_a^*(t) - \sigma^{is^*}(t)\right] \, u_a^{s^*}(t) = 0 \qquad \forall a = (i, j), s. \tag{12.51}$$

Denote $\Theta_a^{s^*}(t)$ as the difference between the minimal instantaneous travel time from node $j$ to destination $s$ and the instantaneous travel time from node $i$ to destination $s$ plus the instantaneous travel time on link $a$ at time $t$. It follows that

$$\Theta_a^{s^*}(t) = \sigma^{js^*}(t) + c_a^*(t) - \sigma^{is^*}(t) \qquad \forall a, s; a = (i, j). \tag{12.52}$$

Thus, the link-time-based instantaneous DUO route choice conditions can be summarized as follows:

$$\Theta_a^{s^*}(t) \geq 0 \qquad \forall a = (i, j), s; \tag{12.53}$$

$$u_a^{s^*}(t) \, \Theta_a^{s^*}(t) = 0 \qquad \forall a = (i, j), s; \tag{12.54}$$

$$u_a^s(t) \geq 0 \qquad \forall a = (i, j), s. \tag{12.55}$$

Now, we state a lemma concerning the relationship between the link-time-based and the route-time-based instantaneous DUO route choice conditions.

*Lemma 12.1.* Link-time-based instantaneous DUO route choice conditions (12.53)-(12.55) imply route-time-based instantaneous DUO route choice conditions (12.3)-(12.5).

*Proof.*

If we consider node $i$ as an origin $r$, then link-time-based instantaneous DUO route choice conditions (12.53)-(12.55) apply to O-D pair $rs$. Thus, equations (12.53)-(12.55) can be written as

$$\sigma^{js^*}(t) + c_a^*(t) - \sigma^{rs^*}(t) \geq 0 \qquad \forall a = (r,j), r, s; \tag{12.56}$$

$$u_a^{rs^*}(t) \left[\sigma^{js^*}(t) + c_a^*(t) - \sigma^{rs^*}(t)\right] = 0 \qquad \forall a = (r,j), r, s; \tag{12.57}$$

$$u_a^{rs}(t) \geq 0 \qquad \forall a = (r,j), r, s \tag{12.58}$$

Suppose there are $P$ routes from origin $r$ to destination $s$ via link $a$. Since equation (12.74) applies to all links $a$ exiting origin $r$, we can define an instantaneous travel time on route $p$ as $\psi_p^{rs^*}(t)$. If link $a$ is on route $p$, this instantaneous route travel time $\psi_p^{rs^*}(t)$ must be greater than or equal to the minimal instantaneous travel time from node $j$ to destination $s$ plus the instantaneous travel time on link $a$ at time $t$. It follows that

$$\psi_p^{rs^*}(t) \geq \sigma^{js^*}(t) + c_a^*(t) \tag{12.59}$$

By equation (12.56), we have

$$\psi_p^{rs^*}(t) \geq \sigma^{rs^*}(t) \tag{12.60}$$

which is identical to equation (12.3). Furthermore, if there is a flow on link $a$, i.e., $u_a^{rs^*}(t) > 0$, then by equations (12.56)-(12.57), we have

$$\sigma^{js^*}(t) + c_a^*(t) - \sigma^{rs^*}(t) = 0 \tag{12.61}$$

Specify $p$ as the route via link $a$ and the minimal instantaneous travel time subroute from node $j$ to destination $s$. Thus, the flow $f_p^{rs^*}(t)$ on route $p$ is positive at time $t$ and we have

$$\psi_p^{rs^*}(t) = \sigma^{js^*}(t) + c_a^*(t) \tag{12.62}$$

In other words, we have

$$f_p^{rs^*}(t) \left[\psi_p^{rs^*}(t) - \sigma^{rs^*}(t)\right] = 0 \tag{12.63}$$

For any positive link inflow $u_a^{rs^*}(t) > 0$, we can generate a corresponding positive route inflow $f_p^{rs^*}(t)$ and the corresponding equation (12.63). Thus, equation (12.63) applies to any positive route inflow $f_p^{rs^*}(t)$. We conclude that it applies to any route between any O-D pair $rs$. Note that the above equation applies to any zero route inflow $f_p^{rs^*}(t) = 0$ as well. Therefore, equations (12.60) and (12.63) imply the route-time-based instantaneous DUO route choice conditions (12.3)-(12.5). The proof is complete.

### 12.3.2 The Link-Time-Based VI Model

The equivalent variational inequality formulation of link-time-based instantaneous DUO route choice conditions (12.53)-(12.55) may be stated as follows.

*Theorem 12.3.* The dynamic traffic flow pattern satisfying network constraint set (12.6)-(12.17) is in a link-time-based instantaneous DUO route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \sum_s \sum_a \Theta_a^{s^*}(t) \ \left[u_a^s(t) - u_a^{s^*}(t)\right] \ dt \ \geq \ 0 \tag{12.64}$$

*Proof of Necessity.*

We need to prove that link-time-based instantaneous DUO route choice conditions (12.53)-(12.55) imply variational inequality (12.64). For any link $a$, a feasible inflow at time $t$ is

$$u_a^s(t) \geq 0. \qquad \forall a = (i,j), s. \tag{12.65}$$

Multiplying equation (12.65) and equation (12.53) we have

$$u_a^s(t) \ \Theta_a^{s^*}(t) \ \geq 0 \qquad \forall a, s; a = (i,j). \tag{12.66}$$

We subtract equation (12.54) from equation (12.66) and obtain

$$\left[u_a^s(t) - u_a^{s^*}(t)\right] \ \Theta_a^{s^*}(t) \ \geq 0 \qquad \forall a, s; a = (i,j). \tag{12.67}$$

Summing equation (12.67) for all links $a$ and all destinations $s$, it follows that

$$\sum_s \sum_a \left[u_a^s(t) - u_a^{s^*}(t)\right] \ \Theta_a^{s^*}(t) \ \geq 0 \qquad \text{where } a = (i,j). \tag{12.68}$$

Integrating the above equation from 0 to $T$, we obtain variational inequality (12.64).

*Proof of Sufficiency.*

We need to prove that any solution $u_a^{s^*}(t)$ to variational inequality (12.64) satisfies link-time-based instantaneous DUO route choice conditions (12.53)-(12.55). We know that the first and third instantaneous DUO route choice conditions (12.53) and (12.55) hold by definition. Thus, we need to prove that the second instantaneous DUO route choice condition (12.54) also holds.

Assume that the second instantaneous DUO route choice condition (12.54) does not hold only for a link $b = (l, m)$ for a destination $n$ during a time interval $[d-\delta, d+\delta] \in [0, T]$, i.e.,

$$u_b^{n^*}(t) > 0 \quad \text{and} \quad \Theta_b^{n^*}(t) > 0 \quad t \in [d-\delta, d+\delta] \tag{12.69}$$

Thus, we have

$$u_b^{n^*}(t)\ \Theta_b^{n^*}(t) > 0 \tag{12.70}$$

where

$$\Theta_b^{n^*}(t) = \sigma^{mn^*}(t) + c_b^*(t) - \sigma^{ln^*}(t) > 0 \qquad \text{where } b = (l, m). \tag{12.71}$$

Note that the second instantaneous DUO route choice condition (12.54) holds for all links other than link $b = (l, m)$ for destination $n$ at time $t$. Equation (12.54) also holds for link $b = (l, m)$ for destinations $s \neq n$ at time $t$ and for link $b = (l, m)$ for destinations $n$ at time $t \notin [d-\delta, d+\delta]$. It follows that

$$\int_0^T \sum_s \sum_a \Theta_a^{s^*}(t)\ u_a^{s^*}(t)\ dt \ = \ \int_{d-\delta}^{d+\delta} u_b^{n^*}(t)\ \ \Theta_b^{n^*}(t)\ dt\ > 0 \tag{12.72}$$

We note that all other terms in the above equation vanish because of instantaneous DUO route choice condition (12.54).

For each O-D pair $rs$, we can always find one minimal travel time route $p$ for vehicles departing origin $r$ at time $t$, where route $p$ was evaluated under optimal flow pattern $\{u_a^{s^*}(t)\}$. For each link $a$ on this route $p$, the first instantaneous DUO route choice condition (12.53) becomes an equality by definition. It follows that

$$\Theta_a^{s^*}(t) = \sigma^{js^*}(t) + c_a^*(t) - \sigma^{is^*}(t) = 0 \qquad \forall a, s; a = (i, j); a \in p. \tag{12.73}$$

Next, we need to find a set of feasible inflows $u_a^s(t)$ so that the following equations

$$u_a^s(t)\ \Theta_a^{s^*}(t) = 0 \qquad \forall a, s; a = (i, j) \tag{12.74}$$

always hold. We adjust all the departure flows $f^{rs}(t)$ for all O-D pairs at time $t$. For each O-D pair $rs$ at each time $t$, we assign O-D departure flow $f^{rs}(t)$ to the minimal travel time route $p$, which was evaluated under the optimal flow pattern $\{u_a^{s^*}(t)\}$. This will generate a set of feasible inflow patterns $\{u_a^s(t)\} = \sum_r u_a^{rs}(t)$ which always satisfies equations (12.73) and (12.74) because flows are not assigned to routes with non-minimal instantaneous travel times which were evaluated under the optimal flow pattern $\{u_a^{s^*}(t)\}$. Summing equations (12.74) for all links $a$ and all destinations $s$, it follows that

$$\sum_s \sum_a u_a^s(t)\ \Theta_a^{s^*}(t) = 0 \qquad \text{where } a = (i, j). \tag{12.75}$$

Integrating the above equation (12.75) for interval $[0,T]$, we have

$$\int_0^T \sum_s \sum_a u_a^s(t)\, \Theta_a^{s^*}(t)\, dt \;= 0 \tag{12.76}$$

We subtract equation (12.72) from equation (12.76) and obtain

$$\int_0^T \sum_s \sum_a \Theta_a^{s^*}(t)\; \left[u_a^s(t) - u_a^{s^*}(t)\right]\; dt \;< 0 \tag{12.77}$$

The above equation contradicts variational inequality (12.64). Therefore, any optimal solution $\{u_a^{s^*}(t)\}$ to variational inequality (12.64) satisfies the second instantaneous DUO route choice condition (12.54). Since we proved the necessity and sufficiency of the equivalence of variational inequality (12.64) to link-time-based instantaneous DUO route choice conditions (12.53)-(12.55), the proof is complete.

## 12.4 A Multi-Group Link-Cost-Based VI Model of Instantaneous Route Choice

In this section, we consider a link-based multi-group VI model for the instantaneous dynamic user-optimal problem. The dynamic network constraints have to be written for each group of travelers designated by index $m$. The constraint set for this problem is identical to the constraint set (12.36)-(12.47) for the route-based multi-group VI model in Section 12.2. As before, the basic difference between the two models is that the variational inequality of the multi-group link-cost-based model is formulated using link-based flow variables instead of route-based variables as in the multi-group route-cost-based VI model.

### 12.4.1 Multi-Group Link-Cost-Based Conditions

The instantaneous route disutility function is dependent on the instantaneous route travel time, fuel consumption enroute and automobile operating cost, etc. Denote $\tilde{c}_{ma}(t)$ as the instantaneous disutility function for group $m$ travelers entering link $a$ at time $t$. It follows that

$$\tilde{c}_{ma}(t) = \alpha_{ma} + \beta_{ma}\; c_a(t) \qquad \forall m, a \tag{12.78}$$

where $\alpha_{ma}$ is a fixed instantaneous disutility parameter for group $m$ travelers on link $a$ and $\beta_{ma}$ is a parameter to transform instantaneous link travel time $c_a(t)$ into the disutility of group $m$ travelers. Denote $\tilde{\psi}_{mp}^{rs}(t)$ as the instantaneous route travel disutility for group $m$ travelers from origin $r$ to destination $s$ at time $t$. The instantaneous route travel disutility for all allowable routes $p$ between decision node $i$ and destination $s$ is computed as follows:

$$\tilde{\psi}_{mp}^{is}(t) = \sum_{a \in isp} \tilde{c}_{ma}(t) \qquad \forall m, p, i, s \tag{12.79}$$

where the summation is over all links $a$ on route $p$. We then define a multi-group link-cost-based instantaneous DUO state as follows.

**Multi-Group Link-Cost-Based Instantaneous DUO State:** *If, for any departure flow of group m from each decision node to each destination node at each instant of time, the instantaneous travel* disutilities *equal the minimal instantaneous route travel* disutility, *the dynamic traffic flow over the network is in a multi-group link-cost-based instantaneous dynamic user-optimal state.*

Define $\tilde{\sigma}_m^{is}(t)$ as the minimal instantaneous route travel disutility for group $m$ travelers from node $i$ to destination $s$ at time $t$. The asterisk denotes that the travel disutility is computed using multi-group link-cost-based instantaneous DUO traffic flows. For group $m$ travelers on link $a = (i, j)$, the minimal instantaneous travel disutility $\tilde{\sigma}_m^{is^*}(t)$ from node $i$ to destination $s$ should be equal to or less than the minimal instantaneous travel disutility $\tilde{\sigma}_m^{js^*}(t)$ from node $j$ to destination $s$ plus the instantaneous link travel disutility $\tilde{c}_{ma}^*(t)$ at time instant $t$. It follows that

$$\tilde{\sigma}_m^{js^*}(t) + \tilde{c}_{ma}^*(t) \geq \tilde{\sigma}_m^{is^*}(t) \qquad \forall m, a = (i, j), s. \qquad (12.80)$$

If any departure flow of group $m$ from node $i$ to destination $s$ enters link $a$ at time $t$, or $u_{ma}^s(t) > 0$, then the multi-group link-cost-based instantaneous DUO route choice conditions require that link $a$ is on the minimal instantaneous travel disutility route. In other words, the instantaneous minimal travel disutility $\tilde{\sigma}_m^{is^*}(t)$ for vehicles departing node $i$ toward destination $s$ at time $t$ should equal the minimal instantaneous travel disutility $\tilde{\sigma}_m^{js^*}(t)$ for vehicles departing node $j$ to destination $s$ plus the instantaneous link travel disutility $\tilde{c}_{ma}^*(t)$ at time instant $t$. It follows that

$$\tilde{\sigma}_m^{js^*}(t) + \tilde{c}_{ma}^*(t) = \tilde{\sigma}_m^{is^*}(t), \quad \text{if } u_{ma}^{s^*}(t) > 0 \quad \forall m, a = (i, j), s. \qquad (12.81)$$

The above equations are also equivalent to the following:

$$\left[\tilde{\sigma}_m^{js^*}(t) + \tilde{c}_{ma}^*(t) - \tilde{\sigma}_m^{is^*}(t)\right] u_{ma}^{s^*}(t) = 0 \qquad \forall m, a = (i, j), s. \qquad (12.82)$$

Denote $\tilde{\Theta}_{ma}^{s^*}(t)$ as the difference between the minimal instantaneous travel disutility from node $j$ to destination $s$ and the instantaneous travel disutility from node $i$ to destination $s$ plus the instantaneous travel disutility on link $a$ for group $m$ at time $t$. It follows that

$$\tilde{\Theta}_{ma}^{s^*}(t) = \tilde{\sigma}_m^{js^*}(t) + \tilde{c}_{ma}^*(t) - \tilde{\sigma}_m^{is^*}(t) \qquad \forall m, a, s; a = (i, j). \qquad (12.83)$$

Thus, the multi-group link-cost-based instantaneous DUO route choice conditions can be summarized as follows:

$$\tilde{\Theta}_{ma}^{s^*}(t) \geq 0 \qquad \forall m, a = (i, j), s; \qquad (12.84)$$

$$u_{ma}^{s^*}(t)\, \tilde{\Theta}_{ma}^{s^*}(t) = 0 \qquad \forall m, a = (i, j), s; \tag{12.85}$$

$$u_{ma}^{s}(t) \geq 0 \qquad \forall m, a = (i, j), s. \tag{12.86}$$

Similar to Lemma 12.1, we can prove that multi-group link-cost-based instantaneous DUO route choice conditions (12.84)-(12.86) imply route-based multi-group instantaneous DUO route choice conditions (12.33)-(12.35).

### 12.4.2 The Multi-Group Link-Cost-Based VI Model

The equivalent variational inequality formulation of multi-group link-cost-based instantaneous DUO route choice conditions (12.84)-(12.86) may be stated as follows.

> *Theorem 12.4.* The dynamic traffic flow pattern satisfying network constraint set (12.36)-(12.47) is in a multi-group link-cost-based instantaneous DUO route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \sum_s \sum_{ma} \tilde{\Theta}_{ma}^{s^*}(t) \left[u_{ma}^{s}(t) - u_{ma}^{s^*}(t)\right] dt \geq 0 \tag{12.87}$$

The proofs of necessity and sufficiency follow in the same way as for the single group case.

## 12.5 Relationships Between VI Models and Optimization Models

As illustrated in Chapter 3, VI models can be reformulated as optimization models under certain symmetry conditions. We show in this section that the VI model can be reformulated as an optimal control problem which is identical to the optimal control models with similar constraints presented in Chapter 5. We will not discuss each VI model in this chapter, but focus our analysis on the link-time-based VI model for the instantaneous DUO route choice problem of Section 12.3. Similar analyses can be performed for other VI models for various instantaneous DUO route choice problems.

Consider the following VI problem from Section 12.3:

$$\int_0^T \sum_s \sum_a \Theta_a^{s^*}(t) \left[u_a^{s}(t) - u_a^{s^*}(t)\right] dt \geq 0 \tag{12.88}$$

In order to present a partitionable VI, we need to transform the original VI into a partitionable VI using some new definitions as follows. Recall that the instantaneous link travel time function can be expressed as

$$c_a(t) = g_{1a}[x_a(t), u_a(t)] + g_{2a}[x_a(t), v_a(t)] \qquad \forall a. \tag{12.89}$$

Denote an auxiliary link travel time function $\lambda_a^{s^*}(t)$ as

$$\lambda_a^{s^*}(t) = g_{2a}[x_a(t), v_a(t)] + \sigma^{js^*}(t) \qquad \forall a, s; a = (i,j). \tag{12.90}$$

Recall that in Section 12.3, we defined another auxiliary link travel time function

$$\Theta_a^{s^*}(t) = \sigma^{js^*}(t) + c_a^*(t) - \sigma^{is^*}(t) \qquad \forall a, s; a = (i,j). \tag{12.91}$$

Substituting equations (12.89)-(12.90) into equation (12.91), we obtain

$$\begin{aligned} \Theta_a^{s^*}(t) &= \sigma^{js^*}(t) + g_{1a}[x_a(t), u_a(t)] + g_{2a}[x_a(t), v_a(t)] - \sigma^{is^*}(t) \\ &= g_{1a}[x_a(t), u_a(t)] - \sigma^{is^*}(t) + \lambda_a^{s^*}(t) \geq 0 \\ & \qquad \forall a, s; a = (i,j). \end{aligned} \tag{12.92}$$

Based on equation (12.90), we define a related auxiliary link travel time function

$$\bar{\Theta}_a^{s^*}(t) = g_{2a}[x_a(t), v_a(t)] + \sigma^{js^*}(t) - \lambda_a^{s^*}(t) = 0 \qquad \forall a, s; a = (i,j). \tag{12.93}$$

Using the above new definitions, link-time-based instantaneous DUO route choice conditions (12.53)-(12.55) are rewritten as equivalent conditions as follows.

$$\Theta_a^{s^*}(t) \geq 0 \qquad \forall a = (i,j), s; \tag{12.94}$$

$$\bar{\Theta}_a^{s^*}(t) = 0 \qquad \forall a = (i,j), s; \tag{12.95}$$

$$u_a^{s^*}(t)\, \Theta_a^{s^*}(t) = 0 \qquad \forall a = (i,j), s; \tag{12.96}$$

$$v_a^{s^*}(t)\, \bar{\Theta}_a^{s^*}(t) = 0 \qquad \forall a = (i,j), s; \tag{12.97}$$

$$u_a^s(t) \geq 0 \qquad \forall a = (i,j), s; \tag{12.98}$$

$$v_a^s(t) \geq 0 \qquad \forall a = (i,j), s. \tag{12.99}$$

Equation (12.94) is equivalent to equation (12.53), and equation (12.96) is equivalent to equation (12.54). Then, the link-time-based variational inequality (12.88) or Theorem 12.3 can be restated as an equivalent VI in the following theorem.

*Theorem 12.5.* The dynamic traffic flow pattern satisfying network constraint set (12.6)-(12.17) is in a link-time-based instantaneous DUO route choice state if and only if it satisfies the variational inequality problem:

$$\begin{aligned} &\int_0^T \sum_s \sum_a \Big\{ \Theta_a^{s^*}(t)\ \Big[u_a^s(t) - u_a^{s^*}(t)\Big] \\ &+ \ \bar{\Theta}_a^{s^*}(t)\ \Big[v_a^s(t) - v_a^{s^*}(t)\Big] \Big\}\ dt \ \geq\ 0 \end{aligned} \tag{12.100}$$

The second term in the above variational inequality equals zero. It is placed within the VI so that the reformulation of the VI as an optimal control problem can be performed more easily. Therefore, the above VI is equivalent to the link-time-based instantaneous DUO route choice conditions (12.53)-(12.55) or (12.94)-(12.99). The proofs of necessity and sufficiency are straightforward and not given here.

Substituting definitions (12.92) and (12.93) into equation (12.100), variational inequality (12.100) is equivalent to

$$\begin{aligned}
& \int_0^T \sum_s \sum_a \Big\{ \Big[ g_{1a}[x_a(t), u_a(t)] - \sigma^{is^*}(t) + \lambda_a^{s^*}(t) \Big] \Big[ u_a^s(t) - u_a^{s^*}(t) \Big] \\
+ & \Big[ g_{2a}[x_a(t), v_a(t)] - \lambda_a^{s^*}(t) + \sigma^{js^*}(t) \Big] \Big[ v_a^s(t) - v_a^{s^*}(t) \Big] \Big\} \, dt \\
= & \int_0^T \sum_a \Big\{ g_{1a}[x_a(t), u_a(t)] \Big[ u_a(t) - u_a^*(t) \Big] \\
+ & g_{2a}[x_a(t), v_a(t)] \Big[ v_a(t) - v_a^*(t) \Big] \Big\} \, dt \\
+ & \int_0^T \sum_s \sum_a \Big\{ \Big[ -\sigma^{is^*}(t) + \lambda_a^{s^*}(t) \Big] \Big[ u_a^s(t) - u_a^{s^*}(t) \Big] \\
+ & \Big[ -\lambda_a^{s^*}(t) + \sigma^{js^*}(t) \Big] \Big[ v_a^s(t) - v_a^{s^*}(t) \Big] \Big\} \, dt \geq 0 \qquad (12.101)
\end{aligned}$$

We now show that a relaxation or diagonalization procedure can be designed so that the above VI can be formulated as an optimal control model in each relaxation iteration. In other words, our optimal control model in Chapter 5 is a special case of this VI. Now we assume that the actual link travel time $\tau_a(t)$ is fixed temporarily in the flow propagation constraints at each relaxation iteration. Then, the cross-effects of flow variables at different time instants can be separable at each iteration. In other words, a Jacobian submatrix of the instantaneous link travel time $c_a(t)$ with respect to the inflow $u_a(t)$ for each time instant can be written as

$$\nabla_u c_a^*(t) = \begin{bmatrix} \frac{\partial g_{11}(t)}{\partial u_1(t)} & 0 & \cdots & 0 \\ 0 & \frac{\partial g_{12}(t)}{\partial u_2(t)} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \frac{\partial g_{1n}(t)}{\partial u_n(t)} \end{bmatrix}$$

where $n$ is the total number of links in the network. Obviously, the above matrix is symmetric. We note that $x_a(t)$ does not enter the Jacobian submatrix because it is a state variable. Another Jacobian sub-matrix of the instantaneous link travel time $c_a(t)$ with respect to the exiting flow $v_a(t)$ for each time instant can be written as

$$\nabla_v c_a^*(t) = \begin{bmatrix} \frac{\partial g_{21}(t)}{\partial v_1(t)} & 0 & \cdots & 0 \\ 0 & \frac{\partial g_{22}(t)}{\partial v_2(t)} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \frac{\partial g_{2n}(t)}{\partial v_n(t)} \end{bmatrix}$$

which is also symmetric. Then, at each time instant $t$, there is an optimization problem which is equivalent to the integrand of variational inequality (12.100):

$$\begin{aligned}
\text{Min} \quad & \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t), \omega]\, d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t), \omega]\, d\omega \right\} \\
+ & \sum_a \sum_s \left\{ u_a^s(t) \left[ -\sigma^{is^*}(t) + \lambda_a^{s^*}(t) \right] + v_a^s(t) \left[ -\lambda_a^{s^*}(t) + \sigma^{js^*}(t) \right] \right\} \\
= & \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t), \omega]\, d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t), \omega]\, d\omega \right\} \\
- & \sum_a \sum_s [\sigma^{is^*}(t)\; u_a^s(t) + \sigma^{js^*}(t)\; v_a^s(t)] \\
+ & \sum_a \sum_s \lambda_a^{s^*}(t)\; [u_a^s(t) - v_a^s(t)] \qquad (12.102)
\end{aligned}$$

Reorganizing the above equation based on each node $j$, we have

$$\begin{aligned}
\text{Min} \quad & \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t), \omega] d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t), \omega] d\omega \right\} \\
+ & \sum_s \sum_a \lambda_a^{s^*}(t)\; [u_a^s(t) - v_a^s(t)] - \sum_s \sigma^{rs^*}(t) \sum_{a \in A(r)} u_a^s(t) \\
+ & \sum_s \sum_{j \neq rs} \sigma^{js^*}(t)\; [\sum_{a \in B(j)} v_a^s(t) - \sum_{a \in A(j)} u_a^s(t)] \\
+ & \sum_s \sigma^{ss^*}(t) \sum_{a \in B(s)} v_a^s(t) \\
= & \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t), \omega] d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t), \omega] d\omega \right\} \\
+ & \sum_{rs} \sum_a \lambda_a^{s^*}(t)[u_a^{rs}(t) - v_a^{rs}(t)] - \sum_{r \neq s} \sum_s \sigma^{rs^*}(t) \sum_{a \in A(r)} u_a^{rs}(t) \\
+ & \sum_{rs} \sum_{j \neq rs} \sigma^{js^*}(t)[\sum_{a \in B(j)} v_a^{rs}(t) - \sum_{a \in A(j)} u_a^{rs}(t)] \\
+ & \sum_r \sum_{s \neq r} \sigma^{ss^*}(t) \sum_{a \in B(s)} v_a^{rs}(t)
\end{aligned}$$

The above equation is equivalent to the following partial Hamiltonian function with flow conservation constraints only:

$$\begin{aligned}
\mathcal{H} &= \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t),\omega]d\omega + \int_0^{v_a(t)} g_{2a}[x_a(t),\omega]d\omega \right\} \\
&+ \sum_{rs}\sum_{ap} \lambda_a^{s*}(t)[u_{ap}^{rs}(t) - v_{ap}^{rs}(t)] \\
&+ \sum_{r\neq s}\sum_{s} \sigma^{rs*}(t)[f^{rs}(t) - \sum_{a\in A(r)}\sum_{p} u_{ap}^{rs}(t)] \\
&+ \sum_{rs}\sum_{j\neq rs}\sum_{p} \sigma^{js*}(t)[\sum_{a\in B(j)} v_{ap}^{rs}(t) - \sum_{a\in A(j)} u_{ap}^{rs}(t)] \\
&+ \sum_{r}\sum_{s\neq r} \sigma^{ss*}(t)[\sum_{a\in B(s)}\sum_{p} v_{ap}^{rs}(t) - e^{rs}(t)]
\end{aligned}$$

where the flow propagation constraint is not included. Thus, we can simply state the objective function of the optimization program as

$$\begin{aligned}
\text{Min} \quad & \int_0^T \sum_a \left\{ \int_0^{u_a(t)} g_{1a}[x_a(t),\omega]\, d\omega \right. \\
& + \left. \int_0^{v_a(t)} g_{2a}[x_a(t),\omega]\, d\omega \right\} dt \qquad (12.103)
\end{aligned}$$

because other terms in the partial Hamiltonian function are associated with link and node flow conservation equations. Therefore, link-time-based variational inequality (12.88) can be reformulated as an optimal control problem with objective function (12.103) and constraints (12.6)-(12.17) at each relaxation iteration. In other words, we have demonstrated that our original optimal control model in Chapter 5 is a special relaxation or diagonalization problem of VI formulation (12.88). We note that the actual link travel time $\tau_a(t)$ is fixed temporarily in the flow propagation constraints at each relaxation iteration.

The solution procedure in Chapter 6 is in fact the relaxation algorithm for solving VI model (12.88). Therefore, similar solution procedures can be developed for solving the link-based multi-group VI model for the instantaneous DUO route choice problem.

Next, we consider a special case in which the instantaneous link travel time function depends on the state variable only, namely,

$$c_a(t) = c_a[x_a(t)] \qquad \forall a. \qquad (12.104)$$

Likewise, the actual link travel time function is:

$$\tau_a(t) = \tau_a[x_a(t)] \qquad \forall a. \qquad (12.105)$$

The same derivation and analysis from (12.90) to (12.101) applies here. However, it is not necessary to design a relaxation procedure in this case. Because link travel time functions depend on the state variable only, the cross-effects of flow variables at different time instants are separable. In other words, the two Jacobian submatrices of the instantaneous link travel time $c_a(t)$ with respect to the inflow $u_a(t)$ and exit flow $v_a(t)$ for each time instant can be written as

$$\nabla_u c_a^*(t) = 0$$

$$\nabla_v c_a^*(t) = 0$$

which are symmetric. Thus, we can obtain the following optimization program:

$$\text{Min} \quad \int_0^T \sum_a u_a(t) c_a[x_a(t)]\, dt \tag{12.106}$$

s.t. constraints (12.6)-(12.17).

The above optimal control problem is equivalent to variational inequality (12.88). Equation (12.106) is an objective function for a kind of dynamic system-optimal route choice problem. It is a special case of our general instantaneous DUO route choice problems when assumptions (12.104)-(12.105) hold.

## 12.6 Notes

Variational inequality formulation approaches originated with static transportation network problems. The static user-optimal route choice problem was formulated as an equivalent set of inequalities by Smith (1979). Later, Dafermos (1980) developed an elastic demand model with disutility functions using the VI approach. An elastic demand model with demand functions was introduced by Dafermos and Nagurney (1984b). Fisk and Boyce (1983) also presented alternative VI formulations for network equilibrium travel choice problems. Nagurney (1993) summarized the modeling and algorithmic aspects of VI models for static traffic assignment problems. Recently, Friesz et al (1993) formulated a VI model for the simultaneous departure time/route choice problem. Smith (1993) also presented a route-based VI formulation using the packet representation of vehicle groups. Both dynamic models are route-based, which require explicit route enumeration for formulation and solution.

As discussed in Sections 12.1-12.4, route-based and link-based definitions of instantaneous DUO are not necessarily equivalent. Route-based definitions do not provide any routing strategies at intermediate decision nodes and have less applicability compared to the link-based definitions in ATIS systems. Furthermore, link-based DUO definitions imply route-based DUO definitions and are more realistic in terms of users' route choice behavior. Therefore, in this book and in studies of dynamic incident management, we focus on models formulated according to the link-based DUO definitions.

# Chapter 13

# Variational Inequality Models of Ideal Dynamic User-Optimal Route Choice Problems

In this chapter, we present both route-based and link-based variational inequality models for the ideal dynamic user-optimal route choice problem. In Section 13.1, a route-time-based VI model for ideal DUO route choice is proposed. This model is the most straight-forward formulation of route-time-based, ideal DUO route choice conditions. In Section 13.2, a multi-group route-time-based VI model is developed. In this model, each group of travelers is associated with a disutility function. Thus, the route-based ideal DUO route choice conditions are defined for each group of travelers on the basis of travel disutilities instead of travel times only.

Route-time-based VI models have an intuitive interpretation. However, they encounter a computational difficulty in terms of explicit route enumeration. A link-time-based VI model is therefore proposed in Section 13.3. We prove that the link-time-based ideal DUO route choice conditions imply the route-time-based ideal DUO route choice conditions. In parallel to the route-based VI models, a multi-group link-based VI model is also presented in Section 13.4.

In Section 13.5, the relationships between VI models and optimization models are discussed. As an example, the link-time-based VI model is reformulated as optimal control problem. Thus, an algorithm without explicit route enumeration can be designed to solve this VI.

## 13.1 A Route-Time-Based VI Model of Ideal Route Choice

### 13.1.1 Route-Time-Based Conditions

Recall from Chapter 4 that the actual link travel time is the travel time over a link actually experienced by vehicles. The actual travel time $\tau_a(t)$ over link $a$ for vehicles entering link $a$ at time $t$ is assumed to be dependent on the number of vehicles $x_a(t)$, the inflow $u_a(t)$ and the exit flow $v_a(t)$ on link $a$ at time $t$. It follows that

$$\tau_a(t) = \sum_{a \in rsp} \tau_a[x_a(t), u_a(t), v_a(t)] \qquad \forall a \tag{13.1}$$

Suppose we have a set of routes $\{p\}$ and $f_p^{rs}(t)$ is the route inflow from origin $r$ to destination $s$ at time $t$. Denote $\eta_p^{rs}(t)$ as the actual travel time from origin $r$ to destination $s$ over route $p$ at time $t$, and $\pi^{rs}(t)$ as the minimal actual route travel time from origin $r$ to destination $s$ at time $t$. In Chapter 4 we defined a recursive formula to compute the route travel time $\eta_p^{rs}(t)$ for each allowable route as follows. Assume route $p$ consists of nodes $(r, 1, 2, \cdots, i, \cdots, s)$. Denote $\eta_p^{ri}(t)$ as the travel time *actually* experienced over route $p$ from origin $r$ to node $i$ by vehicles departing origin $r$ at time $t$. Then, a recursive formula for route travel time $\eta_p^{rs}(t)$ is:

$$\eta_p^{ri}(t) = \eta_p^{r(i-1)}(t) + \tau_a[t + \eta_p^{r(i-1)}(t)] \qquad \forall p, r, i; i = 1, 2, \cdots, s;$$

where link $a = (i-1, i)$.

If the actual link travel time $\tau_a(t)$ is determined, the minimal actual O-D travel time $\pi^{rs}(t)$ can be computed as $\pi^{rs}(t) = \min_p \eta_p^{rs}(t)$. $\pi^{rs}(t)$ is a functional of all link flow variables at time $\omega$: $\pi^{rs}(t) = \pi^{rs}[u(\omega), v(\omega), x(\omega)|\omega \geq t]$. This functional is neither a state variable nor a control variable, and it is not fixed. This functional is not available in closed form.

For any route $p$ from origin $r$ to destination $s$, route $p$ is defined as being used at time $t$ if $f_p^{rs}(t) > 0$. In this chapter, we assume that the time-dependent origin-destination trip pattern is known *a priori*. Thus, the departure times of travelers are given. From Chapter 7, recall the definition of the ideal dynamic user-optimal (DUO) state as follows.

**Travel-Time-Based Ideal DUO State:** *If, for each O-D pair at each instant of time, the actual travel times for all routes that are being used equal the minimal actual route travel time, the dynamic traffic flow over the network is in a travel-time-based ideal dynamic user-optimal state.*

The route-time-based ideal DUO route choice conditions which are equivalent to the above definition are defined as follows.

$$\eta_p^{rs*}(t) - \pi^{rs*}(t) \geq 0 \qquad \forall p, r, s; \tag{13.2}$$

$$f_p^{rs^*}(t)\left[\eta_p^{rs^*}(t) - \pi^{rs^*}(t)\right] = 0 \qquad \forall p, r, s; \tag{13.3}$$

$$f_p^{rs}(t) \geq 0 \qquad \forall p, r, s. \tag{13.4}$$

The above definition of an ideal DUO state and its corresponding route choice conditions are defined for each O-D pair. They are not defined for each decision node-destination pair as in Chapter 4 (which is equivalent to link-based definition in Section 13.3). Thus, rerouting strategies are not provided for travelers at intermediate intersections.

## 13.1.2 Dynamic Network Constraints

The constraint set for this route-time-based VI model is summarized as follows.

Relationship between state and control variables:

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \tag{13.5}$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r; s \neq r; \tag{13.6}$$

Flow conservation constraints:

$$f^{rs}(t) = \sum_{a \in A(r)} \sum_{p} u_{ap}^{rs}(t) \qquad \forall r, s; \tag{13.7}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \tag{13.8}$$

$$\sum_{a \in B(s)} \sum_{p} v_{ap}^{rs}(t) = e^{rs}(t) \qquad \forall r, s; s \neq r; \tag{13.9}$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[t + \tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; \tag{13.10}$$

Definitional constraints:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \tag{13.11}$$

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \qquad \forall a; \tag{13.12}$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \quad u_{ap}^{rs}(t) \geq 0, \quad v_{ap}^{rs}(t) \geq 0 \quad \forall a, p, r, s; \tag{13.13}$$

$$e_p^{rs}(t) \geq 0, \quad E_p^{rs}(t) \geq 0, \quad \forall p, r, s; \tag{13.14}$$

Boundary conditions:

$$E_p^{rs}(0) = 0, \quad \forall p, r, s; \tag{13.15}$$

$$x_{ap}^{rs}(0) = 0, \quad \forall a, p, r, s. \tag{13.16}$$

### 13.1.3 The Route-Time-Based VI Model

The equivalent variational inequality formulation of route-time-based ideal DUO route choice conditions (13.2)-(13.4) may be stated as follows.

*Theorem 13.1.* The dynamic traffic flow pattern satisfying network constraint set (13.5)-(13.16) is in a route-time-based ideal DUO route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \sum_{rs} \sum_p \eta_p^{rs^*}(t) \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] dt \geq 0 \tag{13.17}$$

*Proof of Necessity.*

We need to prove that the ideal DUO route choice conditions (13.2)-(13.4) imply variational inequality (13.17). For any route $p$, a feasible inflow at time $t$ is

$$f_p^{rs}(t) \geq 0. \tag{13.18}$$

Multiplying ideal DUO route choice condition (13.2) by the above equation, we have

$$f_p^{rs}(t) \left[\eta_p^{rs^*}(t) - \pi^{rs^*}(t)\right] \geq 0 \quad \forall p, r, s. \tag{13.19}$$

We subtract equation (13.3) from equation (13.19) and obtain

$$\left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] \left[\eta_p^{rs^*}(t) - \pi^{rs^*}(t)\right] \geq 0 \quad \forall p, r, s. \tag{13.20}$$

Summing equation (13.20) for all routes $p$ and all O-D pairs $rs$, it follows that

$$\begin{aligned}
&\sum_{rs} \sum_p \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] \left[\eta_p^{rs^*}(t) - \pi^{rs^*}(t)\right] \\
=\ & \sum_{rs} \sum_p \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] \eta_p^{rs^*}(t) - \sum_{rs} \pi^{rs^*}(t) \sum_p \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] \\
=\ & \sum_{rs} \sum_p \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] \eta_p^{rs^*}(t) \geq 0
\end{aligned} \tag{13.21}$$

where the flow conversation equation

$$\sum_p f_p^{rs}(t) = \sum_p f_p^{rs^*}(t) = f^{rs}(t)$$

holds for each O-D $rs$ at each time $t$. Integrating the above equation (13.21) from 0 to $T$, we obtain variational inequality (13.17).

*Proof of Sufficiency.*

We need to prove that any solution $f_p^{rs^*}(t)$ to variational inequality (13.17) satisfies ideal DUO route choice conditions (13.2)-(13.4). We know that the first and third ideal DUO route choice conditions (13.2) and (13.4) hold by definition. Thus, we need to prove that the second ideal DUO route choice condition (13.3) also holds.

Assume that the second ideal DUO route choice condition (13.3) does not hold only for a route $q$ for O-D pair $kn$ during time interval $[t_1-\delta, t_1+\delta] \in [0,T]$, i.e.,

$$f_q^{kn^*}(t) > 0 \quad \text{and} \quad \eta_q^{kn^*}(t) - \pi^{kn^*}(t) > 0 \quad \forall t \in [t_1-\delta, t_1+\delta] \quad (13.22)$$

Since the second ideal DUO route choice condition (13.3) holds for all routes other than route $q$ for O-D pair $kn$ at time $t$, it follows that

$$\begin{aligned} & \int_0^T \sum_{rs}\sum_p f_p^{rs^*}(t)\,[\eta_p^{rs^*}(t) - \pi^{rs^*}(t)]\,dt \\ = & \int_{t_1-\delta}^{t_1+\delta} f_q^{kn^*}(t)\ [\eta_q^{kn^*}(t) - \pi^{kn^*}(t)]\,dt\ > 0 \end{aligned} \quad (13.23)$$

We note that all other terms in the above equation vanish because of ideal DUO route choice condition (13.3).

For each O-D pair $rs$, we can always find one minimal actual travel time route $l$ for vehicles departing origin $r$ at time $t$, where route $l$ was evaluated under the optimal flow pattern $\{f_p^{rs^*}(t)\}$. For this route $l$, the first ideal DUO route choice condition (13.2) becomes an equality by definition. It follows that

$$\eta_l^{rs^*}(t) - \pi^{rs^*}(t) = 0 \qquad \forall l, r, s. \quad (13.24)$$

Next, we need to find a set of feasible route inflows $f_p^{rs}(t)$ so that the following equations

$$f_p^{rs}(t)\ \left[\eta_p^{rs^*}(t) - \pi^{rs^*}(t)\right] = 0 \qquad \forall p, r, s \quad (13.25)$$

always hold. We consider all the departure flows $f^{rs}(t)$ for all O-D pairs at each time $t$. For each O-D pair $rs$ at each time $t$, we assign O-D departure flow $f^{rs}(t)$ to the minimal travel time route $l$, which was evaluated under the optimal flow pattern $\{f_p^{rs^*}(t)\}$. This will generate a set of feasible route inflow patterns $\{f_p^{rs}(t)\}$ which always satisfies equation (13.25) because flows are not assigned to routes with non-minimal travel times which were evaluated under the optimal route inflow pattern $\{f_p^{rs^*}(t)\}$. Summing equations (13.25) for all routes $p$ and all O-D pairs $rs$, it follows that

$$\sum_{rs}\sum_{p} f_p^{rs}(t)\ \left[\eta_p^{rs^*}(t) - \pi^{rs^*}(t)\right] = 0 \tag{13.26}$$

Integrating the above equation for time period $[0, T]$, we have

$$\int_0^T \sum_{rs}\sum_{p} f_p^{rs}(t)\ \left[\eta_p^{rs^*}(t) - \pi^{rs^*}(t)\right]\ dt\ = 0 \tag{13.27}$$

We subtract equation (13.23) from equation (13.27) and obtain

$$\begin{aligned}
&\int_0^T \sum_{rs}\sum_{p} \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right]\ \left[\eta_p^{rs}(t) - \pi^{rs}(t)\right]\ dt \\
=\ &\int_0^T \left\{\sum_{rs}\sum_{p} \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right]\ \eta_p^{rs}(t) \right. \\
-\ &\left. \sum_{rs} \pi^{rs}(t) \sum_{p} \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right]\right\}\ dt \\
=\ &\int_0^T \sum_{rs}\sum_{p} \eta_p^{rs^*}(t)\ \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right]\ dt\ <\ 0
\end{aligned} \tag{13.28}$$

where the flow conservation

$$\sum_{p} f_p^{rs}(t) = \sum_{p} f_p^{rs^*}(t) = f^{rs}(t)$$

holds for each O-D pair $rs$ at each time $t$ so that the second term vanishes. The above equation contradicts VI problem (13.17). Therefore, any optimal solution $\{f_p^{rs^*}(t)\}$ to variational inequality (13.17) satisfies the second ideal DUO route choice condition (13.3). Since we proved the necessity and sufficiency of the equivalence of variational inequality (13.17) to route-time-based ideal DUO route choice conditions (13.2)-(13.4), the proof is complete.

## 13.2 A Multi-Group Route-Cost-Based VI Model of Ideal Route Choice

In this section, we define an ideal dynamic user-optimal (DUO) model based on travel costs or disutilities instead of travel times. To be consistent with Chapter 12, we still stratify travelers into $M$ groups for each O-D pair. When $M = 1$, the following definition and VI model reduces to a single group model, but one which is different from the above VI model based on actual travel times.

### 13.2.1 Multi-Group Route-Cost-Based Conditions

The actual route disutility function depends on the actual route travel time, fuel consumption enroute and automobile operating cost, etc. Denote $\tilde{\tau}_{ma}(t)$ as the actual disutility function for travelers of group $m$ entering link $a$ at time $t$. It follows that

$$\tilde{\tau}_{ma}(t) = \alpha_{ma} + \beta_{ma}\ \tau_a(t) \qquad \forall m, a \tag{13.29}$$

where $\alpha_{ma}$ is a fixed actual disutility parameter for group $m$ travelers on link $a$ and $\beta_{ma}$ is a parameter to transform actual link travel time $\tau_a(t)$ into the disutility of group $m$ travelers. Denote $\tilde{\eta}^{rs}_{mp}(t)$ as the actual route travel disutility for group $m$ travelers from origin $r$ to destination $s$ at time $t$, and $\tilde{\pi}^{rs}_m(t)$ as the minimal actual route travel disutility for group $m$ travelers from origin $r$ to destination $s$ at time $t$. We also need to use a recursive formula to compute the route travel disutility $\tilde{\eta}^{rs}_{mp}(t)$ for all allowable routes. Assume route $p$ consists of nodes $(r, 1, 2, \cdots, i, \cdots, s)$. Denote $\tilde{\eta}^{rj}_{mp}(t)$ as the travel disutility *actually* experienced over route $p$ from origin $r$ to node $j$ by vehicles departing origin $r$ at time $t$. Then, a recursive formula for route travel disutility $\tilde{\eta}^{rs}_{mp}(t)$ is:

$$\tilde{\eta}^{rj}_{mp}(t) = \tilde{\eta}^{r(j-1)}_{mp}(t) + \tilde{\tau}_{ma}[t + \eta^{r(j-1)}_p(t)] \qquad \forall m, p, r, j; j = 1, 2, \cdots, s;$$

where link $a = (j-1, j)$ and time $[t + \eta^{r(j-1)}_p(t)]$ is the arrival time at link $a$ for group $m$ travelers. We then define a multi-group travel-cost-based ideal DUO state as follows.

> **Multi-Group Travel-Cost-Based Ideal DUO State:** *If, for each group m and each O-D pair at each instant of time, the actual travel* disutilities *for all routes that are being used equal the minimal ideal route travel* disutility, *the dynamic traffic flow over the network is in a multi-group travel-cost-based ideal dynamic user-optimal state.*

Recall that $f^{rs}_{mp}(t)$ is the route inflow of group $m$ from origin $r$ to destination $s$ at time $t$. The equivalent multi-group ideal DUO route choice conditions can be summarized as follows:

$$\tilde{\eta}^{rs^*}_{mp}(t) - \tilde{\pi}^{rs^*}_m(t) \geq 0 \qquad \forall m, p, r, s; \tag{13.30}$$

$$f^{rs^*}_{mp}(t) \left[\tilde{\eta}^{rs^*}_{mp}(t) - \tilde{\pi}^{rs^*}_m(t)\right] = 0 \qquad \forall m, p, r, s; \tag{13.31}$$

$$f^{rs}_{mp}(t) \geq 0 \qquad \forall m, p, r, s. \tag{13.32}$$

### 13.2.2 Dynamic Network Constraints

The revised constraint set for the multi-group route-cost-based model is summarized as follows.

Relationship between state and control variables:

$$\frac{dx^{rs}_{map}}{dt} = u^{rs}_{map}(t) - v^{rs}_{map}(t) \qquad \forall m, a, p, r, s; \tag{13.33}$$

$$\frac{dE^{rs}_{mp}(t)}{dt} = e^{rs}_{mp}(t) \qquad \forall m, p, r; s \neq r; \tag{13.34}$$

Flow conservation constraints:

$$f^{rs}_{m}(t) = \sum_{a \in A(r)} \sum_{p} u^{rs}_{map}(t) \qquad \forall m, r, s; \tag{13.35}$$

$$\sum_{a \in B(j)} v^{rs}_{map}(t) = \sum_{a \in A(j)} u^{rs}_{map}(t) \qquad \forall j, m, p, r, s; j \neq r, s; \tag{13.36}$$

$$\sum_{a \in B(s)} \sum_{p} v^{rs}_{map}(t) = e^{rs}_{m}(t) \qquad \forall m, r, s; s \neq r; \tag{13.37}$$

Flow propagation constraints:

$$x^{rs}_{ap}(t) = \sum_{b \in \tilde{p}} \{x^{rs}_{bp}[t + \tau_a(t)] - x^{rs}_{bp}(t)\} + \{E^{rs}_{p}[t + \tau_a(t)] - E^{rs}_{p}(t)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; \tag{13.38}$$

Definitional constraints:

$$\sum_{mrsp} u^{rs}_{map}(t) = u_a(t), \qquad \sum_{mrsp} v^{rs}_{map}(t) = v_a(t), \qquad \forall a; \tag{13.39}$$

$$\sum_{mrsp} x^{rs}_{map}(t) = x_a(t), \qquad \forall a; \tag{13.40}$$

Nonnegativity conditions:

$$x^{rs}_{map}(t) \geq 0, \quad u^{rs}_{map}(t) \geq 0, \quad v^{rs}_{map}(t) \geq 0 \quad \forall m, a, p, r, s; \tag{13.41}$$

$$e^{rs}_{mp}(t) \geq 0, \qquad E^{rs}_{mp}(t) \geq 0, \qquad \forall m, p, r, s; \tag{13.42}$$

Boundary conditions:

$$E^{rs}_{mp}(0) = 0, \qquad \forall m, p, r, s; \tag{13.43}$$

$$x^{rs}_{map}(0) = 0, \qquad \forall m, a, p, r, s. \tag{13.44}$$

### 13.2.3 The Multi-Group Route-Cost-Based VI Model

The equivalent variational inequality formulation of multi-group route-cost-based ideal DUO route choice conditions (13.30)-(13.32) may be stated as follows.

*Theorem 13.2.* The dynamic traffic flow pattern satisfying network constraint set (13.33)-(13.44) is in a multi-group travel-cost-based ideal DUO route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \sum_{rs} \sum_{mp} \tilde{\eta}_{mp}^{rs*}(t) \left[ f_{mp}^{rs}(t) - f_{mp}^{rs*}(t) \right] dt \geq 0 \tag{13.45}$$

The proof of necessity and sufficiency for variational inequality (13.45) follows in the same manner as in Section 13.1.3 for the single group route-time-based case.

## 13.3 A Link-Time-Based VI Model of Ideal Route Choice

The set of dynamic network constraints for the link-time-based VI model is identical to constraint set (13.5)-(13.16) of the route-time-based VI model in Section 13.1. The basic difference between the two models is that the VI itself of the link-time-based VI model is formulated using link-based flow variables instead of route-based variables as in the route-time-based VI model.

### 13.3.1 Link-Time-Based Conditions

Similar to Chapter 12, we introduce a new set of ideal DUO route choice conditions based on link and node variables, instead of route-based variables. As in Section 13.1.1,, the definition of the travel-time-based ideal DUO route choice state is given as follows.

**Travel-Time-Based Ideal DUO State:** *If, for each O-D pair at each instant of time, the* actual *travel times experienced by travelers departing at the same time are equal and minimal, the dynamic traffic flow over the network is in a travel-time-based* ideal *dynamic user-optimal state.*

Note also that the link-time-based ideal DUO state is defined in a somewhat different way than the link-time-based instantaneous DUO state in Section 12.3.1. In the case of an ideal DUO state, the equilibration of route travel times is stated for each O-D pair instead of each decision node-destination pair as in the case of instantaneous DUO state because the ideal DUO is focused on the optimal state of finishing the entire journey.

We now write the equivalent mathematical inequalities for the above definition using link variables, in contrast to the route-based formulation in Section 13.1. In this case, for any route from origin $r$ to destination $s$, link $a$ is defined as being used at time $t$ if $u_a^{rs}(t) > 0$. Define $\pi^{ri^*}(t)$ as the minimal travel time *actually* experienced by vehicles departing origin $r$ to node $i$ at time $t$, the asterisk denoting that the travel time is computed using ideal DUO traffic flows. For link $a = (i,j)$, the minimal travel time $\pi^{rj^*}(t)$ from origin $r$ to $j$ should be equal to or less than the minimal travel time $\pi^{ri^*}(t)$ from origin $r$ to $i$ plus the actual link travel time $\tau_a[t + \pi^{ri}(t)]$ at time instant $[t + \pi^{ri}(t)]$, where this time instant is the earliest clock time when the flow departing origin $r$ at time $t$ can enter link $a$. It follows that

$$\pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)] \geq \pi^{rj^*}(t) \qquad \forall a = (i,j), r.$$

If, for each O-D pair $rs$, any departure flow from origin $r$ at time $t$ enters link $a$ at the earliest clock time $[t + \pi^{ri^*}(t)]$, or $u_a^{rs}[t + \pi^{ri^*}(t)] > 0$, then the ideal DUO route choice conditions require that link $a$ is on the minimal travel time route. In other words, the minimal travel time $\pi^{rj^*}(t)$ from origin $r$ to $j$ should equal the minimal travel time $\pi^{ri^*}(t)$ from origin $r$ to $i$ plus the actual link travel time $\tau_a[t + \pi^{ri^*}(t)]$ at time instant $[t + \pi^{ri^*}(t)]$. It follows that

$$\pi^{rj^*}(t) = \pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)], \quad \text{if } u_a^{rs^*}[t + \pi^{ri^*}(t)] > 0 \quad \forall a = (i,j), r, s.$$

The above equation is also equivalent to the following:

$$\left[\pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)] - \pi^{rj^*}(t)\right] \; u_a^{rs^*}[t + \pi^{ri^*}(t)] = 0 \qquad \forall a = (i,j), r, s.$$

Thus, the link-time-based ideal DUO route choice conditions can be summarized as follows:

$$\pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)] - \pi^{rj^*}(t) \; \geq \; 0 \qquad \forall a = (i,j), r; \qquad (13.46)$$

$$u_a^{rs^*}[t + \pi^{ri^*}(t)] \; \left[\pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)] - \pi^{rj^*}(t)\right] \; = \; 0 \qquad \forall a = (i,j), r, s; \qquad (13.47)$$

$$u_a^{rs}[t + \pi^{ri^*}(t)] \geq 0 \qquad \forall a = (i,j), r, s. \qquad (13.48)$$

*Lemma 13.1: Link-time-based ideal DUO route choice conditions (13.46)-(13.48) imply route-time-based ideal DUO route choice conditions (13.2)-(13.4).*

*Proof:*

We need to prove for each O-D pair $rs$ that under link-time-based ideal DUO route choice conditions (13.46)-(13.48) any vehicle flows departing from origin $r$ at time $t$ must arrive at destination $s$ at the same time by using the minimal actual travel time routes.

For simplicity, we first consider the case with only two route departure flows $f_1^{rs}(t) > 0$, $f_2^{rs}(t) > 0$ for one O-D pair $rs$ at time $t$. It follows that

$$f_1^{rs}(t) + f_2^{rs}(t) = f^{rs}(t) \tag{13.49}$$

Suppose $f_1^{rs}(t)$, $f_2^{rs}(t)$ take routes 1 and 2, respectively. Routes 1 and 2 are minimal-travel-time routes generated under the link-time-based ideal DUO route choice conditions. For simplicity, assume that route 1 comprises 4 links: $1 = (r,h), 2 = (h,i), \cdots, 4 = (j,s)$; and route 2 comprises 5 links: $5 = (r,k), 6 = (k,l), \cdots, 9 = (m,s)$. Note that routes 1 and 2 may have overlapping links. Also note that this assumption can be generalized to any route with any number of links. Using link-time-based ideal DUO route choice conditions (13.46)-(13.48), we have

$$u_1^{rs^*}(t) > 0, u_2^{rs}[t + \pi^{rh^*}(t)] > 0, \cdots, u_4^{rs}[t + \pi^{rj^*}(t)] > 0 \tag{13.50}$$

$$u_5^{rs^*}(t) > 0, u_6^{rs}[t + \pi^{rk^*}(t)] > 0, \cdots, u_9^{rs}[t + \pi^{rm^*}(t)] > 0 \tag{13.51}$$

This is because route 1 and route 2 are generated under link-time-based ideal DUO route choice conditions (13.46)-(13.48) so that there are inflows into links $1, 2, \cdots, 9$ over route 1 and route 2. If route 1 and route 2 do not have overlapping links, the inflow on each link over route 1 and 2 is positive at the instant of time when departure flows arrive at the link. It follows that

$$u_{11}^{rs^*}(t) > 0, u_{21}^{rs}[t + \pi^{rh^*}(t)] > 0, \cdots, u_{41}^{rs}[t + \pi^{rj^*}(t)] > 0 \tag{13.52}$$

$$u_{52}^{rs^*}(t) > 0, u_{62}^{rs}[t + \pi^{rk^*}(t)] > 0, \cdots, u_{92}^{rs}[t + \pi^{rm^*}(t)] > 0 \tag{13.53}$$

where the second subscripts 1 and 2 represent routes 1 and 2, respectively. Note that the instants of time when departure flows arrive at the links are ensured by link-time-based ideal DUO route choice conditions (13.46)-(13.48). For example, $[t + \pi^{rh^*}(t)]$ is the instant of time when departure flow $f_1^{rs}(t)$ arrives at link 2. In other words, if departure flows $f_1^{rs}(t) > 0$, $f_2^{rs}(t) > 0$ satisfy link-time-based ideal DUO route choice conditions (13.46)-(13.48), we obtain equations (13.52)-(13.53).

If routes 1 and 2 have overlapping links, conditions (13.52)-(13.53) still hold. For example, if link 2 is identical to link 6 (routes 1 and 2 are overlapping on link 2 or link 6), $[t + \pi^{rh^*}(t)] = [t + \pi^{rk^*}(t)]$ is the instant of time when departure flows arrive at link 2. Both flows must experience the same link travel time $\tau_2[t+\pi^{rh^*}(t)]$ and exit link 2 at the same time $[t+\pi^{ri^*}(t)]$. Then the inflows on subsequent links over routes 1 and 2 still satisfy equations (13.52)-(13.53).

Denote the arrival flows over routes 1 and 2 as $e_1^{rs}[t + \pi^{rs^*}(t)]$ and $e_2^{rs}[t + \pi^{rs^*}(t)]$, which are associated with departure flows $f_1^{rs}(t)$ and $f_2^{rs}(t)$, respectively. Note that routes 1 and 2 are minimal travel time routes. Using route-based flow propagation constraints (13.10) for the last links $4 = (j, s)$ and $9 = (m, s)$ over routes 1 and 2, we obtain that

$$e_1^{rs}[t + \pi^{rs^*}(t)] > 0, \;\; e_2^{rs}[t + \pi^{rs^*}(t)] > 0 \tag{13.54}$$

where

$$\pi^{rs^*}(t) = \pi^{rj^*}(t) + \tau_4[t + \pi^{rj^*}(t)] = \pi^{rm^*}(t) + \tau_9[t + \pi^{rm^*}(t)]$$

Note that

$$u_{11}^{rs^*}(t), u_{21}^{rs}[t + \pi^{rh^*}(t)], \cdots, u_{41}^{rs}[t + \pi^{rj^*}(t)], e_1^{rs}[t + \pi^{rs^*}(t)]$$

and

$$u_{52}^{rs^*}(t), u_{62}^{rs}[t + \pi^{rk^*}(t)], \cdots, u_{92}^{rs}[t + \pi^{rm^*}(t)], e_2^{rs}[t + \pi^{rs^*}(t)]$$

are the two sets of inflows over routes 1 and 2, respectively. Since these flows are positive, we conclude that the departure flows $f_1^{rs}(t)$ and $f_2^{rs}(t)$ arrive at destination $s$ at the same time $[t + \pi^{rs^*}(t)]$. Thus, the link-time-based ideal DUO route choice conditions guarantee for O-D pair $rs$ that flows departing at time $t$ have the same arrival time $[t + \pi^{rs^*}(t)]$.

If we consider a general case having multiple route departure flows $f_p^{rs}(t) > 0$ for any O-D pair $rs$ at time $t$, the above analysis still applies to any positive departure flow over any route $p$ between O-D pair $rs$. Therefore, link-time-based ideal DUO route choice conditions (13.46)-(13.48) imply route-time-based ideal DUO route choice conditions (13.2)-(13.4).

### 13.3.2 The Link-Time-Based Model

Denote $\Omega_a^{rj^*}(t)$ as the difference of the minimal travel time from $r$ to $j$ and the travel time from $r$ to $j$ via minimal travel time route from $r$ to $i$ and link $a$ for vehicles departing from origin at time $t$. It follows that

$$\Omega_a^{rj^*}(t) = \pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)] - \pi^{rj^*}(t) \qquad \forall a, r; a = (i, j). \tag{13.55}$$

We then rewrite the link-time-based ideal DUO route choice conditions as:

$$\Omega_a^{rj^*}(t) \geq 0 \qquad \forall a = (i, j), r; \tag{13.56}$$

$$u_a^{rs^*}[t + \pi^{ri^*}(t)]\, \Omega_a^{rj^*}(t) = 0 \qquad \forall a = (i, j), r, s; \tag{13.57}$$

$$u_a^{rs}[t + \pi^{ri^*}(t)] \geq 0 \qquad \forall a = (i, j), r, s. \tag{13.58}$$

Then, the equivalent variational inequality formulation of link-time-based ideal DUO route choice conditions (13.56)-(13.58) may be stated as follows.

*Theorem 13.3.* The dynamic traffic flow pattern satisfying constraints (13.5)-(13.16) is in a link-time-based ideal DUO route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \sum_{rs}\sum_a \Omega_a^{rj^*}(t) \left\{u_a^{rs}[t+\pi^{ri^*}(t)] - u_a^{rs^*}[t+\pi^{ri^*}(t)]\right\} dt \geq 0 \quad (13.59)$$

*Proof of Necessity.*

We need to prove that link-time-based ideal DUO route choice conditions (13.56)-(13.58) imply variational inequality (13.59). For any link $a$, a feasible inflow at time $[t+\pi^{ri^*}(t)]$ is

$$u_a^{rs}[t+\pi^{ri^*}(t)] \geq 0 \quad (13.60)$$

Multiplying equation (13.60) and DUO route choice condition (13.56), we have

$$u_a^{rs}[t+\pi^{ri^*}(t)]\, \Omega_a^{rj^*}(t) \geq 0 \qquad \forall a, r, s; a = (i,j). \quad (13.61)$$

We subtract the second ideal DUO route choice condition (13.57) from equation (13.61) and obtain

$$\left\{u_a^{rs}[t+\pi^{ri^*}(t)] - u_a^{rs^*}[t+\pi^{ri^*}(t)]\right\} \Omega_a^{rj^*}(t) \geq 0 \quad \forall a, r, s; a = (i,j). \quad (13.62)$$

Summing equation (13.62) for all links $a$ and all O-D pairs $rs$, it follows that

$$\sum_{rs}\sum_a \left\{u_a^{rs}[t+\pi^{ri^*}(t)] - u_a^{rs^*}[t+\pi^{ri^*}(t)]\right\} \Omega_a^{rj^*}(t) \geq 0 \quad (13.63)$$

Integrating the above equation from 0 to $T$, we obtain variational inequality (13.59).

*Proof of Sufficiency.*

We need to prove that any solution $u_a^{rs^*}[t+\pi^{ri^*}(t)]$ to variational inequality (13.59) satisfies link-time-based ideal DUO route choice conditions (13.56)-(13.58). We know that the first and third ideal DUO route choice conditions (13.56) and (13.58) hold by definition. Thus, we need to prove that the second ideal DUO route choice condition (13.57) also holds.

Assume that the second ideal DUO route choice condition (13.57) does not hold only for a link $b = (l,m)$ for O-D pair $kn$ during time interval $[t_1 - \delta, t_1 + \delta] \in [0,T]$, i.e.,

$$u_b^{kn^*}[t+\pi^{kl^*}(t)] > 0 \qquad \text{and} \qquad \Omega_b^{km^*}(t) > 0 \quad (13.64)$$

In other words, we have

$$u_b^{kn^*}[t+\pi^{kl^*}(t)]\, \Omega_b^{km^*}(t) > 0 \quad (13.65)$$

Since the second ideal DUO route choice condition (13.57) holds for all cases other than link $b = (l,m)$ for O-D pair $kn$ during time interval $[t_1 - \delta, t_1 + \delta]$, it follows that

$$\int_0^T \sum_{rs}\sum_a \Omega_a^{rj^*}(t)\; u_a^{rs^*}[t+\pi^{ri^*}(t)]\; dt$$

$$= \int_{t_1-\delta}^{t_1+\delta} u_b^{kn^*}[t+\pi^{kl^*}(t)]\;\; \Omega_b^{km^*}(t)\; dt\; > 0 \qquad (13.66)$$

We note that all other terms in the above equation vanish because of ideal DUO route choice condition (13.57).

For each O-D pair $rs$, we can always find one minimal travel time route $p$ for vehicles departing origin $r$ at time $t$, where route $p$ is evaluated under the optimal flow pattern $\{u_a^{rs^*}[t+\pi^{ri^*}(t)]\}$. For this route $p$, the first ideal DUO route choice condition (13.56) becomes an equality by definition. It follows that

$$\Omega_a^{rj^*}(t) = 0 \qquad \forall a, r, s; a = (i,j); a \in p. \qquad (13.67)$$

Next, we need to find a set of feasible inflows $u_a^{rs}[t+\pi^{ri^*}(t)]$ so that the following equations

$$u_a^{rs}[t+\pi^{ri^*}(t)]\; \Omega_a^{rj^*}(t) = 0 \qquad \forall a, r, s; a = (i,j) \qquad (13.68)$$

always hold. We adjust all the departure flows $f^{rs}(t)$ for all O-D pairs at time $t$. For each O-D pair $rs$ at each time $t$, we assign O-D departure flow $f^{rs}(t)$ to the minimal travel time route $p$, which was evaluated under the optimal flow pattern $\{u_a^{rs^*}[t+\pi^{ri^*}(t)]\}$. This will generate a set of feasible inflow patterns $\{u_a^{rs}[t+\pi^{ri^*}(t)]\}$ which always satisfies equations (13.67) and (13.68) because flows are not assigned to routes with non-minimal actual travel times which were evaluated under the optimal flow pattern $\{u_a^{rs^*}[t+\pi^{ri^*}(t)]\}$. Summing equations (13.68) for all links $a$ and all O-D pairs $rs$, it follows that

$$\sum_{rs}\sum_a u_a^{rs}[t+\pi^{ri^*}(t)]\; \Omega_a^{rj^*}(t) = 0 \qquad \text{where } a = (i,j). \qquad (13.69)$$

Integrating the above equation from 0 to $T$, we obtain

$$\int_0^T \sum_{rs}\sum_a u_a^{rs}[t+\pi^{ri^*}(t)]\; \Omega_a^{rj^*}(t)\; dt\; = 0 \qquad (13.70)$$

We subtract equation (13.66) from equation (13.70) and obtain

$$\int_0^T \sum_{rs}\sum_a \Omega_a^{rj^*}(t)\; \Big\{u_a^{rs}[t+\pi^{ri^*}(t)] - u_a^{rs^*}[t+\pi^{ri^*}(t)]\Big\}\; dt\; < 0 \quad (13.71)$$

The above equation contradicts variational inequality (13.59). Therefore, any optimal solution $\{u_a^{rs^*}[t+\pi^{ri^*}(t)]\}$ to variational inequality (13.59) satisfies the second ideal DUO route choice condition (13.57). Since we proved the necessity and sufficiency of the equivalence of variational inequality (13.59) to link-time-based ideal DUO route choice conditions (13.56)-(13.58), the proof is complete.

## 13.4 A Multi-Group Link-Cost-Based VI Model of Ideal Route Choice

In this section, we consider a multi-group ideal dynamic user-optimal (DUO) model based on travel *costs or disutilities* instead of travel *times*. To be consistent with Chapter 12, we stratify travelers into $M$ groups for each O-D pair. When $M = 1$, the following definition and VI model reduce to a single group model, but which is different from the VI model based on actual travel times. The constraint set for this problem is identical to the constraint set (13.33)-(13.44) for the multi-group route-time-based VI model in Section 13.2.

### 13.4.1 Multi-Group Link-Cost-Based Conditions

The disutility function depends on travel time, fuel consumption, operating cost, etc. Denote $\tilde{\tau}_{ma}(t)$ as the actual disutility function for group $m$ travelers entering link $a$ at time $t$. It follows that

$$\tilde{\tau}_{ma}(t) = \alpha_{ma} + \beta_{ma}\ \tau_a(t) \qquad \forall m, a \tag{13.72}$$

where $\alpha_{ma}$ is a fixed actual disutility parameter for group $m$ travelers on link $a$ and $\beta_{ma}$ is a parameter to transform actual link travel time $\tau_a(t)$ into the disutility of group $m$ travelers.

Denote $\tilde{\eta}^{rs}_{mp}(t)$ as the actual route travel disutility for group $m$ travelers from origin $r$ to destination $s$ at time $t$. Also denote $\tilde{\pi}^{rs}_m(t)$ as the minimal actual route travel disutility for group $m$ travelers from origin $r$ to destination $s$ at time $t$, and $\bar{\pi}^{rs}_m(t)$ as the corresponding actual route travel time for group $m$ travelers departing origin $r$ to destination $s$ at time $t$. We also need to use a recursive formula to compute the route travel disutility $\tilde{\eta}^{rs}_{mp}(t)$ for all allowable routes. Assume route $p$ consists of nodes $(r, 1, 2, \cdots, i, \cdots, s)$. Denote $\tilde{\eta}^{rj}_{mp}(t)$ as the travel disutility *actually* experienced over route $p$ from origin $r$ to node $j$ by vehicles departing origin $r$ at time $t$. Then, a recursive formula for route travel disutility $\tilde{\eta}^{rs}_{mp}(t)$ is:

$$\tilde{\eta}^{rj}_{mp}(t) = \tilde{\eta}^{r(j-1)}_{mp}(t) + \tilde{\tau}_{ma}[t + \eta^{r(j-1)}_p(t)] \qquad \forall m, p, r, j; j = 1, 2, \cdots, s;$$

where link $a = (j-1, j)$ and time $[t + \eta^{r(j-1)}_p(t)]$ is the arrival time instant at link $a$ for group $m$ travelers. Recall the definition of the multi-group travel-cost-based ideal DUO state in Section 13.2 which also applies to the link-cost-based problem.

> **Multi-Group Travel-Cost-Based Ideal DUO State:** *If, for each group and each O-D pair at each instant of time, the actual travel* disutilities *for all routes that are being used equal the minimal actual route travel* disutility, *the dynamic traffic flow over the network is in a multi-group ideal dynamic user-optimal state.*

We now write the equivalent mathematical inequalities for the above definition using link variables, in contrast to the route-based formulation in Section 13.2. For group $m$ on link $a = (i, j)$, the minimal travel disutility $\tilde{\pi}_m^{rj^*}(t)$ from origin $r$ to $j$ should be equal or less than the minimal travel disutility $\tilde{\pi}_m^{ri^*}(t)$ from origin $r$ to $i$ plus the actual link travel disutility $\tilde{\tau}_{ma}[t + \bar{\pi}_m^{ri}(t)]$ at time instant $[t + \bar{\pi}_m^{ri}(t)]$, where this time instant is the earliest clock time when the flow departing origin $r$ at time $t$ can enter link $a$ over the minimal travel disutility route. It follows that

$$\tilde{\pi}_m^{ri^*}(t) + \tilde{\tau}_{ma}[t + \bar{\pi}_m^{ri^*}(t)] \geq \tilde{\pi}_m^{rj^*}(t) \qquad \forall m, a = (i, j), r.$$

If, for each group $m$ and each O-D pair $rs$, any departure flow from origin $r$ at time $t$ enters link $a$ at the earliest clock time $[t + \bar{\pi}_m^{ri^*}(t)]$, or $u_{ma}^{rs}[t + \bar{\pi}_m^{ri^*}(t)] > 0$, then the multi-group ideal DUO route choice conditions require that link $a$ is on the minimal travel disutility route. In other words, the minimal travel disutility $\tilde{\pi}_m^{rj^*}(t)$ from origin $r$ to $j$ should equal to the minimal travel disutility $\tilde{\pi}_m^{ri^*}(t)$ from origin $r$ to $i$ plus the actual link travel disutility $\tilde{\tau}_{ma}[t + \bar{\pi}_m^{ri^*}(t)]$ at time instant $[t + \bar{\pi}_m^{ri^*}(t)]$. It follows that

$$\tilde{\pi}_m^{rj^*}(t) = \tilde{\pi}_m^{ri^*}(t) + \tilde{\tau}_{ma}[t + \bar{\pi}_m^{ri^*}(t)], \quad \text{if } u_{ma}^{rs^*}[t + \bar{\pi}_m^{ri^*}(t)] > 0, \quad \forall m, a = (i, j), r, s.$$

The above equation is also equivalent to the following:

$$\left[\tilde{\pi}_m^{ri^*}(t) + \tilde{\tau}_{ma}[t + \bar{\pi}_m^{ri^*}(t)] - \tilde{\pi}_m^{rj^*}(t)\right] u_{ma}^{rs^*}[t + \bar{\pi}_m^{ri^*}(t)] = 0, \quad \forall m, a = (i, j), r, s.$$

Thus, the multi-group link-cost-based ideal DUO route choice conditions can be summarized as follows:

$$\tilde{\pi}_m^{ri^*}(t) + \tilde{\tau}_{ma}[t + \bar{\pi}_m^{ri^*}(t)] - \tilde{\pi}_m^{rj^*}(t) \geq 0 \qquad \forall m, a = (i, j), r; \qquad (13.73)$$

$$\left[\tilde{\pi}_m^{ri^*}(t) + \tilde{\tau}_{ma}[t + \bar{\pi}_m^{ri^*}(t)] - \tilde{\pi}_m^{rj^*}(t)\right] u_{ma}^{rs^*}[t + \bar{\pi}_m^{ri^*}(t)] = 0$$
$$\forall m, a = (i, j), r, s; \qquad (13.74)$$

$$u_{ma}^{rs}[t + \bar{\pi}_m^{ri^*}(t)] \geq 0 \qquad \forall m, a = (i, j), r, s. \qquad (13.75)$$

Similar to Lemma 13.1, we can easily prove that the multi-group link-cost-based ideal DUO route choice conditions imply the multi-group route-cost based ideal DUO route choice conditions. The detailed proof is omitted here.

### 13.4.2 The Multi-Group Link-Cost-Based VI Model

For group $m$ travelers, denote $\tilde{\Omega}_{ma}^{rj^*}(t)$ as the difference of the minimal travel disutility from $r$ to $j$ and the travel disutility from $r$ to $j$ via minimal travel disutility route from $r$ to $i$ and link $a$ for vehicles departing from origin $r$ at time $t$. It follows that

$$\tilde{\Omega}_{ma}^{rj^*}(t) = \tilde{\pi}_m^{ri^*}(t) + \tilde{\tau}_{ma}[t + \bar{\pi}_m^{ri^*}(t)] - \tilde{\pi}_m^{rj^*}(t) \qquad \forall m, a, r; a = (i, j). \quad (13.76)$$

We then rewrite the multi-group link-cost-based ideal DUO route choice conditions as follows:

$$\tilde{\Omega}_{ma}^{rj*}(t) \geq 0 \qquad \forall m, a=(i,j), r; \tag{13.77}$$

$$u_{ma}^{rs*}[t+\bar{\pi}_m^{ri*}(t)]\,\tilde{\Omega}_{ma}^{rj*}(t)=0 \qquad \forall m, a=(i,j), r, s; \tag{13.78}$$

$$u_{ma}^{rs}[t+\bar{\pi}_m^{ri*}(t)] \geq 0 \qquad \forall m, a=(i,j), r, s. \tag{13.79}$$

The equivalent variational inequality formulation of multi-group link-cost-based ideal DUO route choice conditions (13.77)-(13.79) may be stated as follows.

*Theorem 13.4.* The dynamic traffic flow pattern satisfying network constraint set (13.33)-(13.44) is in a multi-group link-cost-based ideal DUO route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \sum_{rs}\sum_{ma} \tilde{\Omega}_{ma}^{rj*}(t)\,\left\{u_{ma}^{rs}[t+\bar{\pi}_m^{ri*}(t)] - u_{ma}^{rs*}[t+\bar{\pi}_m^{ri*}(t)]\right\}\,dt \;\geq\; 0 \tag{13.80}$$

The proof follows in a manner similar to Section 13.3.2 (single group, link-time-based) except that the arrival time at each link $[t+\pi_m^{ri*}(t)]$ is determined by the travel time $\bar{\pi}_m^{ri*}(t)$ over the minimal cost route in stead of the minimal travel time $\pi_m^{ri*}(t)$.

## 13.5 Relationships Between VI Models and Optimization Models

We now consider the relationship between VI models and optimization models. As in Chapter 12, we will not discuss each VI model in this chapter. As an example, we focus our analysis on the link-time-based VI model for the ideal DUO route choice problem. We show in the following that the VI model can be reformulated as an optimal control problem which is similar to the optimal control models presented in Chapter 5. Similar analyses can be performed for the other VI models for the various ideal DUO route choice problems.

In this section, we discuss the following VI problem:

$$\int_0^T \sum_{rs}\sum_{a} \Omega_a^{rj*}(t)\,\left\{u_a^{rs}[t+\pi^{ri*}(t)] - u_a^{rs*}[t+\pi^{ri*}(t)]\right\}\,dt \;\geq\; 0 \tag{13.81}$$

To simplify our analysis, we assume the time period $[0,T]$ is long enough so that all departure flows can be cleared at final time $T$. In other words, any positive departure from origin $r$ at time $t$ will arrive at destination $s$ at time $t+\pi^{rs*}(t) \leq T$.

We consider a simplified link travel time function as follows:

$$\tau_a(t) = \tau_a[x_a(t), u_a(t)] \qquad \forall a. \tag{13.82}$$

As shown in Chapter 16, some link travel time functions for arterial roads and freeway segments do not dependent on exit flow explicitly. Thus, the above assumption is reasonable.

In order to present a partitionable VI, we need to transform the original VI into a partitionable VI using some new definitions as follows. For link $a$ and O-D pair $rs$ at time instant $[t + \pi^{ri^*}(t)]$, denote an auxiliary link travel time function $\lambda_a^{rs^*}[t + \pi^{ri^*}(t)]$ as

$$\lambda_a^{rs^*}[t + \pi^{ri^*}(t)] = -\pi^{rj^*}(t) \qquad \forall a, r, s; a = (i, j). \tag{13.83}$$

Recall that in Section 13.3, for link $a$ and O-D pair $rs$ at time instant $[t + \pi^{ri^*}(t)]$, we defined an auxiliary link travel time function

$$\Omega_a^{rj^*}(t) = \pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)] - \pi^{rj^*}(t) \qquad \forall a, r; a = (i, j). \tag{13.84}$$

Substituting equation (13.83) into equation (13.84), we obtain

$$\Omega_a^{rj^*}(t) = \pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)] + \lambda_a^{rs^*}[t + \pi^{ri^*}(t)] \geq 0$$

$$\forall a, r, s; a = (i, j). \tag{13.85}$$

For link $a$ and O-D pair $rs$ at time instant $[t + \pi^{ri^*}(t)]$, we define a related auxiliary link travel time function as

$$\bar{\Omega}_a^{rj^*}(t) = -\lambda_a^{rs^*}[t + \pi^{ri^*}(t)] + \pi^{rj^*}(t) = 0 \qquad \forall a, r, s; a = (i, j). \tag{13.86}$$

Using the above new definitions, link-time-based ideal DUO route choice conditions (13.56)-(13.58) are rewritten as equivalent conditions as follows.

$$\Omega_a^{rj^*}(t) \geq 0 \qquad \forall a = (i, j), r, s; \tag{13.87}$$

$$\bar{\Omega}_a^{rj^*}(t) = 0 \qquad \forall a = (i, j), r, s; \tag{13.88}$$

$$u_a^{rs^*}[t + \pi^{ri^*}(t)]\, \Omega_a^{rj^*}(t) = 0 \qquad \forall a = (i, j), r, s; \tag{13.89}$$

$$v_a^{s^*}[t + \pi^{ri^*}(t)]\, \bar{\Omega}_a^{rj^*}(t) = 0 \qquad \forall a = (i, j), r, s; \tag{13.90}$$

$$u_a^{rs}[t + \pi^{ri^*}(t)] \geq 0 \qquad \forall a = (i, j), r, s; \tag{13.91}$$

$$v_a^{rs}[t + \pi^{ri^*}(t)] \geq 0 \qquad \forall a = (i, j), r, s. \tag{13.92}$$

We note that equation (13.87) is equivalent to equation (13.56) and equation (13.89) is equivalent to equation (13.57). Then, the link-time-based variational inequality (13.81) in Theorem 13.3 can be restated as an equivalent VI in the following theorem.

*Theorem 13.5.* The dynamic traffic flow pattern satisfying network constraint set (13.5)-(13.16) is in a travel-time-based ideal DUO route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \sum_{rs} \sum_a \Big\{ \Omega_a^{rj^*}(t) \ \Big[u_a^{rs}[t+\pi^{ri^*}(t)] - u_a^{rs^*}[t+\pi^{ri^*}(t)]\Big]$$
$$+ \quad \bar{\Omega}_a^{rj^*}(t) \ \Big[v_a^{rs}[t+\pi^{ri^*}(t)] - v_a^{rs^*}[t+\pi^{ri^*}(t)]\Big]\Big\} \ dt \ \geq \ 0 \qquad (13.93)$$

The second term in the above variational inequality equals 0. It is placed within the VI so that the reformulation of the VI as an optimal control problem can be performed more easily. Therefore, the above VI is equivalent to link-time-based ideal DUO route choice conditions (13.56)-(13.58) or (13.87)-(13.92). The proofs of necessity and sufficiency are straightforward and not given here.

Substituting definitions (13.85) and (13.86) into equation (13.93), variational inequality (13.93) is equivalent to

$$\int_0^T \sum_{rs} \sum_a \ \Big\{ \Big[\pi^{ri^*}(t) + \tau_a[t+\pi^{ri^*}(t)] + \lambda_a^{rs^*}[t+\pi^{ri^*}(t)]\Big]$$
$$\cdot \Big[u_a^{rs}[t+\pi^{ri^*}(t)] - u_a^{rs^*}[t+\pi^{ri^*}(t)]\Big]$$
$$+ \quad \Big[-\lambda_a^{rs^*}[t+\pi^{ri^*}(t)] + \pi^{rj^*}(t)\Big] \ \Big[v_a^{rs}[t+\pi^{ri^*}(t)] - v_a^{rs^*}[t+\pi^{ri^*}(t)]\Big]\Big\} \ dt$$
$$= \quad \int_0^T \sum_r \sum_a \ \tau_a[t+\pi^{ri^*}(t)] \ \Big[u_a^r[t+\pi^{ri^*}(t)] - u_a^{r^*}[t+\pi^{ri^*}(t)]\Big] \ dt$$
$$+ \quad \int_0^T \sum_{rs} \sum_a \ \Big\{ \Big[\pi^{ri^*}(t) + \lambda_a^{rs^*}[t+\pi^{ri^*}(t)]\Big]$$
$$\cdot \Big[u_a^{rs}[t+\pi^{ri^*}(t)] - u_a^{rs^*}[t+\pi^{ri^*}(t)]\Big]$$
$$+ \quad \Big[-\lambda_a^{rs^*}[t+\pi^{ri^*}(t)] + \pi^{rj^*}(t)\Big]$$
$$\cdot \Big[v_a^{rs}[t+\pi^{ri^*}(t)] - v_a^{rs^*}[t+\pi^{ri^*}(t)]\Big]\Big\} \ dt \ \geq \ 0 \qquad (13.94)$$

We now show that a double relaxation or diagonalization procedure can be designed so that the above VI can be formulated as an optimal control model in each relaxation iteration. At the first-level relaxation, we assume that the actual link travel time $\tau_a(t)$ in the flow propagation constraints and the resulting minimal actual travel times $\pi^{ri}(t)$ in the above variational inequality are fixed temporarily at each relaxation iteration. Then, the cross-effects of flow variables at different time instants can be separated at each iteration. We define a new time variable $\xi_a^r = t + \pi^{ri^*}(t)$ where $\pi^{ri^*}(t)$ is fixed temporarily at each relaxation iteration. Suppose there are $R$ origins. Then, for each link

$a$ and origin $r$ at time instant $\xi_a^r$, we have

$$u_a(\xi_a^r) = u_a^1(\xi_a^r) + u_a^2(\xi_a^r) + \cdots + u_a^R(\xi_a^r) \tag{13.95}$$

The actual link travel time $\tau_a(\xi_a^r)$ at time instant $\xi_a^r$ can be expressed as

$$\tau_a(\xi_a^r) = \tau_a[u_a^1(\xi_a^r), \cdots, u_a^R(\xi_a^r), x_a^1(\xi_a^r), \cdots, x_a^R(\xi_a^r)] \tag{13.96}$$

Since the variational inequality for the ideal DUO route choice is different from the variational inequality for instantaneous DUO route choice, the cross-effects of origin-specific link flow variables are asymmetric and can not be eliminated. Thus, we need to design a second-level relaxation for this VI. Thus, we can derive a similar optimal control problem as in the instantaneous case. To this end, we fix temporarily all other link flow variables $u_a(\xi_a^{r'})$, $x_a(\xi_a^{r'})$ $(r' \neq r)$ in equation (13.96) at each second-level relaxation iteration. Denote $n$ as the total number of links in the network. Then, a Jacobian submatrix of the actual link travel time $\tau_a(\xi_a^r)$ with respect to the inflow $u_a^r(\xi_a^r)$ for each time instant $\xi_a^r$ can be written as

$$\nabla_u \tau_a^r(\xi_a^r) = \begin{bmatrix} \frac{\partial \tau_1(\xi_a^r)}{\partial u_1^r(\xi_a^r)} & 0 & \cdots & 0 \\ 0 & \frac{\partial \tau_2(\xi_a^r)}{\partial u_2^r(\xi_a^r)} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \frac{\partial \tau_n(\xi_a^r)}{\partial u_n^r(\xi_a^r)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial \tau_1(\xi_a^r)}{\partial u_1(\xi_a^r)} & 0 & \cdots & 0 \\ 0 & \frac{\partial \tau_2(\xi_a^r)}{\partial u_2(\xi_a^r)} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \frac{\partial \tau_n(\xi_a^r)}{\partial u_n(\xi_a^r)} \end{bmatrix}$$

where the cross-effects of all other link flow variables $u_a(\xi_a^r)$, $x_a(\xi_a^r)$ $(r' \neq r)$ can be eliminated. Obviously, the above matrix is symmetric. We note that $x_a(\xi_a^r)$ does not enter the Jacobian submatrix because it is a state variable.

At each second-level relaxation iteration, another Jacobian sub-matrix of the actual link travel time $\tau_a(\xi_a^r)$ with respect to the exiting flow $v_a^r(\xi_a^r)$ for each time instant $\xi_a^r$ can be written as

$$\nabla_v \tau_a(\xi_a^r) = \begin{bmatrix} \frac{\partial \tau_1(\xi_a^r)}{\partial v_1^r(\xi_a^r)} & 0 & \cdots & 0 \\ 0 & \frac{\partial \tau_2(\xi_a^r)}{\partial v_2^r(\xi_a^r)} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \frac{\partial \tau_n(\xi_a^r)}{\partial v_n^r(\xi_a^r)} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial \tau_1(\xi_a^r)}{\partial v_1(\xi_a^r)} & 0 & \cdots & 0 \\ 0 & \frac{\partial \tau_2(\xi_a^r)}{\partial v_2(\xi_a^r)} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & \frac{\partial \tau_n(\xi_a^r)}{\partial v_n(\xi_a^r)} \end{bmatrix} = \mathbf{0}$$

which is also symmetric. Then, at each time instant $\xi_a^r$, there is an optimization problem which is equivalent to variational inequality (13.93), as follows:

$$\begin{aligned} \text{Min} \quad & \sum_r \sum_a \int_0^{u_a^r(\xi_a^r)} \tau_a[u_a^1(\xi_a^r), \cdots, \omega, \cdots, u_a^R(\xi_a^r), x_a^1(\xi_a^r), \cdots, x_a^R(\xi_a^r)]\, d\omega \\ + & \sum_{rs} \sum_a u_a^{rs}(\xi_a^r) \left[\pi^{ri^*}(\xi_a^r) + \lambda_a^{rs^*}(\xi_a^r)\right] \\ + & \sum_{rs} \sum_a v_a^{rs}(\xi_a^r) \left[-\lambda_a^{rs^*}(\xi_a^r) + \pi^{rj^*}(\xi_a^r)\right] \end{aligned} \tag{13.97}$$

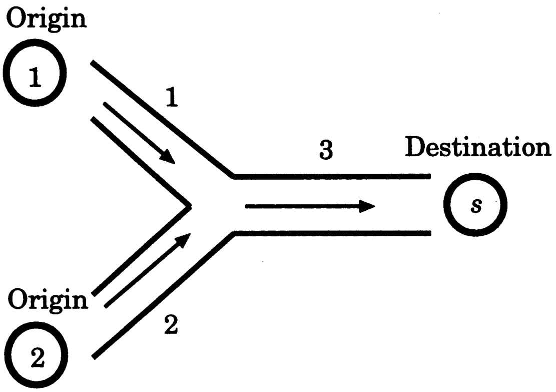



Figure 13.1: Example Network with Two Origins

We now use the simple network in Figure 13.1 to illustrate the above analysis. The actual travel time $\tau_a(\xi_a^r)$ on link 3 at time instant $\xi_a^1$ can be expressed as

$$\tau_3(\xi_a^1) = \tau_3[u_a^1(\xi_3^1), u_3^2(\xi_3^1), x_3^1(\xi_3^1), x_3^2(\xi_3^1)] \tag{13.98}$$

At each second-level relaxation iteration, link flow variables $u_3^2(\xi_3^1)$ and $x_3^2(\xi_3^1)$ associated with origin 2 have to be fixed temporarily in equation (13.98). Then, we can find an optimization problem which is equivalent to the variational inequality. In the optimization problem, the first two terms for link 3 at each time instant $\xi_3^1$ are

$$\text{Min} \quad \int_0^{u_3^1(\xi_3^1)} \tau_3[\omega, u_3^2(\xi_3^1), x_3^1(\xi_3^1), x_3^2(\xi_3^1)]\, d\omega$$

$$+ \int_0^{u_3^2(\xi_3^1)} \tau_3[u_3^1(\xi_3^1), \omega, x_3^1(\xi_3^1), x_3^2(\xi_3^1)]\, d\omega \qquad (13.99)$$

Note that in both terms, all flow variables are defined for time $\xi_3^1$.

Reorganizing equation (13.97) based on each node $j$, we have

$$\begin{aligned}
\text{Min} \quad & \sum_r \sum_a \int_0^{u_a^r(\xi_a^r)} \tau_a[u_a^1(\xi_a^r), \cdots, \omega, \cdots, u_a^R(\xi_a^r), x_a^1(\xi_a^r), \cdots, x_a^R(\xi_a^r)]\, d\omega \\
+ & \sum_{rs} \sum_a \lambda_a^{s^*}(\xi_a^r)[u_a^{rs}(\xi_a^r) - v_a^{rs}(\xi_a^r)] - \sum_s \pi^{rs^*}(\xi_a^r) \sum_{a \in A(r)} u_a^{rs}(\xi_a^r) \\
+ & \sum_s \sum_{j \neq rs} \sigma^{js^*}(\xi_a^r)[\sum_{a \in B(j)} v_a^s(\xi_a^r) - \sum_{a \in A(j)} u_a^s(\xi_a^r)] \\
+ & \sum_{s \neq r} \sigma^{ss^*}(\xi_a^r) \sum_{a \in B(s)} v_a^s(\xi_a^r) \\
= \quad & \sum_r \sum_a \int_0^{u_a^r(\xi_a^r)} \tau_a[u_a^1(\xi_a^r), \cdots, \omega, \cdots, u_a^R(\xi_a^r), x_a^1(\xi_a^r), \cdots, x_a^R(\xi_a^r)]\, d\omega \\
+ & \sum_{rs} \sum_a \lambda_a^{s^*}(\xi_a^r)[u_a^{rs}(\xi_a^r) - v_a^{rs}(\xi_a^r)] - \sum_{r \neq s} \sum_s \sigma^{rs^*}(\xi_a^r) \sum_{a \in A(r)} u_a^{rs}(\xi_a^r) \\
+ & \sum_{rs} \sum_{j \neq rs} \sigma^{js^*}(\xi_a^r)[\sum_{a \in B(j)} v_a^{rs}(\xi_a^r) - \sum_{a \in A(j)} u_a^{rs}(\xi_a^r)] \\
+ & \sum_r \sum_{s \neq r} \sigma^{ss^*}(\xi_a^r) \sum_{a \in B(s)} v_a^{rs}(\xi_a^r)
\end{aligned}$$

The above equation is equivalent to the following partial Hamiltonian function with flow conservation constraints only:

$$\begin{aligned}
\mathcal{H} = & \sum_r \sum_a \int_0^{u_a^r(\xi_a^r)} \tau_a[u_a^1(\xi_a^r), \cdots, \omega, \cdots, u_a^R(\xi_a^r), x_a^1(\xi_a^r), \cdots, x_a^R(\xi_a^r)]\, d\omega \\
+ & \sum_{rs} \sum_a \lambda_a^{s^*}(\xi_a^r)[u_a^{rs}(\xi_a^r) - v_a^{rs}(\xi_a^r)] \\
+ & \sum_{r \neq s} \sum_s \sigma^{rs^*}(\xi_a^r)[f^{rs}(\xi_a^r) - \sum_{a \in A(r)} u_a^{rs}(\xi_a^r)] \\
+ & \sum_{rs} \sum_{j \neq rs} \sum_p \sigma^{js^*}(\xi_a^r)[\sum_{a \in B(j)} v_{ap}^{rs}(\xi_a^r) - \sum_{a \in A(j)} u_{ap}^{rs}(\xi_a^r)] \\
+ & \sum_r \sum_{s \neq r} \sigma^{ss^*}(\xi_a^r)[\sum_{a \in B(s)} \sum_p v_{ap}^{rs}(\xi_a^r) - e^{rs}(\xi_a^r)]
\end{aligned}$$

Note the flow propagation constraint is not included. Thus, we can simply state the objective function of the optimization program as

$$\text{Min} \quad \int_0^T \sum_{r,a} \int_0^{u_a^r(\xi_a^r)} \tau_a[u_a^1(\xi_a^r), \cdots, \omega, \cdots, u_a^R(\xi_a^r), x_a^1(\xi_a^r), \cdots, x_a^R(\xi_a^r)]\, d\omega\, dt \quad (13.100)$$

because other terms in the partial Hamiltonian function are associated with link and node flow conservation equations. Note that $\xi_a^r = t + \pi^{ri^*}(t)$ where $\pi^{ri^*}(t)$ is fixed temporarily at each second-level relaxation iteration. Also note that all flow variables at time instants $\xi_a^r > T$ are zero. Therefore, link-time-based variational inequality (13.81) can be reformulated as an optimal control problem with objective function (13.100) and constraints (13.5)-(13.16) at each double relaxation iteration. In other words, we have demonstrated that our original optimal control model in Chapter 5 is a special relaxation or diagonalization problem of VI formulation (13.81). We note that the actual link travel time $\tau_a(t)$ is fixed temporarily in the flow propagation constraints at each first-level relaxation iteration. We also note that the relaxation for the ideal DUO VI model is different from the instantaneous DUO VI model. The double relaxation procedure is summarized in Figure 13.2.

## 13.6 Notes

If there is only one origin, the cross-effects between origins can be eliminated. Thus, the second-level relaxation can be dropped and the above optimal control problem will be identical to the optimal control model in Chapter 5 for instantaneous DUO route choice. Furthermore, if there is only one destination, a similar conclusion can be drawn because the VI formulation and the relaxation for the above origin-based model applies to the destination-based model as well. This conclusion has important implications for freeway corridor models when the CBD is considered as one destination. We speculate that under one destination, our origin-based model would also lead to the conclusion that both instantaneous DUO and ideal DUO models yield the same results. However, this conclusion needs more theoretical study and a numerical demonstration.

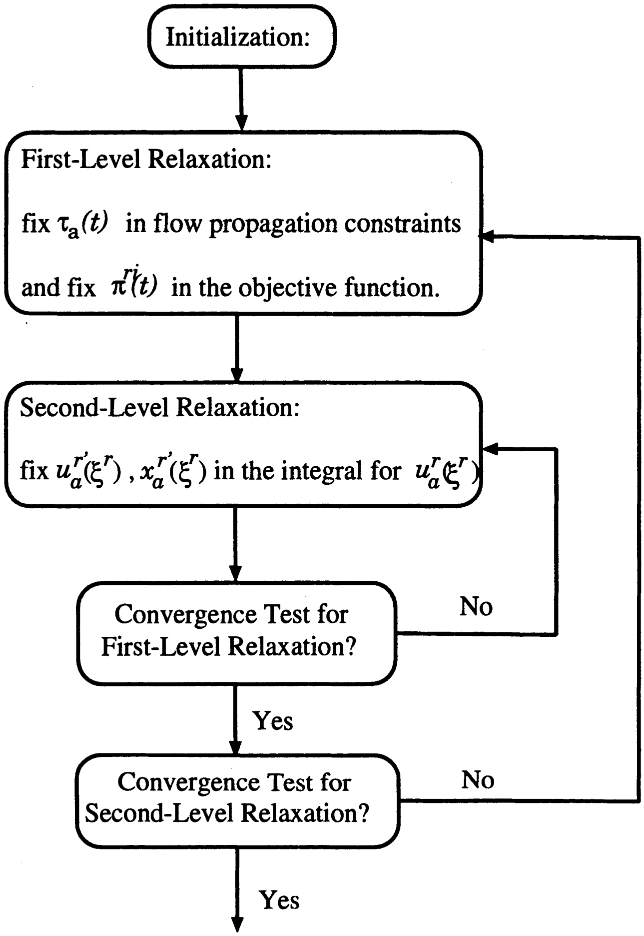



Figure 13.2: Flowchart of the Double Relaxation Procedure

# Chapter 14

# Variational Inequality Models of Dynamic Departure Time/ Route Choice Problems

In this chapter, we consider an ideal situation where all travelers are equipped with navigation devices and fully comply with the dynamic user-optimal criterion when choosing routes, departure times and modes. We first present a dynamic, user-optimal departure time/route choice model for a general network with multiple origin-destination pairs. We model this choice problem by specifying that a given number of travelers are ready for departure between each origin-destination pair at time 0. However, their departure times may be delayed to reduce their overall travel costs. A route-based variational inequality model for joint departure time/route choice is presented in Section 14.1. In a parallel fashion, a link-based variational inequality model is proposed in Section 14.2. The relationship between the variational inequality models and the optimization models is discussed in Section 14.3.

## 14.1 A Route-Based VI Model of Departure Time/Route Choice

A number of vehicles are ready to depart at an initial time 0, but these drivers may prefer to delay their departure times in order to reduce their driving times. Drivers are assumed to make their departure time choices so as to minimize their individual disutility functions defined on travel time and pre-trip delay. Of course, the change of departure flow rates will change the traffic in the network so that the travel times for other travelers could change.

### 14.1.1 Route-Based Conditions

We first consider the joint departure time/route choice conditions. A disutility function $\mathcal{U}_p^{rs}(t)$ based on departure times is defined for travelers departing from origin $r$ to destination $s$ over route $p$ at time $t$. This disutility function represents a weighted sum of:

1. waiting time at the origin node;
2. driving time during the trip;
3. a bonus for early arrival or a penalty for late arrival.

Consider the flow which originates at node $r$ at time $t$ and is destined for node $s$. There is a set of routes $\{p\}$ between O-D pair $rs$. Define $\eta_p^{rs}(t)$ as the travel time actually experienced over route $p$ by vehicles departing origin $r$ toward destination $s$ at time $t$. We use a recursive formula to compute the route travel time $\eta_p^{rs}(t)$ for each allowable route. Assume route $p$ consists of nodes $(r, 1, \cdots, i-1, i, \cdots, s)$. Denote $\eta_p^{ri}(t)$ as the travel time actually experienced over route $p$ from origin $r$ to node $i$ by vehicles departing from origin $r$ at time $t$. Then, a recursive formula for route travel time $\eta_p^{rs}(t)$ is:

$$\eta_p^{ri}(t) = \eta_p^{r(i-1)}(t) + \tau_a[t + \eta_p^{r(i-1)}(t)] \qquad \forall p, r, i; i = 1, 2, \cdots, s;$$

where link $a = (i-1, i)$.

We define one unit of disutility to equal one unit of in-vehicle driving time, and one unit of waiting time prior to departure to be equivalent to $\alpha$ units of disutility ($\alpha \leq 1$); $\alpha$ could be negative since staying at home may have positive utility. Since all travelers are able to depart at time 0, $\alpha t$ is the disutility for a departure at time $t$ due to waiting. Furthermore, we assume there is a desired arrival time interval $[t_{rs}^* - \Delta_{rs}, t_{rs}^* + \Delta_{rs}]$ for travelers at each destination $s$, where $t_{rs}^*$ is the center of the required arrival time interval (e.g. work starting time) associated with travelers departing from origin $r$ toward destination $s$. $\Delta_{rs}$ represents the arrival time flexibility at destination $s$ for travelers departing from origin $r$ toward destination $s$.

The disutility function for the route-based joint departure time and route choice problem is constructed as

$$\mathcal{U}_p^{rs}(t) = \alpha t + \eta_p^{rs}(t) + \mathcal{V}_p^{rs}[t, \eta_p^{rs}(t); t_{rs}^*] \qquad \forall p, r, s, \tag{14.1}$$

where $t$ is the departure time of travelers and $\mathcal{V}_p^{rs}[t, \eta_p^{rs}(t); t_{rs}^*]$ is the disutility for early or late arrival which is defined as follows

$$\mathcal{V}_p^{rs}[t, \eta_p^{rs}(t); t_{rs}^*] =$$

$$\begin{cases} \gamma_1[t + \eta_p^{rs}(t) - t_{rs}^* + \Delta_{rs}^*]^2 & \text{if } t + \eta_p^{rs}(t) < t_{rs}^* - \Delta_{rs}^* \quad \text{(early arrival)} \\ 0 & \text{if } |t + \eta_p^{rs}(t) - t_{rs}^*| \leq \Delta_{rs}^* \\ \gamma_2[t + \eta_p^{rs}(t) - t_{rs}^* - \Delta_{rs}^*]^2 & \text{if } t + \eta_p^{rs}(t) > t_{rs}^* + \Delta_{rs}^* \quad \text{(late arrival)} \end{cases}$$

where $t$ is the departure time of travelers and $\gamma_1$ and $\gamma_2$ are parameters ($\gamma_2 \gg \alpha$). $\gamma_1$ is negative because early arrival should be encouraged rather than discouraged. This arrival time disutility function is shown in Figure 14.1.

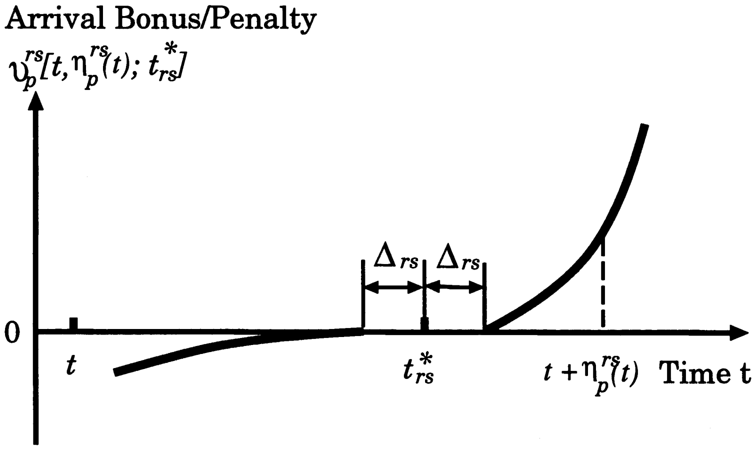



Figure 14.1: Arrival Time Disutility

The dynamic user-optimal departure time/route choice conditions require that for each O-D pair $rs$ at any time $t$, if there is a positive departure flow $f_p^{rs^*}(t) > 0$ over route $p$, the disutility $\mathcal{U}_p^{rs^*}(t)$ for route $p$ must equal the minimal $rs$ disutility $\mathcal{U}_{min}^{rs^*}$ over time $t$. Furthermore, if the departure flow $f_p^{rs^*}(t)$ over route $p$ equals 0 at time $t$, the disutility $\mathcal{U}_p^{rs^*}(t)$ over route $p$ at time $t$ must be greater than or equal to the minimal $rs$ disutility $\mathcal{U}_{min}^{rs^*}$. The route-based DUO departure time/route choice conditions can be written as

$$\mathcal{U}_p^{rs^*}(t) - \mathcal{U}_{min}^{rs^*} \geq 0 \qquad \forall p, r, s; \tag{14.2}$$

$$f_p^{rs^*}(t)\ [\mathcal{U}_p^{rs^*}(t) - \mathcal{U}_{min}^{rs^*}] = 0 \qquad \forall p, r, s; \tag{14.3}$$

$$f_p^{rs}(t) \geq 0 \qquad \forall p, r, s. \tag{14.4}$$

where the asterisk denotes that the travel disutility is computed using DUO departure flows and route flows.

## 14.1.2 Dynamic Network Constraints

In this section, the constraint set for our dynamic user-optimal departure time/route choice problem is first summarized as follows.

Relationships between state and control variables:

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \tag{14.5}$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall r, s, p; \tag{14.6}$$

$$\frac{dF_p^{rs}(t)}{dt} = f_p^{rs}(t) \qquad \forall r, s, p; \tag{14.7}$$

Flow conservation constraints:

$$f_p^{rs}(t) = \sum_{a \in A(r)} \delta_{ap}^{rs} u_{ap}^{rs}(t) \qquad \forall p, r, s; \tag{14.8}$$

$$e_p^{rs}(t) = \sum_{a \in B(s)} \delta_{ap}^{rs} v_{ap}^{rs}(t) \qquad \forall p, r, s; \tag{14.9}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \tag{14.10}$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[t + \tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall r, s, p, j; a \in B(j); j \neq r; \tag{14.11}$$

Definitional constraints:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \tag{14.12}$$

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \qquad \sum_{rs} x_a^{rs}(t) = x_a(t), \qquad \forall a, r, s; \tag{14.13}$$

$$\sum_p E_p^{rs}(t) = E^{rs}(t), \qquad \sum_p F_p^{rs}(t) = F^{rs}(t), \qquad \forall r, s; \tag{14.14}$$

$$\sum_p f_p^{rs}(t) = f^{rs}(t), \qquad \sum_p e_p^{rs}(t) = e^{rs}(t), \qquad \forall r, s; \tag{14.15}$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \quad u_{ap}^{rs}(t) \geq 0, \quad v_{ap}^{rs}(t) \geq 0 \quad \forall a, p, r, s; \tag{14.16}$$

$$e_p^{rs}(t) \geq 0, \quad f_p^{rs}(t) \geq 0, \quad E_p^{rs}(t) \geq 0, \quad F_p^{rs}(t) \geq 0 \quad \forall p, r, s; \tag{14.17}$$

$$f^{rs}(t) \geq 0, \qquad F^{rs}(t) \geq 0 \qquad \forall r, s; \tag{14.18}$$

Boundary conditions:

$$F^{rs}(T) \text{ given} \qquad \forall r, s; \tag{14.19}$$

$$E_p^{rs}(0) = 0, \quad F_p^{rs}(0) = 0 \quad \forall p, r, s; \quad x_{ap}^{rs}(0) = 0, \quad \forall a, p, r, s. \tag{14.20}$$

The first three constraints (14.5)-(14.7) are state equations for each link $a$ and for the cumulative effects at origins and destinations. Equations (14.8)-(14.10) are flow conservation constraints at each node including origins and destinations. Other constraints include flow propagation constraints, definitional constraints, nonnegativity and boundary conditions. In summary, the control variables are $u_{ap}^{rs}(t)$, $v_{ap}^{rs}(t)$, $e_p^{rs}(t)$, and $f_p^{rs}(t)$; the state variables are $x_{ap}^{rs}(t)$, $E_p^{rs}(t)$, and $F_p^{rs}(t)$; the functionals are $\pi^{rs}(t)$.

### 14.1.3 The Route-Based VI Model

The equivalent variational inequality formulation of route-based DUO departure time/route choice conditions (14.2)-(14.4) may be stated as follows.

*Theorem 14.1.* The dynamic traffic flow satisfying constraints (14.5)-(14.20) is in a route-based DUO departure time/route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \left\{ \sum_{rs} \sum_p \mathcal{U}_p^{rs*}(t) \left[ f_p^{rs}(t) - f_p^{rs*}(t) \right] \right\} dt \geq 0 \qquad (14.21)$$

*Proof of Necessity.*

We need to prove that route-based DUO departure time/route choice conditions (14.2)-(14.4) imply variational inequality (14.21). Multiplying equations (14.2) and (14.4), we have

$$f_p^{rs}(t) \, \{\mathcal{U}_p^{rs*}(t) - \mathcal{U}_{min}^{rs*}\} \geq 0 \qquad \forall p, r, s. \qquad (14.22)$$

We subtract equation (14.3) from equation (14.22) and obtain

$$[\mathcal{U}^{rs*}(t) - \mathcal{U}_{min}^{rs*}] \, [f_p^{rs}(t) - f_p^{rs*}(t)] \geq 0 \qquad \forall p, r, s. \qquad (14.23)$$

Summing equation (14.23) for all routes $p$ and all O-D pairs $rs$, it follows that

$$\sum_{rs} \sum_p [\mathcal{U}_p^{rs*}(t) - \mathcal{U}_{min}^{rs*}] \, [f_p^{rs}(t) - f_p^{rs*}(t)] \geq 0 \qquad (14.24)$$

Integrating the above equation (14.24) from time 0 to $T$, we have

$$\int_0^T \sum_{rs} \sum_p [\mathcal{U}_p^{rs*}(t) - \mathcal{U}_{min}^{rs*}] \, [f_p^{rs}(t) - f_p^{rs*}(t)] \, dt \geq 0 \qquad (14.25)$$

or

$$\int_0^T \sum_{rs} \left\{ \sum_p \mathcal{U}_p^{rs*}(t) \, [f_p^{rs}(t) - f_p^{rs*}(t)] \right.$$

$$\left. - \sum_{rs} \sum_p \mathcal{U}_{min}^{rs*} \, [f_p^{rs}(t) - f_p^{rs*}(t)] \right\} dt \geq 0 \qquad (14.26)$$

By the definition of departure flows, we have

$$\int_0^T \sum_p f_p^{rs}(t) \, dt = F^{rs}(T) = \int_0^T \sum_p f_p^{rs*}(t) \, dt$$

Thus, the second term of equation (14.26) is 0 and equation (14.26) becomes variational inequality (14.21).

*Proof of Sufficiency.*

We need to prove that any solutions $f_p^{rs^*}(t)$ to variational inequality (14.21) satisfy DUO departure time/route choice conditions (14.2)-(14.4). We know that the first and third DUO departure time/route choice conditions (14.2) and (14.4) hold by definition. Thus, we only need to prove that the second DUO departure time/route choice condition (14.3) also holds.

Assume that DUO departure time/route choice condition (14.3) does not hold only for route $q$ between O-D pair $kn$ during a short time interval $[d-\delta, d+\delta] \in [0, T]$, i.e.,

$$f_q^{kn^*}(t) > 0 \qquad \text{and} \qquad \mathcal{U}_q^{kn^*}(t) - \mathcal{U}_{min}^{kn^*} > 0 \tag{14.27}$$

or

$$f_q^{kn^*}(t) \; \{\mathcal{U}_q^{kn^*}(t) - \mathcal{U}_{min}^{kn^*}\} > 0 \tag{14.28}$$

Integrating equation (14.28) from $(d-\delta)$ to $(d+\delta)$, we have

$$\int_{d-\delta}^{d+\delta} f_q^{kn^*}(t) \; \{\mathcal{U}_q^{kn^*}(t) - \mathcal{U}_{min}^{kn^*}\} \; dt > 0 \tag{14.29}$$

or

$$\int_{d-\delta}^{d+\delta} f_q^{kn^*}(t) \, \mathcal{U}_q^{kn^*}(t) \; dt \; > \; \int_{d-\delta}^{d+\delta} f_q^{kn^*}(t) \, \mathcal{U}_{min}^{kn^*} \; dt \tag{14.30}$$

We now need to find a set of feasible departure route inflows $f_p^{rs}(t)$ which will contradict variational inequality (14.21).

For routes between O-D pairs $rs \neq kn$, we allow the feasible departure route inflows $f_p^{rs}(t)$ to equal the optimal departure route inflows $f_p^{rs^*}(t)$ at each instant of time. For O-D pair $kn$, we can always find a route $l$ with the minimal disutility $\mathcal{U}_l^{kn^*}(t) = \mathcal{U}_{min}^{kn^*}$ for a short time interval $[b-\epsilon, b+\epsilon] \in [0, T]$. We note that route $l$ is evaluated under the optimal departure route inflow pattern $\{f_p^{rs^*}(t)\}$. For other routes $p \neq q, l$ between O-D pair $kn$, we allow the feasible departure route inflows $f_p^{kn}(t)$ to equal the optimal departure route inflows $f_p^{kn^*}(t)$ at each instant of time. For routes $q, l$ between O-D pair $kn$, we allow the feasible departure route inflows $f_q^{kn}(t)$, $f_l^{kn}(t)$ to equal optimal departure route inflows $f_q^{kn^*}(t)$ and $f_l^{kn^*}(t)$ for times outside the time interval $[d-\delta, d+\delta]$ and $[b-\epsilon, b+\epsilon]$, respectively. Furthermore, we shift all departure inflows over route $q$ during time interval $[d-\delta, d+\delta]$ to route $l$ over the time interval $[b-\epsilon, b+\epsilon]$, during which disutility $\mathcal{U}_l^{kn^*}(t) = \mathcal{U}_{min}^{kn^*}$. It follows that

$$\int_{b-\epsilon}^{b+\epsilon} f_l^{kn}(t)dt - \int_{b-\epsilon}^{b+\epsilon} f_l^{kn^*}(t)dt = \int_{d-\delta}^{d+\delta} f_q^{kn^*}(t)dt \tag{14.31}$$

Variational inequality (14.21) becomes

$$\int_0^T \sum_{rs}\sum_p \mathcal{U}_p^{rs^*}(t)\left[f_p^{rs}(t)-f_p^{rs^*}(t)\right]\,dt$$
$$= \int_0^T \left\{\mathcal{U}_l^{kn^*}(t)\left[f_l^{kn}(t)-f_l^{kn^*}(t)\right]\right.$$
$$+ \left.\mathcal{U}_q^{kn^*}(t)\left[f_q^{kn}(t)-f_q^{kn^*}(t)\right]\right\}\,dt \qquad (14.32)$$

The second term of equation (14.32) becomes

$$\int_0^T \mathcal{U}_q^{kn^*}(t)\left[f_q^{kn}(t)-f_q^{kn^*}(t)\right]\,dt$$
$$= \int_0^T \mathcal{U}_q^{kn^*}(t)\,f_q^{kn}(t)\,dt - \int_0^T \mathcal{U}_q^{kn^*}(t)f_q^{kn^*}(t)\,dt$$
$$= \int_0^T \mathcal{U}_q^{kn^*}(t)\,f_q^{kn}(t)\,dt - \left\{\int_0^{d-\delta} + \int_{d+\delta}^T\right\}\mathcal{U}_q^{kn^*}(t)\,f_q^{kn^*}(t)\,dt$$
$$- \int_{d-\delta}^{d+\delta} \mathcal{U}_q^{kn^*}(t)f_q^{kn^*}(t)\,dt \qquad (14.33)$$

Substituting equation (14.30) into equation (14.33), we have

$$\int_0^T \mathcal{U}_q^{kn^*}(t)\left[f_q^{kn}(t)-f_q^{kn^*}(t)\right]\,dt$$
$$< \int_0^T \mathcal{U}_q^{kn^*}(t)\,f_q^{kn}(t)\,dt - \left\{\int_0^{d-\delta} + \int_{d+\delta}^T\right\}\mathcal{U}_q^{kn^*}(t)\,f_q^{kn^*}(t)\,dt$$
$$- \int_{d-\delta}^{d+\delta} \mathcal{U}_{min}^{kn^*}(t)f_q^{kn^*}(t)\,dt \qquad (14.34)$$

The first term of equation (14.32) becomes

$$\int_0^T \mathcal{U}_l^{kn^*}(t)\,\left[f_l^{kn}(t)-f_l^{kn^*}(t)\right]\,dt$$
$$= \int_{b-\epsilon}^{b+\epsilon} \mathcal{U}_l^{kn^*}(t)\,\left[f_l^{kn}(t)-f_l^{kn^*}(t)\right]\,dt$$
$$= \int_{b-\epsilon}^{b+\epsilon} \mathcal{U}_{min}^{kn^*}(t)\,\left[f_l^{kn}(t)-f_l^{kn^*}(t)\right]\,dt$$
$$= \mathcal{U}_{min}^{kn^*}(t)\int_{b-\epsilon}^{b+\epsilon}\left[f_l^{kn}(t)-f_l^{kn^*}(t)\right]\,dt \qquad (14.35)$$

where $\mathcal{U}_l^{kn^*}(t) = \mathcal{U}_{min}^{kn^*}(t)$ for time interval $[b-\epsilon, b+\epsilon]$. Substituting equation (14.31) into equation (14.35), we have

$$\int_0^T \mathcal{U}_l^{kn^*}(t) \left[f_l^{kn}(t) - f_l^{kn^*}(t)\right] dt = \mathcal{U}_{min}^{kn^*}(t) \int_{d-\delta}^{d+\delta} f_q^{kn^*}(t)\, dt \quad (14.36)$$

Summing equations (14.34) and (14.36) and substituting into equation (14.32), we obtain

$$\begin{aligned} & \int_0^T \sum_{rs}\sum_p \mathcal{U}_p^{rs^*}(t) \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] dt \\ < \quad & \int_0^T \mathcal{U}_q^{kn^*}(t)\, f_q^{kn}(t)\, dt - \left\{\int_0^{d-\delta} + \int_{d+\delta}^T\right\} \mathcal{U}_q^{kn^*}(t)\, f_q^{kn^*}(t)\, dt \quad (14.37) \end{aligned}$$

Furthermore, since $f_q^{kn}(t) = 0$ for time interval $[d-\delta, d+\delta]$ and $f_q^{kn}(t) = f_q^{kn^*}(t)$ for time intervals $[0, d-\delta]$, and $[d+\delta, T]$, we simplify equation (14.37) as

$$\begin{aligned} & \int_0^T \sum_{rs}\sum_p \mathcal{U}_p^{rs^*}(t) \left[f_p^{rs}(t) - f_p^{rs^*}(t)\right] dt \\ < \quad & \left\{\int_0^{d-\delta} + \int_{d+\delta}^T\right\} \mathcal{U}_q^{kn^*}(t)\, f_q^{kn}(t)\, dt - \left\{\int_0^{d-\delta} + \int_{d+\delta}^T\right\} \mathcal{U}_q^{kn^*}(t)\, f_q^{kn^*}(t)\, dt \\ = \quad & \left\{\int_0^{d-\delta} + \int_{d+\delta}^T\right\} \mathcal{U}_q^{kn^*}(t) \left[f_q^{kn}(t) - f_q^{kn^*}(t)\right] dt = 0 \quad (14.38) \end{aligned}$$

The above equation contradicts variational inequality (14.21). Therefore, any optimal solutions $\{f_p^{rs^*}(t)\}$ to variational inequality (14.21) satisfy the second DUO departure time/route choice condition (14.3). Since we have proved the necessity and sufficiency of the equivalence of variational inequality (14.21) to the route-based DUO departure time/route choice conditions (14.2)-(14.4), the proof is complete.

## 14.2 A Link-Based VI Model of Departure Time/Route Choice

Because dynamic traffic flow does not have a constant flow rate during propagation over links and routes, the route-based VI can not be transformed into a link-based VI. Thus, it is very difficult to develop a solution algorithm for a route-based VI without explicit route enumeration. In Chapter 13, we presented a link-based VI model for the ideal DUO route choice problem so that route enumeration can be avoided in both the formulation and the solution procedure. This approach allows the dynamic VI route choice model to be applied to realistic transportation networks.

Using a similar approach, we extend the dynamic route choice model to include departure time choice as well. A link-based ideal dynamic user-optimal (DUO) departure time/route choice model is presented for a network with multiple origin-destination pairs in this section. Since this VI model is link-based, it has computational advantages over route-based models.

The set of dynamic network constraints for the link-based VI model is identical to constraint set (14.5)-(14.20) of the route-based VI model in Section 14.1. The basic difference between the two models is that the link-based VI model is formulated using link-based flow variables instead of route-based variables.

## 14.2.1 Link-Based Conditions

### Departure Time Choice Conditions

We first consider the departure time choice problem. A disutility function $\mathcal{U}^{rs}(t)$ based on departure times is defined for travelers departing from origin $r$ to destination $s$ at time $t$. Denote $\pi^{rs}(t)$ as the minimal travel time experienced by vehicles departing from origin $r$ to destination $s$ at time $t$. We also define the disutility for early or late arrival as follows

$$\mathcal{V}^{rs}[t, \pi^{rs}(t); t^*_{rs}] =$$

$$\begin{cases} \gamma_1[t + \pi^{rs}(t) - t^*_{rs} + \Delta^*_{rs}]^2 & \text{if } t + \pi^{rs}(t) < t^*_{rs} - \Delta^*_{rs} \quad \text{(early arrival)} \\ 0 & \text{if } |t + \pi^{rs}(t) - t^*_{rs}| \leq \Delta^*_{rs} \\ \gamma_2[t + \pi^{rs}(t) - t^*_{rs} - \Delta^*_{rs}]^2 & \text{if } t + \pi^{rs}(t) > t^*_{rs} + \Delta^*_{rs} \quad \text{(late arrival)} \end{cases}$$

where $t$ is the departure time of travelers and $\gamma_1$ and $\gamma_2$ are parameters ($\gamma_1 \leq 0$, $\gamma_2 \gg \alpha$). Thus, the disutility function for the joint departure time and route choice problem is constructed as

$$\mathcal{U}^{rs}(t) = \alpha t + \pi^{rs}(t) + \mathcal{V}^{rs}[t, \pi^{rs}(t); t^*_{rs}] \qquad \forall r, s, \tag{14.39}$$

where $t$ is the departure time of travelers.

The dynamic user-optimal departure time choice conditions require that for each O-D pair $rs$ at any time $t$, if there is a positive departure flow $f^{rs}(t) > 0$, the disutility $\mathcal{U}^{rs}(t)$ must equal the minimal $rs$ disutility $\mathcal{U}^{rs}_{min}$ over time $t$. Furthermore, if the departure flow $f^{rs}(t)$ equals 0 at time $t$, the disutility $\mathcal{U}^{rs}(t)$ at time $t$ must be greater than or equal to the minimal $rs$ disutility $\mathcal{U}^{rs}_{min}$. The DUO departure time choice conditions can be written as

$$\mathcal{U}^{rs^*}(t) - \mathcal{U}^{rs}_{min} \geq 0 \qquad \forall r, s; \tag{14.40}$$

$$f^{rs^*}(t)\ \{\mathcal{U}^{rs^*}(t) - \mathcal{U}^{rs}_{min}\} = 0 \qquad \forall r, s; \tag{14.41}$$

$$f^{rs}(t) \geq 0 \qquad \forall r, s. \tag{14.42}$$

where the asterisk denotes that the travel disutility is computed using DUO departure flows.

### Ideal DUO Route Choice Conditions

We then consider the route choice problem. The actual travel time $\tau_a[x_a(t), u_a(t), v_a(t)]$, or simply $\tau_a(t)$, over link $a$ is assumed to be dependent on the number of vehicles $x_a(t)$, the inflow $u_a(t)$ and the exit flow $v_a(t)$ on link $a$ at time $t$. We assume the travel time $\tau_a(t)$ is the sum of two components: 1) a flow-dependent cruise time $g_{1a}[x_a(t), u_a(t)]$ over the uncongested part of link $a$ and 2) a queuing delay $g_{2a}[x_a(t), v_a(t)]$ at the end of link $a$. It follows that

$$\tau_a(t) = g_{1a}[x_a(t), u_a(t)] + g_{2a}[x_a(t), v_a(t)]. \qquad (14.43)$$

The two components $g_{1a}[x_a(t), u_a(t)]$ and $g_{2a}[x_a(t), v_a(t)]$ of the time-dependent link travel time function $\tau_a[x_a(t), u_a(t), v_a(t)]$ are assumed to be nonnegative and differentiable with respect to $x_a(t), u_a(t)$ and $x_a(t), v_a(t)$, respectively. Recall that the travel-time-based ideal DUO route choice state is defined as:

> **Travel-Time-Based Ideal DUO State:** *If, for each O-D pair at each instant of time, the* actual *travel times experienced by travelers departing at the same time are equal and minimal, the dynamic traffic flow over the network is in a travel-time-based ideal dynamic user-optimal state.*

For vehicles departing from origin $r$ at time $t$, denote $\Omega_a^{rj^*}(t)$ as the difference between the minimal travel time from $r$ to $j$ and the travel time from origin $r$ to node $j$ via the minimal travel time route from origin $r$ to node $i$ and link $a$. It follows that

$$\Omega_a^{rj^*}(t) = \pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)] - \pi^{rj^*}(t) \qquad \forall a, r; a = (i, j). \qquad (14.44)$$

Thus, the link-time-based ideal DUO route choice conditions are:

$$\Omega_a^{rj^*}(t) \geq 0 \qquad \forall a = (i, j), r; \qquad (14.45)$$

$$u_a^{rs^*}[t + \pi^{ri^*}(t)]\, \Omega_a^{rj^*}(t) = 0 \qquad \forall a = (i, j), r, s; \qquad (14.46)$$

$$u_a^{rs}[t + \pi^{ri^*}(t)] \geq 0 \qquad \forall a = (i, j), r, s. \qquad (14.47)$$

### Joint DUO Departure Time/Route Choice Conditions

In order to simplify the presentation, we rewrite the combined link-based DUO departure time/route choice conditions as follows:

$$\Omega_a^{rj^*}(t) \geq 0 \qquad \forall a = (i, j), r; \qquad (14.48)$$

$$u_a^{rs^*}[t + \pi^{ri^*}(t)]\, \Omega_a^{rj^*}(t) = 0 \qquad \forall a = (i, j), r, s; \qquad (14.49)$$

$$u_a^{rs}[t + \pi^{ri^*}(t)] \geq 0 \qquad \forall a = (i, j), r, s; \qquad (14.50)$$

$$\mathcal{U}^{rs^*}(t) - \mathcal{U}_{min}^{rs^*} \geq 0 \qquad \forall r, s; \qquad (14.51)$$

$$f^{rs^*}(t)\, \{\mathcal{U}^{rs^*}(t) - \mathcal{U}_{min}^{rs^*}\} = 0 \qquad \forall r, s; \qquad (14.52)$$

$$f^{rs}(t) \geq 0 \qquad \forall r, s. \qquad (14.53)$$

where $\mathcal{U}_{min}^{rs}$ is the minimal $rs$ disutility over time $t$.

### 14.2.2 The Link-Based VI Model

The equivalent variational inequality formulation of link-based DUO departure time/route choice conditions (14.48)-(14.53) may be stated as follows.

*Theorem 14.2.* Dynamic traffic flow satisfying constraints (14.5)-(14.20) is in a DUO departure time/route choice state if and only if it satisfies the variational inequality problem:

$$\int_0^T \left\{ \sum_{rs} \sum_a \Omega_a^{rj^*}(t) \left\{ u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)] \right\} + \sum_{rs} \mathcal{U}^{rs^*}(t) \left\{ f^{rs}(t) - f^{rs^*}(t) \right\} \right\} dt \geq 0 \tag{14.54}$$

*Proof of Necessity.*

We need to prove that DUO departure time/route choice conditions (14.48)-(14.53) imply variational inequality (14.54). We first discuss the ideal DUO route choice conditions (14.48)-(14.50). Multiplying equations (14.48) and (14.50), we have

$$u_a^{rs}[t + \pi^{ri^*}(t)]\, \Omega_a^{rj^*}(t) \geq 0 \qquad \forall a, r, s; a = (i, j). \tag{14.55}$$

Subtracting equation (14.49) from equation (14.55), we obtain

$$\left\{ u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)] \right\} \Omega_a^{rj^*}(t) \geq 0 \quad \forall a, r, s; a = (i, j). \tag{14.56}$$

Summing equation (14.56) for all links $a$ and all O-D pairs $rs$, it follows that

$$\sum_{rs} \sum_a \left\{ u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)] \right\} \Omega_a^{rj^*}(t) \geq 0 \qquad \text{where } a = (i, j). \tag{14.57}$$

Integrating the above equation (14.57) from time 0 to $T$, we have

$$\int_0^T \sum_{rs} \sum_a \left\{ u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)] \right\} \Omega_a^{rj^*}(t)\, dt \geq 0 \tag{14.58}$$

We next discuss DUO departure time choice conditions (14.51)-(14.53). Multiplying equations (14.51) and (14.53), we have

$$f^{rs}(t) \{\mathcal{U}^{rs^*}(t) - \mathcal{U}^{rs}_{min}\} \geq 0 \qquad \forall r, s. \tag{14.59}$$

We subtract equation (14.52) from equation (14.59) to obtain

$$\mathcal{U}^{rs^*}(t) \{f^{rs}(t) - f^{rs^*}(t)\} - \mathcal{U}^{rs}_{min} \{f^{rs}(t) - f^{rs^*}(t)\} \geq 0 \qquad \forall r, s. \tag{14.60}$$

Summing equation (14.60) for all O-D pairs $rs$, it follows that

$$\sum_{rs} \mathcal{U}^{rs^*}(t)\,\{f^{rs}(t) - f^{rs^*}(t)\} - \sum_{rs} \mathcal{U}^{rs}_{min}\,\{f^{rs}(t) - f^{rs^*}(t)\} \geq 0 \quad (14.61)$$

Integrating the above equation from time 0 to $T$, we have

$$\int_0^T \sum_{rs} \mathcal{U}^{rs^*}(t)\,\{f^{rs}(t) - f^{rs^*}(t)\}\,dt - \int_0^T \sum_{rs} \mathcal{U}^{rs}_{min}\,\{f^{rs}(t) - f^{rs^*}(t)\}\,dt \geq 0 \quad (14.62)$$

or

$$\int_0^T \sum_{rs} \mathcal{U}^{rs^*}(t)\,\{f^{rs}(t) - f^{rs^*}(t)\}\,dt - \sum_{rs} \mathcal{U}^{rs}_{min} \int_0^T \{f^{rs}(t) - f^{rs^*}(t)\}\,dt \geq 0 \quad (14.63)$$

By the definition of departure flows, we have

$$\int_0^T f^{rs}(t)\,dt = F^{rs}(T) = \int_0^T f^{rs^*}(t)\,dt$$

Thus, the second term of equation (14.63) is 0 and equation (14.63) becomes

$$\int_0^T \sum_{rs} \mathcal{U}^{rs^*}(t)\,\{f^{rs}(t) - f^{rs^*}(t)\}\,dt \geq 0 \quad (14.64)$$

Combining equations (14.58) and (14.64), we obtain variational inequality (14.54).

*Proof of Sufficiency.*

We need to prove that any solutions $u_a^{rs^*}[t + \pi^{ri^*}(t)]$ and $f^{rs^*}(t)$ to variational inequality (14.54) satisfy DUO departure time/route choice conditions (14.48)-(14.53). We know that the first and third ideal DUO route choice conditions (14.48) and (14.50) hold by definition. The fourth and sixth DUO departure time choice conditions (14.51) and (14.53) also hold by definition. Thus, we only need to prove that the second ideal DUO route choice condition (14.49) and the fifth DUO departure time choice condition (14.52) also hold.

In order to prove the above statement, we need to prove that the following three cases are not true.

1. The second ideal DUO route choice condition (14.49) does not hold, but the fifth DUO departure time choice condition (14.52) does hold.

2. The fifth DUO departure time choice condition (14.52) does not hold, but the second ideal DUO route choice condition (14.49) does hold.

3. Both the second ideal DUO route choice condition (14.49) and the fifth DUO departure time choice condition (14.52) do not hold.

**Case 1**

The proof for Case 1 is similar to that for a pure ideal DUO route choice problem. Assume that the second ideal DUO route choice condition (14.49) does not hold only for a link $b = (l, m)$ for O-D pair $pq$ during a short time interval $[d - \delta, d + \delta]$ where $[d - \delta, d + \delta] \in [0, T]$, i.e.,

$$u_b^{pq^*}[t + \tau^{kl^*}(t)] > 0 \qquad \text{and}$$
$$\Omega_b^{km^*}(t) = \tau^{kl^*}(t) + \tau_b[t + \tau^{kl^*}(t)] - \tau^{km^*}(t) > 0 \tag{14.65}$$

Multiplying the above two equations, we have

$$u_b^{pq^*}[t + \tau^{kl^*}(t)]\, \Omega_b^{km^*}(t) > 0 \tag{14.66}$$

The first term in the variational inequality (14.54) becomes

$$\int_0^T \sum_{rs} \sum_a \Omega_a^{rj^*}(t)\, u_a^{rs^*}[t + \pi^{ri^*}(t)]\, dt$$
$$= \int_{d-\delta}^{d+\delta} u_b^{pq^*}[t + \tau^{kl^*}(t)]\ \Omega_b^{km^*}(t)\, dt > 0 \tag{14.67}$$

We note that all other terms in the above equation vanish because the ideal DUO route choice condition (14.49) holds for other links and O-D pairs at each time instant and for link $b = (l, m)$ for O-D pair $pq$ at time instants which are not within time interval $[d - \delta, d + \delta]$.

For each O-D pair $rs$, we can always find one minimal travel time route $k$ for vehicles departing origin $r$ at time $t$, which was evaluated under the optimal flow pattern $\{u_a^{rs^*}[t + \pi^{ri^*}(t)]\}$. For this route $k$, the first ideal DUO route choice condition (14.48) becomes equality by definition. It follows that

$$\Omega_a^{rj^*}(t) = \pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)] - \pi^{rj^*}(t) = 0$$

$$\forall a, r, s; a = (i, j); a \in k. \tag{14.68}$$

Next, we need to find a set of feasible inflows $u_a^{rs}[t+\pi^{ri^*}(t)]$ so that the following equation

$$u_a^{rs}[t + \pi^{ri^*}(t)]\, \Omega_a^{rj^*}(t) = 0 \qquad \forall a, r, s; a = (i, j) \tag{14.69}$$

always holds. We choose the feasible departure flows $f^{rs}(t)$ to equal the optimal departure flows $f^{rs^*}(t)$ for all O-D pairs $rs$ at each instant of time. Thus, the second term in (14.54) will vanish. We also need to re-route all feasible departure flows $f^{rs}(t)$ for all O-D pairs at each instant of time. For each O-D pair $rs$, we assign the feasible O-D departure flow $f^{rs}(t)$ to the minimal travel time route $k$, which was evaluated under the optimal flow patterns $\{u_a^{rs^*}[t + \pi^{ri^*}(t)]\}$. This generates a set of feasible inflow patterns $\{u_a^{rs}[t+\pi^{ri^*}(t)]\}$ which always satisfies equation (14.69) (because either $\Omega_a^{rj^*}(t) = 0$ for links on route $k$ or $u_a^{rs}[t + \pi^{ri^*}(t)] = 0$ since no flow is routed onto those links which are not on route $k$). Summing equations (14.69) for all links $a$ and all O-D pairs $rs$, it follows that

$$\sum_{rs}\sum_{a} u_a^{rs}[t + \pi^{ri^*}(t)]\ \Omega_a^{rj^*}(t) = 0 \qquad \text{where } a = (i,j). \tag{14.70}$$

Integrating the above equation, we have

$$\int_0^T \sum_{rs}\sum_{a} \Omega_a^{rj^*}(t)\ u_a^{rs}[t + \pi^{ri^*}(t)]\ dt = 0 \tag{14.71}$$

We subtract equation (14.67) from equation (14.71) and obtain

$$\int_0^T \sum_{rs}\sum_{a} \Omega_a^{rj^*}(t)\ \left\{u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)]\right\}\ dt < 0 \tag{14.72}$$

Note that the second term in (14.54) equals 0. It follows that

$$\int_0^T \left\{\sum_{rs}\sum_{a} \Omega_a^{rj^*}(t)\ \left\{u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)]\right\} + \sum_{rs} \mathcal{U}^{rs^*}(t)\left\{f^{rs}(t) - f^{rs^*}(t)\right\}\right\}\ dt < 0 \tag{14.73}$$

The above equation contradicts variational inequality (14.54). Therefore, any optimal solutions $\{u_a^{rs^*}[t + \pi^{ri^*}(t)]\}$ and $\{f^{rs^*}(t)\}$ to variational inequality (14.54) that satisfy the fifth DUO departure time choice condition (14.52) also satisfy the second ideal DUO route choice condition (14.49).

### Case 2

Assume that the fifth DUO departure time choice condition (14.52) does not hold only for O-D pair $pq$ during a short time interval $[d-\delta, d+\delta]$ where $[d-\delta, d+\delta] \in [0,T]$, i.e.,

$$f^{pq^*}(t) > 0 \qquad \text{and} \qquad \mathcal{U}^{pq^*}(t) - \mathcal{U}^{pq}_{min} > 0 \tag{14.74}$$

or

$$f^{pq^*}(t)\ \{\mathcal{U}^{pq^*}(t) - \mathcal{U}^{pq}_{min}\} > 0 \tag{14.75}$$

We now need to find a set of feasible departure flows $f^{rs}(t)$ and link inflows $u_a^{rs}[t + \pi^{ri^*}(t)]$ which will contradict variational inequality (14.54).

For O-D pairs $rs \neq pq$, we allow the feasible departure flows $f^{rs}(t)$ to equal the optimal departure flows $f^{rs^*}(t)$ at each instant of time. For O-D pair $pq$, we shift all departure flows during time interval $[d - \delta, d + \delta]$ to the time interval $[d + \delta, T]$, during which disutility $\mathcal{U}^{pq^*}(t) = \mathcal{U}^{pq}_{min}$, (if $d + \delta = T$, we shift all departure flows during time interval $[d - \delta, d + \delta]$ to the time interval $[0, d - \delta]$ and the proof will follow.) Thus, the second term of variational inequality (14.54) becomes

$$\begin{aligned}
&\int_0^T \sum_{rs} \mathcal{U}^{rs^*}(t) \left\{ f^{rs}(t) - f^{rs^*}(t) \right\} dt \\
= &\int_0^T \mathcal{U}^{pq^*}(t) \left\{ f^{pq}(t) - f^{pq^*}(t) \right\} dt \\
= &\left\{ \int_0^{d-\delta} + \int_{d+\delta}^T \right\} \mathcal{U}^{pq^*}(t) \left\{ f^{pq}(t) - f^{pq^*}(t) \right\} dt \\
+ &\int_{d-\delta}^{d+\delta} \mathcal{U}^{pq^*}(t) \left\{ f^{pq}(t) - f^{pq^*}(t) \right\} dt \\
= &\, \mathcal{U}^{pq}_{min} \left\{ \int_0^{d-\delta} + \int_{d+\delta}^T \right\} \left\{ f^{pq}(t) - f^{pq^*}(t) \right\} dt \\
- &\int_{d-\delta}^{d+\delta} \mathcal{U}^{pq^*}(t)\, f^{pq^*}(t)\, dt \qquad (14.76)
\end{aligned}$$

Note that in the above equation, $\mathcal{U}^{pq^*}(t) = \mathcal{U}^{pq}_{min}$ for time instants which do not lie within time interval $[d - \delta, d + \delta]$. For any time instant $t$ during time interval $[d - \delta, d + \delta]$, the adjusted feasible departure flow $f^{pq}(t) = 0$. By definition of departure flows, we have

$$\int_0^T f^{pq}(t)\, dt = F^{pq}(T) = \int_0^T f^{pq^*}(t)\, dt$$

Substituting into equation (14.76), it follows that

$$\begin{aligned}
&\left\{ \int_0^{d-\delta} + \int_{d+\delta}^T \right\} \left\{ f^{pq}(t) - f^{pq^*}(t) \right\} dt \\
= &\int_0^T \left\{ f^{pq}(t) - f^{pq^*}(t) \right\} dt - \int_{d-\delta}^{d+\delta} \left\{ f^{pq}(t)dt - f^{pq^*}(t) \right\} dt \\
= &\int_{d-\delta}^{d+\delta} f^{pq^*}(t)\, dt \qquad (14.77)
\end{aligned}$$

Integrating equation (14.75) from $(d - \delta)$ to $(d + \delta)$, we have

$$\int_{d-\delta}^{d+\delta} f^{pq^*}(t) \left\{ \mathcal{U}^{pq^*}(t) - \mathcal{U}^{pq}_{min} \right\} dt > 0 \qquad (14.78)$$

or

$$\int_{d-\delta}^{d+\delta} f^{pq^*}(t)\, \mathcal{U}^{pq^*}(t)\, dt > \int_{d-\delta}^{d+\delta} f^{pq^*}(t)\, \mathcal{U}^{pq}_{min}\, dt \tag{14.79}$$

Substituting equations (14.77) and (14.79) into equation (14.76), we have

$$\mathcal{U}^{pq}_{min} \left\{ \int_0^{d-\delta} + \int_{d+\delta}^{T} \right\} \left\{ f^{pq}(t) - f^{pq^*}(t) \right\} dt - \int_{d-\delta}^{d+\delta} \mathcal{U}^{pq^*}(t)\, f^{pq^*}(t)\, dt$$

$$= \mathcal{U}^{pq}_{min} \int_{d-\delta}^{d+\delta} f^{pq^*}(t)\, dt - \int_{d-\delta}^{d+\delta} \mathcal{U}^{pq^*}(t)\, f^{pq^*}(t)\, dt$$

$$< \mathcal{U}^{pq}_{min} \int_{d-\delta}^{d+\delta} f^{pq^*}(t)\, dt - \int_{d-\delta}^{d+\delta} \mathcal{U}^{pq}_{min}\, f^{pq^*}(t)\, dt = 0 \tag{14.80}$$

Following the above adjustment of feasible departure flows, the link inflows should be adjusted accordingly so as to be feasible. As illustrated in Case 1, for each O-D pair $rs$, we can always find one minimal travel time route $k$ for vehicles departing origin $r$ at time $t$, which was evaluated under the optimal flow pattern $\{u_a^{rs^*}[t + \pi^{ri^*}(t)]\}$. For this route $k$, the first ideal DUO route choice condition (14.48) becomes an equality by definition. It follows that

$$\Omega_a^{rj^*}(t) = \pi^{ri^*}(t) + \tau_a[t + \pi^{ri^*}(t)] - \pi^{rj^*}(t) = 0$$

$$\forall a, r, s; a = (i, j); a \in k. \tag{14.81}$$

Next, we need to find a set of feasible inflows $u_a^{rs}[t+\pi^{ri^*}(t)]$ so that the following equation

$$u_a^{rs}[t + \pi^{ri^*}(t)]\, \Omega_a^{rj^*}(t) = 0 \qquad \forall a, r, s; a = (i, j) \tag{14.82}$$

always holds. For each O-D pair $rs$ at each time instant $t$, we assign the feasible O-D departure flow $f^{rs}(t)$ to the minimal travel time route $k$ only, which was evaluated under the optimal flow pattern $\{u_a^{rs^*}[t+\pi^{ri^*}(t)]\}$. This will generate a feasible inflow pattern $\{u_a^{rs}[t+\pi^{ri^*}(t)]\}$ which always satisfy equation (14.82). Summing equations (14.82) for all links $a$ and all O-D pairs $rs$, it follows that

$$\sum_{rs} \sum_{a} u_a^{rs}[t + \pi^{ri^*}(t)]\, \Omega_a^{rj^*}(t) = 0 \qquad \text{where } a = (i, j). \tag{14.83}$$

Subtracting equation (14.49) from equation (14.83) and integrating the resulting equation, we have

$$\int_0^T \sum_{rs} \sum_{a} \Omega_a^{rj^*}(t) \left\{ u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)] \right\} dt = 0 \tag{14.84}$$

Combining equations (14.76), (14.80) and (14.84), it follows that

$$\int_0^T \left\{ \sum_{rs} \sum_{a} \Omega_a^{rj^*}(t) \left\{ u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)] \right\} \right.$$

$$\left. + \sum_{rs} \mathcal{U}^{rs^*}(t) \left\{ f^{rs}(t) - f^{rs^*}(t) \right\} \right\} dt < 0 \tag{14.85}$$

The above equation contradicts variational inequality (14.54). Therefore, any optimal solutions $\{u_a^{rs^*}[t + \pi^{ri^*}(t)]\}$ and $\{f^{rs^*}(t)\}$ to variational inequality (14.54) that satisfy the second ideal DUO route choice condition (14.49) also satisfy the fifth DUO departure time choice condition (14.52).

**Case 3**

Case 3 includes the following two sub-cases: 3a) conditions (14.49) and (14.52) do not hold for different O-D pairs; 3b) conditions (14.49) and (14.52) do not hold for the same O-D pair.

**Case 3a**

Assume that the second ideal DUO route choice condition (14.49) does not hold for O-D pair $kn$ for time interval $[d_1 - \delta_1, d_1 + \delta_1]$ and the fifth DUO departure time choice condition (14.52) does not hold for O-D pair $pq$ for time interval $[d_2 - \delta_2, d_2 + \delta_2]$. Note that the two O-D pairs are different, but the two time intervals may or may not be different. For O-D pair $kn$, we assume that the second ideal DUO route choice condition (14.49) does not hold only for a link $b = (l, m)$ during time interval $[d_1 - \delta_1, d_1 + \delta_1]$, i.e.,

$$\begin{aligned} &u_b^{kn^*}[t + \tau^{kl^*}(t)] > 0 \qquad \text{and} \\ &\Omega_b^{km^*}(t) = \tau^{kl^*}(t) + \tau_a[t + \tau^{kl}(t)] - \tau^{km^*}(t) > 0 \end{aligned} \tag{14.86}$$

Following the derivation from (14.65) to (14.73) in Case 1, we can find a set of feasible inflows $u_a^{kn}[t + \pi^{ri^*}(t)]$ for O-D pair $kn$ so that the following equation holds:

$$\begin{aligned} \int_0^T &\left\{ \sum_a \Omega_a^{kj^*}(t) \left\{ u_a^{kn}[t + \tau^{ki^*}(t)] - u_a^{kn^*}[t + \tau^{ki^*}(t)] \right\} \right. \\ &+ \left. \mathcal{U}^{kn^*}(t) \left\{ f^{kn}(t) - f^{kn^*}(t) \right\} \right\} dt < 0 \end{aligned} \tag{14.87}$$

Note that in the derivation of the above equation, we follow the process in Case 1 by assuming there is only one O-D pair $kn$.

For O-D pair $pq$, we assume that the fifth DUO departure time choice condition (14.52) does not hold during time interval $[d_2 - \delta_2, d_2 + \delta_2]$.

$$f^{pq^*}(t) > 0 \qquad \text{and} \qquad \mathcal{U}^{pq^*}(t) - \mathcal{U}_{min}^{pq} > 0 \tag{14.88}$$

Following the derivation from (14.74) to (14.85) in Case 2, we can find a set of feasible departure flows $f^{pq}(t)$ and inflows $u_a^{kn}[t + \tau^{ki^*}(t)]$ so that the following equation holds:

$$\begin{aligned} \int_0^T &\left\{ \sum_a \Omega_a^{pj^*}(t) \left\{ u_a^{pq}[t + \tau^{pi^*}(t)] - u_a^{pq^*}[t + \tau^{pi^*}(t)] \right\} \right. \\ &+ \left. \mathcal{U}^{pq^*}(t) \left\{ f^{pq}(t) - f^{pq^*}(t) \right\} \right\} dt < 0 \end{aligned} \tag{14.89}$$

Note that in the derivation of the above equation, we follow the process in Case 2 by assuming there is only one O-D pair $pq$.

For other O-D pairs $rs$, we choose the feasible departure flows $f^{rs}(t)$ to equal the optimal departure flows $f^{rs^*}(t)$ at each instant of time $t$. Thus, the second terms in (14.54) for those O-D pairs will vanish. We also need to re-route all feasible departure flows $f^{rs}(t)$ for those O-D pairs at each instant of time. For each O-D pair $rs$ at each instant of time, we assign the feasible O-D departure flow $f^{rs}(t)$ to the minimal travel time route $h$, which was evaluated under the optimal flow pattern $\{u_a^{rs^*}[t + \pi^{ri^*}(t)]\}$. This will generate a set of feasible inflow patterns $\{u_a^{rs}[t + \pi^{ri^*}(t)]\}$ which always allows the first terms in (14.54) to equal 0 for those O-D pairs $rs$, as illustrated in Case 1.

Combining equations (14.87) and (14.89) and the above analysis,

$$\int_0^T \left\{ \sum_{rs} \sum_{a} \Omega_a^{rj^*}(t) \left\{ u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)] \right\} + \sum_{rs} \mathcal{U}^{rs^*}(t) \left\{ f^{rs}(t) - f^{rs^*}(t) \right\} \right\} dt < 0 \tag{14.90}$$

The above equation contradicts variational inequality (14.54). Therefore, any optimal solutions $\{u_a^{rs^*}[t + \pi^{ri^*}(t)]\}$ and $\{f^{rs^*}(t)\}$ to variational inequality (14.54) will satisfy both the second ideal DUO route choice condition (14.49) and the fifth DUO departure time choice condition (14.52).

### Case 3b

Assume that the second ideal DUO route choice condition (14.49) and the fifth DUO departure time choice condition (14.52) do not hold for the same O-D pair $pq$, but for time intervals $[d_1 - \delta_1, d_1 + \delta_1]$ and $[d_2 - \delta_2, d_2 + \delta_2]$, respectively. Note that the two time intervals can be either identical or different. Since the fifth DUO departure time choice condition (14.52) does not hold during time interval $[d_2 - \delta_2, d_2 + \delta_2]$, it follows that

$$f^{pq^*}(t) > 0 \qquad \text{and} \qquad \mathcal{U}^{pq^*}(t) - \mathcal{U}_{min}^{pq} > 0 \tag{14.91}$$

Following the derivation from (14.74) to (14.80) in Case 2, we can find a set of feasible departure flows $f^{pq}(t)$ so that the following equation holds for O-D pair $pq$:

$$\int_0^T \mathcal{U}^{pq^*}(t) \left\{ f^{pq}(t) - f^{pq^*}(t) \right\} dt < 0 \tag{14.92}$$

Note that in the derivation of the above equation, we follow the process in Case 2 by assuming there is only one O-D pair $pq$.

For O-D pair $pq$, we also need to adjust the link inflow pattern accordingly so that they will be feasible. We assume that the second ideal DUO route choice condition (14.49) does not hold only for a link $b = (l, m)$ during time interval $[d_1 - \delta_1, d_1 + \delta_1]$, i.e.,

$$u_b^{pq^*}[t + \tau^{pl^*}(t)] > 0 \qquad \text{and}$$

$$\Omega_b^{pm^*}(t) = \tau^{pl^*}(t) + \tau_a[t + \tau^{pl}(t)] - \tau^{pm^*}(t) > 0 \tag{14.93}$$

Following the derivation from equations (14.65) to (14.72) in Case 1, we can find a set of feasible inflows $u_a^{pq}[t+\tau^{pi^*}(t)]$ so that the following equation always hold for O-D pair $pq$:

$$\int_0^T \sum_a \Omega_a^{pj^*}(t) \left\{u_a^{pq}[t + \tau^{pi^*}(t)] - u_a^{pq^*}[t + \tau^{pi^*}(t)]\right\} dt \ < \ 0 \tag{14.94}$$

Note that in the derivation of the above equation, we follow the process in Case 1 by assuming there is only one O-D pair $pq$.

Combining equations (14.92) and (14.94), we have

$$\int_0^T \left\{\sum_a \Omega_a^{pj^*}(t) \left\{u_a^{pq}[t + \tau^{pi^*}(t)] - u_a^{pq^*}[t + \tau^{pi^*}(t)]\right\} + \ \mathcal{U}^{pq^*}(t)\left\{f^{pq}(t) - f^{pq^*}(t)\right\}\right\} dt < 0 \tag{14.95}$$

Thus, we found a set of feasible departure flows $f^{pq}(t)$ and inflows $u_a^{pq}[t+\tau^{pi^*}(t)]$ so that the above equation holds for O-D pair $pq$.

For other O-D pairs $rs$, we choose the feasible departure flows $f^{rs}(t)$ to equal the optimal departure flows $f^{rs^*}(t)$ at each instant of time. Thus, the second terms in (14.54) for those O-D pairs vanish. We also reroute the feasible departure flows $f^{rs}(t)$ for those O-D pairs at each time instant. For each O-D pair $rs$, we assign the feasible O-D departure flow $f^{rs}(t)$ to the minimal travel time route $h$, which was evaluated under the optimal flow pattern $\{u_a^{rs^*}[t + \pi^{ri^*}(t)]\}$. This generates a set of feasible inflow patterns $\{u_a^{rs}[t+\pi^{ri^*}(t)]\}$ which always allow the first terms in VI (14.54) for those O-D pairs to equal 0, as illustrated in Case 1.

Combining equations (14.95) and the above analysis, we have

$$\int_0^T \left\{\sum_{rs}\sum_a \Omega_a^{rj^*}(t) \left\{u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)]\right\} + \sum_{rs} \mathcal{U}^{rs^*}(t)\left\{f^{rs}(t) - f^{rs^*}(t)\right\}\right\} dt < 0 \tag{14.96}$$

The above equation contradicts variational inequality (14.54). Therefore, any optimal solutions $\{u_a^{rs^*}[t + \pi^{ri^*}(t)]\}$ and $\{f^{rs^*}(t)\}$ to variational inequality (14.54) will satisfy both the second ideal DUO route choice condition (14.49) and the fifth DUO departure time choice condition (14.52). Since we have proved the necessity and sufficiency of the equivalence of variational inequality (14.54) to DUO departure time/route choice conditions (14.48)-(14.53), the proof is complete.

## 14.3 VI Models and Optimization Models for Departure Time/Route Choice

We now consider the relationship between VI models and optimization models. As in Chapter 13, we do not discuss each VI model in this chapter. As an example, we focus our analysis on the link-based VI model for DUO departure time/route choice problem. We show in the following that under relaxation and some regularity conditions, the VI model can be reformulated as an optimal control problem. Similar analysis can be performed for the route-based VI model for DUO departure time/route choice problem. Therefore, in this section, we discuss the following VI problem:

$$\int_0^T \left\{ \sum_{rs} \sum_{a} \Omega_a^{rj^*}(t) \left\{ u_a^{rs}[t + \pi^{ri^*}(t)] - u_a^{rs^*}[t + \pi^{ri^*}(t)] \right\} \right.$$
$$\left. + \sum_{rs} \mathcal{U}^{rs^*}(t) \left\{ f^{rs}(t) - f^{rs^*}(t) \right\} \right\} dt \geq 0 \tag{14.97}$$

To simplify our analysis, we assume the time period $[0, T]$ is long enough so that all departure flows can be cleared by final time $T$. In other words, any positive departure from origin $r$ at time $t$ will arrive at destination $s$ at time $t + \pi^{rs^*}(t) \leq T$.

We consider a simplified link travel time function as follows:

$$\tau_a(t) = \tau_a[x_a(t), u_a(t)] \qquad \forall a. \tag{14.98}$$

Following a similar derivation in Chapter 13 for the ideal DUO route choice problem, we can design a similar double relaxation or diagonalization procedure so that the above VI can be formulated as an optimal control model in each relaxation iteration. At the first-level relaxation, we assume that the actual link travel time $\tau_a(t)$ in the flow propagation constraints and the resulting minimal actual travel times $\pi^{ri}(t)$ in the above variational inequality are fixed temporarily at each relaxation iteration. Furthermore, the resulting minimal actual travel times $\pi^{rs}(t)$ in the disutility function $\mathcal{U}^{rs}(t)$ are also fixed temporarily at each relaxation iteration. Then, the cross-effects of flow variables at different time instants can be separated at each iteration. We define a new time variable as $\xi_a^r = t + \pi^{ri^*}(t)$ where $\pi^{ri^*}(t)$ is fixed temporarily at each relaxation iteration. Suppose there are $R$ origins. Then, for each link $a$ and origin $r$ at time instant $\xi_a^r$, we have

$$u_a(\xi_a^r) = u_a^1(\xi_a^r) + u_a^2(\xi_a^r) + \cdots + u_a^R(\xi_a^r) \tag{14.99}$$

The actual link travel time $\tau_a(\xi_a^r)$ at time instant $\xi_a^r$ can be expressed as

$$\tau_a(\xi_a^r) = \tau_a[u_a^1(\xi_a^r), \cdots, u_a^R(\xi_a^r), x_a^1(\xi_a^r), \cdots, x_a^R(\xi_a^r)] \tag{14.100}$$

Since the cross-effects of origin-specific link flow variables are asymmetric and cannot be eliminated, we need to design a second-level relaxation for this

VI. In other words, we need to fix temporarily all other link flow variables $u_a(\xi_a^{r'})$, $x_a(\xi_a^{r'})$ $(r' \neq r)$ in Equation (14.100) at each second-level relaxation iteration. Thus, we obtain the objective function of the optimization program as

$$\text{Min} \int_0^T \sum_{r,a} \left\{ \int_0^{u_a^r(\xi_a^r)} \tau_a[u_a^1(\xi_a^r), \cdots, \omega, \cdots, u_a^R(\xi_a^r), x_a^1(\xi_a^r), \cdots, x_a^R(\xi_a^r)] \, d\omega \right.$$

$$\left. + \sum_{rs} f^{rs}(t) \, \mathcal{U}^{rs}(t) \right\} dt \tag{14.101}$$

Note that $\xi_a^r = t + \pi^{ri^*}(t)$ where $\pi^{ri^*}(t)$ is fixed temporarily at each second-level relaxation iteration. Also note that all flow variables at time instants $\xi_a^r > T$ are 0. Therefore, link-based variational inequality (14.97) can be reformulated as an optimal control problem with objective function (14.101) and constraints (14.5)-(14.11) at each double relaxation iteration.

## 14.4 Notes

Several departure time choice models have been proposed by various researchers using different approaches on dynamic traffic networks. Janson (1993) formulated a dynamic user-optimal route choice model in which trips have variable departure times and scheduled arrival times. Friesz et al (1993) presented a joint departure time and route choice model using the variational inequality approach. Ghali and Smith (1993) also considered this problem using a microscopic representation of vehicle streams.

In this chapter, a link-based VI model for DUO departure time/route choice was presented. The necessity and sufficiency proofs of the VI model demonstrate that this model is consistent with the link-based DUO departure time/route choice conditions. Using a link-based VI formulation, explicit route enumeration can be avoided in computation. This feature allows our model to be applied to large-scale dynamic transportation networks with general link travel time functions.

Two major constraints prevent us from applying existing dynamic transportation network models to ATIS systems. The first concerns the accurate representation of travelers' choice behavior. In future extensions, utility functions instead of pure travel times should be used in route choice problems. Different perceptions and compliance with information must be investigated by stratifying travelers into multiple groups. The second concerns the accurate representation of traffic dynamics on each street link. Link traffic dynamics may be very complicated; as pointed out by Newell (1990) and Daganzo (1993), a set of appropriate closed-form link travel time functions might involve the interactions of neighboring link flows. This feature prevents formulating an appropriate optimization model for a realistic departure time/route choice problem. Thus, the general VI formulation approach was proposed for such applications. However, VI models require more computational capability than optimization models.

The proposed link-based VI model for DUO departure time/route choice can be extended to include arrival time choice, destination choice and mode choice as well. Our next step is to develop efficient solution algorithms for the DUO departure time/route choice VI model. We expect that the Frank-Wolfe and diagonalization techniques proposed by Boyce et al (1991) and Ran et al (1993) can be applied to solve this model. Other solution algorithms, such as the projection algorithm, implemented by Nagurney (1986) for static network equilibrium VI models, are also extendable to our dynamic VI problem. We note that the solution algorithm for our DUO departure time/route choice VI model has to be implemented on an expanded time-space network as proposed in Boyce et al (1991). Other important problems, such as incident related dynamic route choice problems and dynamic congestion pricing problems, will be studied as extensions of this VI model.

# Chapter 15

# Dynamic System-Optimal Route Choice and Congestion Pricing

In this chapter, we present several dynamic system-optimal (DSO) route choice models for a network with multiple origin-destination pairs. The constraint set for DSO route choice models can be much more comprehensive, including constraints such as the capacity and oversaturation spillback constraints. However, the more constraints we have, the more difficult will be the solution algorithm. Thus, for large-scale networks, we need to make a trade-off between the reality of formulations and the difficulty of the solution algorithm. The modeling complexity can be pursued as long as realistic traffic flows can be fully represented and reasonable computational times can be achieved.

In a DSO route choice problem, various objective functions can be formulated. Each objective function corresponds to a specific requirement for the overall system. In Section 15.1, we present several typical objective functions for a DSO route choice model. In Section 15.2, a DSO route choice model which minimizes total travel time is formulated and a solution algorithm is presented. In Section 15.3, we discuss a set of DSO route choice models with elastic departure times. Time-optimal models for evacuation purposes are also formulated. In Section 15.4, we consider dynamic congestion pricing strategies which can make a dynamic system-optimal state consistent with a dynamic user-optimal state.

## 15.1 Objective Functions for Dynamic System-Optimal Models

Depending on the objective of the central controller of a Traffic Management Center (TMC), there are various measures of control effectiveness which can be considered as objective functions. In the following, we enumerate several objective functions which are most widely considered in general DSO route choice models:

1. minimize total travel time;
2. minimize total travel cost or disutility;
3. minimize total number of vehicles during time period $[0, T]$;
4. minimize average congestion level during time period $[0, T]$;
5. minimize the length of the congested time period $[0, T]$.

We first consider the problem of minimizing the total travel time of all vehicles within a time period $[0, T]$. Using optimal control theory, the objective function for this dynamic system-optimal route choice problem is formulated as follows:

$$\min \quad \int_0^T \left\{ \sum_a u_a(t)\ \tau_a[x_a(t), u_a(t), v_a(t)] \right\} dt \tag{15.1}$$

In this objective function, if we replace the link travel time function $\tau_a(t)$ with the link travel *cost* function $\tilde{\tau}_a(t)$, then we obtain an objective function which minimizes the total travel cost during time period $[0, T]$:

$$\min \quad \int_0^T \left\{ \sum_a u_a(t)\ \tilde{\tau}_a[x_a(t), u_a(t), v_a(t)] \right\} dt \tag{15.2}$$

Equations (15.1) and (15.2) are different because the link travel cost function $\tilde{\tau}_a(t)$ includes other factors, such as automobile operating cost, link tolls, gasoline consumption, etc. The relationship between equations (15.1) and (15.2) and congestion pricing is discussed in detail in Section 15.4.

We now consider how to minimize the total number of vehicles traveling on the network during time period $[0, T]$. This objective function may be stated as follows:

$$\min \quad \int_0^T \sum_a [u_a(t) - v_a(t)] dt = \sum_a [x_a(T) - x_a(0)] \tag{15.3}$$

Since the initial value $x_a(0)$ is generally given, the above objective function is equivalent to

$$\min \quad \sum_a x_a(T) \tag{15.4}$$

Thus, we obtain an objective function which minimizes the total number of vehicles on the network at the final time $T$. This objective function is useful for reducing the average congestion level of peak-hour traffic when the final time $T$ is set within the peak-hour.

Next, we consider a quantitative definition of the congestion level for the network. The congestion level $D(t)$ is defined as an indicator of average congestion in the network at any time $t$. There are several possible approaches to defining $D(t)$:

1. the ratio of mean travel time to mean free-flow travel time;
2. the ratio of mean free-flow speed to mean flow speed;
3. mean relative density;

where the calculation of the means is flow weighted. For example, we can use the mean relative density as a measure of the level of congestion. The relative density $D_a(t)$ for link $a$ is then defined as

$$D_a(t) = \frac{x_a(t)/l_a}{e_{am}} \qquad \forall a \tag{15.5}$$

where $l_a$ is the link length and $e_{am}$ is the maximum density of traffic on link $a$. Therefore, the congestion level $D(t)$ for the network can be defined as

$$D(t) = \frac{\sum_a u_a(t)\ D_a(t)}{\sum_a u_a(t)} \tag{15.6}$$

where $u_a$ is the inflow on link $a$. Then, the objective function that minimizes the mean relative density during time period $[0, T]$ is expressed as

$$\min \qquad \int_0^T \left\{ \frac{\sum_a u_a(t)\ D_a(t)}{\sum_a u_a(t)} \right\} dt \tag{15.7}$$

The above objective functions can be applied to situations either with or without elastic departure times. For problems with elastic departure times, a special type of DSO route choice problem is to find the minimum time period $[0, T]$ if the total number of departure vehicles is known. This objective function can be stated as follows:

$$\min \qquad T \tag{15.8}$$

This type of DSO route choice problem is discussed in detail in Section 15.3.

## 15.2 Total Travel Time Minimization

### 15.2.1 The Model

We first consider the classic problem of minimizing the total travel time of all vehicles within a time period $[0, T]$. In this model, the O-D departure flows $f^{rs}(t)$ are given. In order to compare the DSO route choice model with the DUO route choice model, we use the constraints for the instantaneous DUO route choice model as given in Chapter 5. Using optimal control theory, the direct optimization model of the DSO route choice problem is formulated as follows.

$$\min_{u,v,x,e,E} \int_0^T \sum_a u_a(t)\ \tau_a[x_a(t), u_a(t), v_a(t)]\ dt \tag{15.9}$$

s.t.

Relationship between state and control variables:

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \tag{15.10}$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r, s \neq r; \tag{15.11}$$

Flow conservation constraints:

$$\sum_{a \in A(r)} \sum_p u_{ap}^{rs}(t) = f^{rs}(t) \qquad \forall r \neq s; s; \tag{15.12}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \tag{15.13}$$

$$\sum_{a \in B(s)} \sum_p v_{ap}^{rs}(t) = e^{rs}(t) \qquad \forall r, s; s \neq r; \tag{15.14}$$

Flow propagation constraints:

$$x_{ap}^{rs}(t) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[t + \tau_a(t)] - x_{bp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; \tag{15.15}$$

Definitional constraints:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \tag{15.16}$$

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \qquad \sum_{rs} x_a^{rs}(t) = x_a(t), \qquad \forall a; \tag{15.17}$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \quad u_{ap}^{rs}(t) \geq 0, \quad v_{ap}^{rs}(t) \geq 0 \quad \forall a, p, r, s; \tag{15.18}$$

$$e_p^{rs}(t) \geq 0, \qquad E_p^{rs}(t) \geq 0, \qquad \forall p, r, s; \tag{15.19}$$

Boundary conditions:

$$E_p^{rs}(0) = 0, \qquad \forall p, r, s; \tag{15.20}$$

$$x_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s. \tag{15.21}$$

The objective function is analogous to the objective function of the well-known static system-optimal (SO) model. The flow propagation constraints are based on the actual link travel time $\tau_a(t)$ instead of its fixed estimate $\bar{\tau}_a(t)$. In summary, the control variables are $u_{ap}^{rs}(t)$, $v_{ap}^{rs}(t)$, and $e_p^{rs}(t)$; the state variables are $x_{ap}^{rs}(t)$ and $E_p^{rs}(t)$.

The objective function can also be defined using link travel costs instead of link travel times, as already noted above. A generalized link travel cost function would include a weighted sum of travel time, atmospheric emissions, gasoline consumption, physical strain of driving, etc. After such a generalized link travel cost function is defined in any practical situation, the link travel time function can be replaced and the above DSO route choice problem directly applied to our purpose.

We note that we cannot conduct an analysis of optimality conditions similar to that for the instantaneous DUO route choice model. Since the actual link travel time $\tau_a(t)$ is a functional of flow variables $u_a(t)$, $v_a(t)$, $x_a(t)$, the first-order necessary conditions are very complex. The resulting marginal link cost has several terms which have cross-effects with link flows on downstream links. This marginal cost is so complex that an analytical expression is not meaningful.

## 15.2.2 Solution Algorithm

This DSO route choice model can be solved using the same algorithm presented in Chapter 6. We need only to revise the link cost functions for the LP subproblem. We reformulate the DSO route choice model as a discrete-time nonlinear program (NLP). Then the diagonalization technique and the Frank-Wolfe algorithm are employed to solve the NLP. In the diagonalization procedure, the estimated link travel time is updated iteratively. Then we apply the Frank-Wolfe technique to solve the NLP. An expanded time-space network is constructed so that each LP subproblem can be decomposed according to O-D pairs and can be viewed as a set of minimal-cost route problems. The flow propagation constraints which represent the relationship of link flows and travel times are satisfied in the modified minimal-cost route search so that only flow conservation constraints for links and nodes remain.

### Discrete DSO Route Choice Model

To convert our DSO route choice problem into an NLP, the assignment time interval $[0, T]$ is subdivided into $K$ small time increments. (The time increments are not necessarily equal.) In each diagonalization iteration, we modify the estimated actual link travel times in the flow propagation constraints in the following way so that each estimated travel time is equal to a multiple of the time increment.

$$\bar{\tau}_a(k) = i \quad \text{if} \quad i - 0.5 \leq \bar{\tau}_a(k) < i + 0.5,$$

where $i$ is an integer and $0 \leq i \leq K$. The optimal control program can then be reformulated as a discrete time NLP as follows:

$$\min_{u,v,x,E} \quad Z = \sum_{k=1}^{K} \sum_{a} u_a(k)\, \tau_a[x_a(k), u_a(k), v_a(k)] \tag{15.22}$$

s.t.

$$x_{ap}^{rs}(k+1) = x_{ap}^{rs}(k) + u_{ap}^{rs}(k) - v_{ap}^{rs}(k) \qquad \forall a, p, r, s; k = 1, \cdots, K; \tag{15.23}$$

$$E^{rs}(k+1) = E^{rs}(k) + \sum_{a \in B(s)} \sum_{p} v_{ap}^{rs}(k) \qquad \forall r; s \neq r; k = 1, \cdots, K; \tag{15.24}$$

$$\sum_{a \in A(r)} \sum_{p} u_{ap}^{rs}(k) = f^{rs}(k) \qquad \forall r \neq s; k = 1, \cdots, K; \tag{15.25}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(k) - \sum_{a \in A(j)} u_{ap}^{rs}(k) = 0 \qquad \forall j, p, r, s; j \neq r, s; k = 1, \cdots, K; \tag{15.26}$$

$$x_{ap}^{rs}(k) = \sum_{b \in \tilde{p}} \{x_{bp}^{rs}[k + \bar{\tau}_a(k)] - x_{bp}^{rs}(k)\} + \{E_p^{rs}[k + \bar{\tau}_a(k)] - E_p^{rs}(k)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; k = 1, \cdots, K+1; \tag{15.27}$$

$$u_{ap}^{rs}(k) \geq 0, \quad v_{ap}^{rs}(k) \geq 0, \quad x_{ap}^{rs}(k+1) \geq 0, \quad \forall a, p, r, s; k = 1, \cdots, K; \tag{15.28}$$

$$E_p^{rs}(k+1) \geq 0, \qquad \forall p, r, s; k = 1, \cdots, K; \tag{15.29}$$

$$E_p^{rs}(1) = 0 \qquad \forall p, r, s; \tag{15.30}$$

$$x_{ap}^{rs}(1) = 0, \qquad \forall a, p, r, s. \tag{15.31}$$

### Diagonalization/Frank-Wolfe Algorithm

Denote the subproblem variables as $p, q, y, \bar{E}$, corresponding to the main problem variables $u, v, x, E$. Applying the Frank-Wolfe algorithm to the minimization of the discretized DSO program requires, at each iteration, a solution of the following linear program (LP):

$$\begin{aligned} \min_{p,q,y,\bar{E}} \hat{Z} &= \nabla_u Z(u, v, x, E)\, p^T + \nabla_v Z(u, v, x, E)\, q^T \\ &+ \nabla_x Z(u, v, x, E)\, y^T + \nabla_E Z(u, v, x, E)\, \bar{E}^T \end{aligned} \tag{15.32}$$

s.t.

$$y_{ap}^{rs}(k+1) = y_{ap}^{rs}(k) + p_{ap}^{rs}(k) - q_{ap}^{rs}(k) \qquad \forall a, p, r, s; k = 1, \cdots, K; \tag{15.33}$$

$$\bar{E}^{rs}(k+1) = \bar{E}^{rs}(k) + \sum_{a \in B(s)} \sum_{p} q_{ap}^{rs}(k) \qquad \forall r; s \neq r; k = 1, \cdots, K; \tag{15.34}$$

$$\sum_{a\in A(r)}\sum_{p} p_{ap}^{rs}(k) = f^{rs}(k) \qquad \forall r \neq s; k = 1,\cdots,K; \tag{15.35}$$

$$\sum_{a\in B(j)} q_{ap}^{rs}(k) - \sum_{a\in A(j)} p_{ap}^{rs}(k) = 0 \qquad \forall j,p,r,s; j \neq r,s; k = 1,\cdots,K; \tag{15.36}$$

$$y_{ap}^{rs}(k) = \sum_{b\in \tilde{p}}\{y_{bp}^{rs}[k+\bar{\tau}_a(k)] - y_{bp}^{rs}(k)\} + \{\bar{E}_p^{rs}[k+\bar{\tau}_a(k)] - \bar{E}_p^{rs}(k)\}$$

$$\forall a \in B(j); j \neq r; p,r,s; k = 1,\cdots,K+1; \tag{15.37}$$

$$p_{ap}^{rs}(k) \geq 0,\quad q_{ap}^{rs}(k) \geq 0,\quad y_{ap}^{rs}(k+1) \geq 0,\quad \forall a,p,r,s; k = 1,\cdots,K; \tag{15.38}$$

$$\bar{E}_p^{rs}(k+1) \geq 0, \qquad \forall p,r,s; k = 1,\cdots,K; \tag{15.39}$$

$$\bar{E}_p^{rs}(1) = 0, \qquad \forall p,r,s; \tag{15.40}$$

$$y_{ap}^{rs}(1) = 0, \qquad \forall a,p,r,s. \tag{15.41}$$

Objective function (15.32) is equivalent to:

$$\begin{aligned}\hat{Z} &= \sum_{k=1}^{K}\sum_{r,s}\sum_{a,p}\left[\frac{\partial Z}{\partial u_{ap}^{rs}(k)}p_{ap}^{rs}(k) + \frac{\partial Z}{\partial v_{ap}^{rs}(k)}q_{ap}^{rs}(k) + \frac{\partial Z}{\partial x_{ap}^{rs}(k+1)}y_{ap}^{rs}(k+1)\right]\\ &+ \sum_{k=1}^{K}\sum_{r,s,p}\frac{\partial Z}{\partial E_p^{rs}(k+1)}\bar{E}_p^{rs}(k+1)\end{aligned} \tag{15.42}$$

The components of the gradient of $Z(u,v,x,E)$ with respect to the control and state variables $u$, $v$, $x$, $E$ are

$$t_{1a}(k) = \frac{\partial Z(u,v,x,E)}{\partial u_a(k)} = \tau_a(k) + u_a(k)\frac{\partial \tau_a(k)}{\partial u_a(k)} \qquad \forall a; k = 1,\cdots,K; \tag{15.43}$$

$$t_{2a}(k) = \frac{\partial Z(u,v,x,E)}{\partial v_a(k)} = u_a(k)\frac{\partial \tau_a(k)}{\partial v_a(k)} \qquad \forall a; k = 1,\cdots,K; \tag{15.44}$$

$$t_{3a}(k) = \frac{\partial Z(u,v,x,E)}{\partial x_a(k)} = u_a(k)\frac{\partial \tau_a(k)}{\partial x_a(k)} \qquad \forall a; k = 2,\cdots,K; \tag{15.45}$$

$$t_{3a}(K+1) = \frac{\partial Z(u,v,x,E)}{\partial x_a(K+1)} = 0 \qquad \forall a; \tag{15.46}$$

$$t_4^{rs}(k) = \frac{\partial \hat{Z}(u,v,x,E)}{\partial E^{rs}(k)} = 0 \qquad \forall r,s; k = 2,\cdots,K+1. \tag{15.47}$$

Therefore, the objective function can be rewritten as

$$\hat{Z} = \sum_{k=1}^{K}\sum_{r,s,a,p}\left[t_{1a}(k)p_{ap}^{rs}(k) + t_{2a}(k)q_{ap}^{rs}(k) + t_{3a}(k+1)y_{ap}^{rs}(k+1)\right] \tag{15.48}$$

As before, we define the diagonalization procedure as the outer iteration and the F-W procedure as the inner iteration in this combined algorithm. Denote the new solution at inner F-W iteration $(n+1)$ as

$$u_a^{(n+1)}(k) = u_a^{(n)}(k) + \alpha^{(n)}[u_a^{(n)}(k) - p_a^{(n)}(k)] \quad \forall a; k = 1, \cdots, K; \quad (15.49)$$

$$v_a^{(n+1)}(k) = v_a^{(n)}(k) + \alpha^{(n)}[v_a^{(n)}(k) - q_a^{(n)}(k)] \quad \forall a; k = 1, \cdots, K; \quad (15.50)$$

$$x_a^{(n+1)}(k) = x_a^{(n)}(k) + \alpha^{(n)}[x_a^{(n)}(k) - y_a^{(n)}(k)] \quad \forall a; k = 1, \cdots, K+1; \quad (15.51)$$

where $\alpha^{(n)}$ is the optimal step size of the one-dimensional search problem in the F-W algorithm at iteration $n$. The one-dimensional search problem is to find step size $\alpha^{(n)}$ that solves

$$\min_{0 \le \alpha^{(n)} \le 1} \sum_{k=1}^{K} \sum_{a} u_a^{(n+1)}(k)\ \tau_a[x_a^{(n+1)}(k), u_a^{(n+1)}(k), v_a^{(n+1)}(k)] \quad (15.52)$$

where $u_a^{(n+1)}(k)$, $v_a^{(n+1)}(k)$, $x_a^{(n+1)}(k)$ must be substituted using the above definitional equations. The algorithm for solving our DSO route choice model can then be stated as follows:

**Step 0: Initialization.**

Find an initial feasible solution $\{x_a^{(1)}(k)\}$, $\{u_a^{(1)}(k)\}$, $\{v_a^{(1)}(k)\}$, $\{E^{(1)}(k)\}$. Set the outer iteration counter $m = 1$.

**Step 1: Diagonalization.**

Find a new estimate of actual link travel time $\bar{\tau}_a^{(m)}(k)$ and solve the DSO program. Set the inner iteration counter $n = 1$.

[Step 1.1]: *Update.* Calculate $t_{1a}(k)$, $t_{2a}(k)$ and $t_{3a}(k)$ using equations (15.43)-(15.46).

[Step 1.2]: *Direction Finding.* Based on $\{t_{1a}(k)\}$, $\{t_{2a}(k)\}$ and $\{t_{3a}(k)\}$ and satisfying flow propagation constraints (15.2.2), search the minimal-cost route forward from each artificial origin to the super destination over an expanded time-space network for each O-D pair $rs$. Perform an all-or-nothing assignment, yielding subproblem solution $\{p_a(k)\}$, $\{q_a(k)\}$, $\{y_a(k)\}$, $\{\bar{E}^{rs}(k)\}$.

[Step 1.3]: *Line Search.* Find the optimal step size $\alpha^{(n)}$ that solves the one-dimensional search problem.

[Step 1.4]: *Move.* Find a new solution by combining $\{u_a(k)\}$, $\{v_a(k)\}$, $\{x_a(k)\}$, $\{E^{rs}(k)\}$ and $\{p_a(k)\}$, $\{q_a(k)\}$, $\{y_a(k)\}$ and $\{\bar{E}^{rs}(k)\}$.

[Step 1.5]: *Convergence Test for Inner Iterations.* If $n$ equals a prespecified number, go to step 2; otherwise, set $n = n + 1$ and go to step 1.1.

**Step 2: Convergence Test for Outer Iterations.**
If $\bar{\tau}_a^{(m)}(k) \simeq \bar{\tau}_a^{(m+1)}(k)$, stop. The current solution, $\{u_a(k)\}$, $\{v_a(k)\}$, $\{x_a(k)\}$ and $\{E^{rs}(k)\}$, is in a near DSO state; otherwise, set $m = m + 1$ and go to step 1.

In order to speed up convergence, an incremental assignment technique is suggested for finding a good initial solution before applying the diagonalization procedure. Since the linear subproblem can be decomposed by each artificial origin-destination pair, this problem is a good candidate for solution with parallel computing techniques.

## 15.3 DSO Route Choice with Elastic Departure Times

We now consider DSO route choice with elastic departure times, which is similar to the DUO departure time/route choice problem in terms of its problem statement. However, these two problems are different in that the DSO route choice with elastic departure times seeks to achieve the system-optimal objective by adjusting both departure time and routes. Thus, this problem is also a simultaneous departure time/route choice problem. The DSO route choice with elastic departure time is easier to formulate and the resulting optimal control program is easier to solve.

In these models, the O-D departure flows $f^{rs}(t)$ are variables. The cumulative number of departure vehicles from origin $r$ to destination $s$ at time $t$ is

$$F^{rs}(t) = \int_0^t f^{rs}(t) \qquad \forall r, s; r \neq s. \tag{15.53}$$

We assume that the total number of departures $F^{rs}(T)$ between O-D pair $rs$ in assignment period $[0, T]$ are given and all vehicles are ready to depart at time 0. In other words, we have

$$\int_0^T f^{rs}(t) = F^{rs}(T) \qquad \text{given} \qquad \forall r, s; r \neq s. \tag{15.54}$$

This type of O-D departure condition is also called an isoperimetric condition and is shown in Figure 15.1.

The auxiliary state equation for departing vehicles is

$$\frac{dF^{rs}(t)}{dt} = f^{rs}(t) \qquad \forall r, s \tag{15.55}$$

where the departure flow rate $f^{rs}(t)$ is an additional control variable and $F^{rs}$ is an additional state variable in the optimal control programs. The initial condition is

$$F^{rs}(0) = 0 \qquad \forall r, s. \tag{15.56}$$

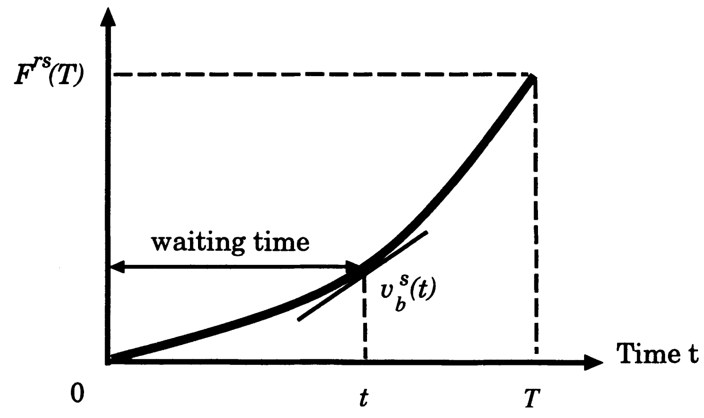



Figure 15.1: Isoperimetric O-D Departure Condition

Since the adjustment of departure times is associated with delays at origins, we adopt the method for handling spillback constraints in Chapter 4 to treat the delays at origins. We create a dummy link $b$ at each origin $r$ to accommodate the vehicles waiting at origin $r$. The state equation for dummy link $b$ at origin $r$ is as follows

$$\frac{dx_b^s(t)}{dt} = f^{rs}(t) - v_b^s(t) \qquad \forall b, r, s; b \in rs. \tag{15.57}$$

We assume the number of spillback vehicles at time 0 is zero. It follows that

$$x_b^s(0) = 0 \qquad \forall b, r, s; b \in rs. \tag{15.58}$$

Thus, the flow conservation equation for origin $r$ should be revised as

$$\sum_{a \in A(r)} u_a^{rs}(t) = v_b^s(t) \qquad \forall r \neq s; b \in rs. \tag{15.59}$$

As a result, the DSO route choice problem with isoperimetric O-D conditions is transferred into a conventional DSO route choice problem with fixed initial link states and fixed final auxiliary states. Of course, we also need the flow propagation constraint for dummy link $b$. Associated with the spillback constraints, link flow capacity constraints are necessary and placed in the constraint set for the DSO route choice problem. We assume there is no upper bound for the queue length $x_b^s(t)$ since an origin always has enough capacity to accommodate vehicles. The queuing delay at the origin is as follows

$$\tau_b(t) = \tau_b[x_b(t), v_b(t)] \qquad \forall b \in rs \tag{15.60}$$

where

$$x_b(t) = \sum_s x_b^s(t) \qquad v_b(t) = \sum_s v_b^s(t)$$

The total delay at origin $r$ is the waiting time plus possible spillback delay, i.e., $t + \tau_b[x_b(t), v_b(t)]$.

### 15.3.1 Fixed Final Time $T$

We first consider the problem of minimizing the total travel time of all vehicles within a time period $[0, T]$. Using optimal control theory, the equivalent optimization model of the dynamic system-optimal route choice problem is formulated as follows.

$$\min_{u,v,x,f,F,e,E} \quad \int_0^T \left\{ \sum_a u_a(t)\ \tau_a[x_a(t), u_a(t), v_a(t)] \right.$$
$$\left. + \quad \sum_r f^r(t)\ \{t + \tau_b[x_b(t), v_b(t)]\} \right\} dt \qquad (15.61)$$

s.t.

Relationship between state and control variables:

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \qquad (15.62)$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r, s \neq r; \qquad (15.63)$$

Flow conservation and spillback constraints:

$$\frac{dx_{bp}^s(t)}{dt} = f_p^{rs}(t) - v_{bp}^s(t) \qquad \forall b, p, r, s; b \in rs. \qquad (15.64)$$

$$\sum_{a \in A(r)} u_{ap}^{rs}(t) = v_{bp}^s(t) \qquad \forall p, r \neq s; b \in rs; \qquad (15.65)$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \qquad (15.66)$$

$$\sum_{a \in B(s)} \sum_p v_{ap}^{rs}(t) = e^{rs}(t) \qquad \forall r, s; s \neq r; \qquad (15.67)$$

Flow propagation constraints for links $a$, $b$:

$$x_{ap}^{rs}(t) = \sum_{d \in \tilde{p}} \{x_{dp}^{rs}[t + \tau_a(t)] - x_{dp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; \qquad (15.68)$$

$$x_{bp}^{s}(t) = \sum_{d \in \tilde{p}} \{x_{dp}^{rs}[t + \tau_b(t)] - x_{dp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_b(t)] - E_p^{rs}(t)\}$$

$$\forall b \in rs; p, r, s; \qquad (15.69)$$

Link Capacity:

$$x_a(t) \leq l_a e_{am} \qquad \forall a; \qquad (15.70)$$

$$v_a(t) \leq v_{am} \qquad \forall a; \qquad v_b(t) \leq v_{bm} \qquad \forall b \in rs; \qquad (15.71)$$

Definitional constraints:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \qquad (15.72)$$

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \qquad \sum_{rs} x_a^{rs}(t) = x_a(t), \qquad \forall a; \qquad (15.73)$$

$$\sum_{s} v_b^{s}(t) = v_b(t), \qquad \sum_{s} x_b^{s}(t) = x_b(t), \qquad \forall b \in rs; \qquad (15.74)$$

$$\sum_{p} f_p^{rs}(t) = f^{rs}(t), \qquad \sum_{s} f^{rs}(t) = f^{r}(t), \qquad \forall r; \qquad (15.75)$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \qquad u_{ap}^{rs}(t) \geq 0, \qquad v_{ap}^{rs}(t) \geq 0 \qquad \forall a, p, r, s; \qquad (15.76)$$

$$f_p^{rs}(t) \geq 0, \quad F_p^{rs}(t) \geq 0, \quad e_p^{rs}(t) \geq 0, \quad E_p^{rs}(t) \geq 0, \quad \forall p, r, s; \qquad (15.77)$$

$$x_{bp}^{s}(t) \geq 0, \qquad v_{bp}^{s}(t) \geq 0, \qquad \forall b \in rs; p, s; \qquad (15.78)$$

Boundary conditions:

$$F^{rs}(T) \qquad \text{given} \qquad \forall r, s; \qquad (15.79)$$

$$E_p^{rs}(0) = 0, \qquad \forall p, r, s; \qquad (15.80)$$

$$x_{ap}^{rs}(0) = 0, \qquad \forall a, p, r, s; \qquad x_{bp}^{s}(0) = 0, \qquad \forall b \in rs; p, s. \qquad (15.81)$$

The above model is similar to the DSO route choice model in Section 15.2 except for the addition of link capacity constraints, oversaturation constraints and the corresponding delays at origins. This model can be solved using a similar diagonalization/F-W algorithm. However, the time-space network expansion should include a set of dummy links for the spillback vehicles and origins.

## 15.3.2 Free Final Time $T$

We next consider the DSO route choice problem with elastic departure time and free final time $T$. We first discuss a model with an objective function which minimizes the total travel time in the network. This formulation is identical to equations (15.61)-(15.81) except that $T$ is now a variable which has to be determined to minimize the total travel time as shown in Figure 15.2. As shown in equation (15.61), the total travel time consists of two parts: 1) travel time in the network; 2) queuing delay at origins. When the time period $[0, T]$ is short (or the vehicles enter the network during a short time period $[0, T]$), the total queuing delay at origins is small, but the total travel time in the network is high because the network is more congested. On the other hand, when the time period $[0, T]$ is long, the total queuing delay at origins becomes larger. Since the departing vehicles are spread out more evenly in a longer time period $[0, T]$), the network is less congested and the total travel time in the network is smaller. Thus, there exists an optimal final time $T^*$ by which the total travel time achieves its minimum.

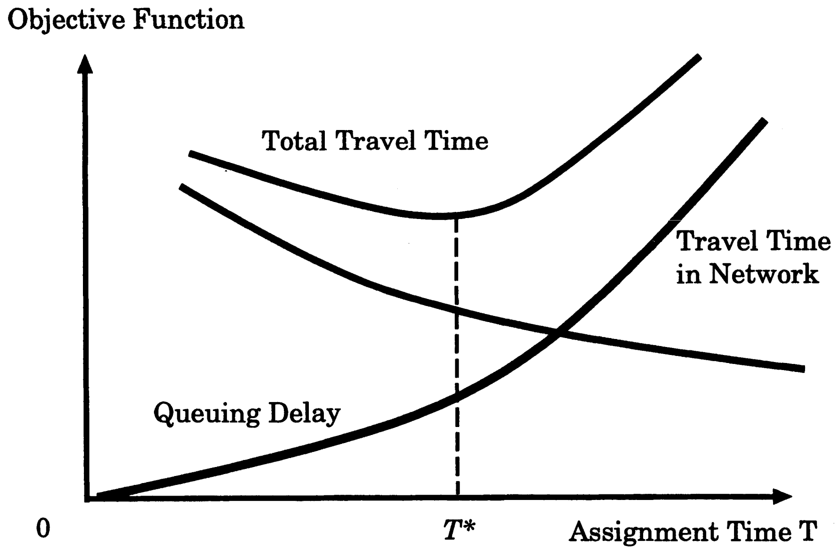



Figure 15.2: Optimal Assignment Time in DSO Route Choice Problem

We note that the queuing delay at origins becomes important in determining the minimal total travel time and the optimal final time $T^*$ in the minimization problem with free optimal time period $T$. If the waiting time or queuing delay at origins is not considered in the objective function, the optimal time period $T$ becomes infinite so that the departure vehicles are assigned onto the network over an infinite time horizon. There will be no congestion in the network and the total congestion delay will approach zero. However, our major interest in the free final time period is to identify the minimum $T$ by which all departing vehicles leave the origins. This problem is now discussed in more detail.

In some emergency situation, we may wish to evacuate all persons from one place to other places within a minimal time period. This typical evacuation problem may have wide applications in emergency situations like hurricanes, earthquakes and fires. In this problem, the performance index of interest is the *elapsed time* to transfer the system from its initial state to a specified state.

The assignment time period $[0, T]$ is free and is a variable in this problem. The objective function is

$$\min \quad T$$

Under this objective function, an extreme requirement is that at the end of the time period there are no vehicles on the network or at least no vehicles on links within a certain range of the emergency areas. Thus, the associated additional boundary constraints for physical links $a$ and dummy links $b$ are

$$x_a(T) = 0 \qquad \forall a, \qquad x_b(T) = 0 \qquad \forall b$$

To be more practical, we may only require that vehicles be cleared at origins at final time $T$. It follows that

$$x_b(T) = 0 \qquad \forall b$$

The equivalent optimal control program of the time-optimal route choice problem is formulated as follows.

$$\min \quad T \tag{15.82}$$

s.t.

Relationship between state and control variables:

$$\frac{dx_{ap}^{rs}}{dt} = u_{ap}^{rs}(t) - v_{ap}^{rs}(t) \qquad \forall a, p, r, s; \tag{15.83}$$

$$\frac{dE_p^{rs}(t)}{dt} = e_p^{rs}(t) \qquad \forall p, r, s \neq r; \tag{15.84}$$

Flow conservation and spillback constraints:

$$\frac{dx_{bp}^{s}(t)}{dt} = f_p^{rs}(t) - v_{bp}^{s}(t) \qquad \forall b, p, r, s; b \in rs. \tag{15.85}$$

$$\sum_{a \in A(r)} u_{ap}^{rs}(t) = v_{bp}^{s}(t) \qquad \forall p, r \neq s; b \in rs; \tag{15.86}$$

$$\sum_{a \in B(j)} v_{ap}^{rs}(t) = \sum_{a \in A(j)} u_{ap}^{rs}(t) \qquad \forall j, p, r, s; j \neq r, s; \tag{15.87}$$

$$\sum_{a \in B(s)} \sum_{p} v_{ap}^{rs}(t) = e^{rs}(t) \qquad \forall r, s; s \neq r; \qquad (15.88)$$

Flow propagation constraints for links $a$, $b$:

$$x_{ap}^{rs}(t) = \sum_{d \in \tilde{p}} \{x_{dp}^{rs}[t + \tau_a(t)] - x_{dp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_a(t)] - E_p^{rs}(t)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; \qquad (15.89)$$

$$x_{bp}^{s}(t) = \sum_{d \in \tilde{p}} \{x_{dp}^{rs}[t + \tau_b(t)] - x_{dp}^{rs}(t)\} + \{E_p^{rs}[t + \tau_b(t)] - E_p^{rs}(t)\}$$

$$\forall b \in rs; p, r, s; \qquad (15.90)$$

Link Capacity:

$$x_a(t) \leq l_a e_{am} \qquad \forall a; \qquad (15.91)$$

$$v_a(t) \leq v_{am} \quad \forall a; \qquad v_b(t) \leq v_{bm} \quad \forall b \in rs; \qquad (15.92)$$

Definitional constraints:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \qquad (15.93)$$

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \qquad \sum_{rs} x_{a}^{rs}(t) = x_a(t), \qquad \forall a; \qquad (15.94)$$

$$\sum_{s} v_{b}^{s}(t) = v_b(t), \qquad \sum_{s} x_{b}^{s}(t) = x_b(t), \qquad \forall b \in rs; \qquad (15.95)$$

$$\sum_{p} f_{p}^{rs}(t) = f^{rs}(t), \qquad \sum_{s} f^{rs}(t) = f^{r}(t), \qquad \forall r; \qquad (15.96)$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \quad u_{ap}^{rs}(t) \geq 0, \quad v_{ap}^{rs}(t) \geq 0 \quad \forall a, p, r, s; \qquad (15.97)$$

$$f_p^{rs}(t) \geq 0, \; F_p^{rs}(t) \geq 0, \; e_p^{rs}(t) \geq 0, \; E_p^{rs}(t) \geq 0, \; \forall p, r, s; \qquad (15.98)$$

$$x_{bp}^{s}(t) \geq 0, \qquad v_{bp}^{s}(t) \geq 0, \qquad \forall b \in rs; p, s; \qquad (15.99)$$

Boundary conditions:

$$F^{rs}(T) \qquad \text{given} \qquad \forall r, s; \qquad (15.100)$$

$$E_p^{rs}(0) = 0, \qquad \forall p, r, s; \qquad (15.101)$$

$$x_{ap}^{rs}(0) = 0, \; \forall a, p, r, s; \; x_{bp}^{s}(0) = 0, \; x_{bp}^{s}(T) = 0, \; \forall b \in rs; p, s. \qquad (15.102)$$

Compared with other DSO route choice models, the time-optimal model has a distinct objective function, min $T$. At the final time $T$, vehicles at origins are cleared as shown in boundary conditions (15.102). Figure 15.3 illustrates the relationship between the optimal final time $T^*$ and cumulative departures between O-D pair $rs$.

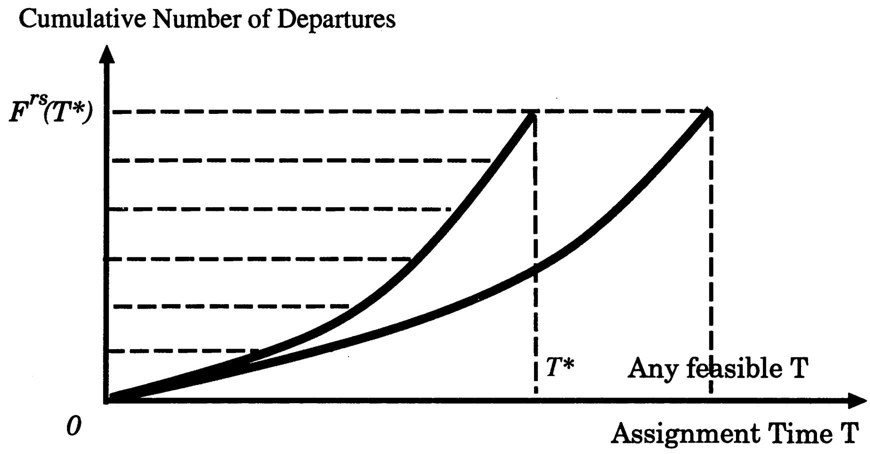



Figure 15.3: Optimal Time Period in Time-Optimal Problem

## 15.4 Dynamic Congestion Pricing

We now consider the dynamic congestion pricing problem in which the dynamic system-optimal objective can be achieved while preserving the dynamic user-optimal route choice properties. Previous research on congestion pricing has concentrated on policy and practical implementation issues. A summary of recent changes in policy is given by Small (1992), emphasizing the changing political acceptance of congestion pricing. The theory of marginal cost pricing has been explored by many researchers in the context of static transportation networks; for example, see Beckmann et al (1956), Dafermos (1972) and Smith (1979). Few theoretical studies on congestion pricing are related to large transportation networks in real time. Recently, some models have been proposed by de Palma and Lindsey (1992) and Ghali and Smith (1993). Both of them use simple networks to explore the properties of dynamic tolls.

We first discuss some possible dynamic congestion pricing strategies. Following a summary of dynamic network constraints, we present two kinds of dynamic toll strategies. Two dynamic link toll models are then formulated as two bilevel programs which are based on two different kinds of route choice assumptions.

### 15.4.1 Various Dynamic Congestion Pricing Strategies

Information and control have been identified as two major approaches to combat traffic congestion in an ATMIS system. Recently, congestion pricing has received increasing attention from policy makers as an effective measure of controlling congestion. With advances in ATMIS technology, real-time congestion pricing is becoming increasingly feasible.

Dynamic tolls can be collected with the application of automatic vehicle identification (AVI) technology. In conjunction with a dynamic route guidance system (DRGS) , effective traffic controls and congestion pricing can be implemented together to influence the routing strategies by either a central controller or individual drivers so that congestion levels in the transportation network can be controlled or adjusted.

The Federal Highway Administration (FHWA) has invited applications from local authorities for demonstration projects on congestion pricing. In the evaluation of strategic IVHS System Architectures, congestion pricing was identified as one of the most important control schemes in future IVHS systems. What seemed impossible only a few years ago seems now possible with the ATMIS technological achievements and the change of policy, which in turn motivates the development of dynamic congestion pricing models for large scale transportation networks. We envision the application of such models as contributing to the evaluation of proposed ATMIS systems. Eventually, such models may prove useful as well in the operation of such systems. In the short run, however, our principal objective is to improve understanding of the properties of dynamic congestion pricing models defined on large, complex road networks.

This section seeks to investigate possible dynamic congestion pricing strategies using dynamic network models. Our focus is on technical aspects rather than policy issues. Furthermore, in this section we concentrate on theoretical models which explore possibilities of congestion pricing in conjunction with dynamic system-optimal route choice models. Traffic controls, including both surface street signal control and freeway ramp control, are assumed to be fixed in the current models. By considering the impact of congestion pricing, we focus our attention on routing strategies instead of traffic signal controls.

Consider an ideal situation with an AVI system installed and an automatic toll debiting system available. We also assume each vehicle is provided with perfect traffic information and travelers will comply with the user-optimal route guidance instructions. As discussed in Section 15.1, system optimal objectives can be defined in a variety of ways in dynamic transportation network problems. Thus, there are more toll strategies in dynamic problems than in their static counterpart. Because of the difficulty of keeping track of each vehicle's route, we focus our attention on non-route based dynamic tolls. This class of dynamic congestion tolls can be classified as link toll and area toll, both defined based on the usage of road capacity. The link toll is charged for vehicles present on that link and is changing from link to link. The area toll is charged for vehicles present within a congested area and is uniform across the area. It applies to any vehicle traveling in the network during the toll time period. In the following, we mainly discuss different modeling aspects for link tolls. We evaluate the performance of the congestion toll policy using the reduction of traffic congestion or total travel time on the network.

Within the link toll category, we consider two types of dynamic congestion pricing strategies under two kinds of route choice behavior assumptions, namely instantaneous DUO and ideal DUO. Before discussing the pricing strategies, we summarize the dynamic network constraints for such models.

## 15.4.2 Dynamic Network Constraints

In the following bilevel programs, the O-D departure flows $f^{rs}(t)$ are given. In order to model a more practical situation, spillback constraints are added in the constraint set. As we discussed in Chapter 4 and the previous section, we define a dummy link $b$ at each origin $r$ to accommodate the spillback vehicles. The queuing delay at origin $r$ is $\tau_b[x_b(t), v_b(t)]$. The dynamic network constraints for our models are summarized as follows.

Relationship between state and control variables:

$$\frac{dx^{rs}_{ap}}{dt} = u^{rs}_{ap}(t) - v^{rs}_{ap}(t) \qquad \forall a, p, r, s; \tag{15.103}$$

$$\frac{dE^{rs}_p(t)}{dt} = e^{rs}_p(t) \qquad \forall p, r, s \neq r; \tag{15.104}$$

Flow conservation and spillback constraints:

$$\frac{dx^{s}_{bp}(t)}{dt} = f^{rs}_p(t) - v^{s}_{bp}(t) \qquad \forall b, p, r, s; b \in rs. \tag{15.105}$$

$$\sum_{a \in A(r)} u^{rs}_{ap}(t) = v^{s}_{bp}(t) \qquad \forall p, r \neq s; b \in rs; \tag{15.106}$$

$$\sum_{a \in B(j)} v^{rs}_{ap}(t) = \sum_{a \in A(j)} u^{rs}_{ap}(t) \qquad \forall j, p, r, s; j \neq r, s; \tag{15.107}$$

$$\sum_{a \in B(s)} \sum_{p} v^{rs}_{ap}(t) = e^{rs}(t) \qquad \forall r, s; s \neq r; \tag{15.108}$$

Flow propagation constraints for links $a$, $b$:

$$x^{rs}_{ap}(t) = \sum_{d \in \tilde{p}} \{x^{rs}_{dp}[t + \tau_a(t)] - x^{rs}_{dp}(t)\} + \{E^{rs}_p[t + \tau_a(t)] - E^{rs}_p(t)\}$$

$$\forall a \in B(j); j \neq r; p, r, s; \tag{15.109}$$

$$x^{s}_{bp}(t) = \sum_{d \in \tilde{p}} \{x^{rs}_{dp}[t + \tau_b(t)] - x^{rs}_{dp}(t)\} + \{E^{rs}_p[t + \tau_b(t)] - E^{rs}_p(t)\}$$

$$\forall b \in rs; p, r, s; \tag{15.110}$$

Link Capacities:

$$x_a(t) \leq l_a e_{am} \qquad \forall a; \tag{15.111}$$

$$v_a(t) \leq v_{am} \quad \forall a; \qquad v_b(t) \leq v_{bm} \quad \forall b \in rs; \tag{15.112}$$

Definitional constraints:

$$\sum_{rsp} u_{ap}^{rs}(t) = u_a(t), \qquad \sum_{rsp} v_{ap}^{rs}(t) = v_a(t), \qquad \forall a; \tag{15.113}$$

$$\sum_{rsp} x_{ap}^{rs}(t) = x_a(t), \qquad \sum_{rs} x_a^{rs}(t) = x_a(t), \qquad \forall a; \tag{15.114}$$

$$\sum_{s} v_b^s(t) = v_b(t), \qquad \sum_{s} x_b^s(t) = x_b(t), \qquad \forall b \in rs; \tag{15.115}$$

$$\sum_{p} f_p^{rs}(t) = f^{rs}(t), \qquad \sum_{s} f^{rs}(t) = f^r(t), \qquad \forall r; \tag{15.116}$$

Nonnegativity conditions:

$$x_{ap}^{rs}(t) \geq 0, \quad u_{ap}^{rs}(t) \geq 0, \quad v_{ap}^{rs}(t) \geq 0 \quad \forall a, p, r, s; \tag{15.117}$$

$$f_p^{rs}(t) \geq 0, \;\; F_p^{rs}(t) \geq 0, \;\; e_p^{rs}(t) \geq 0, \;\; E_p^{rs}(t) \geq 0, \;\; \forall p, r, s; \tag{15.118}$$

$$x_{bp}^{s}(t) \geq 0, \qquad v_{bp}^{s}(t) \geq 0, \qquad \forall b \in rs; p, s; \tag{15.119}$$

Boundary conditions:

$$F^{rs}(T) \qquad \text{given} \qquad \forall r, s; \tag{15.120}$$

$$E_p^{rs}(0) = 0, \qquad \forall p, r, s; \tag{15.121}$$

$$x_{ap}^{rs}(0) = 0, \quad \forall a, p, r, s; \quad x_{bp}^{s}(0) = 0, \quad \forall b \in rs; p, s. \tag{15.122}$$

### 15.4.3 Tolls Based on Instantaneous DUO Route Choice

Denote the time-dependent toll on link $a$ at time $t$ as $\gamma_a(t)$ in dollars. For simplicity, the total link travel cost on link $a$ at time $t$ is

$$\tilde{c}_a(t) = \alpha_a + \beta_a \; c_a(t) + \gamma_a(t) \tag{15.123}$$

where $\alpha_a$ is a term representing the fixed cost (dollars) on link $a$ and $\beta_a$ is a time-independent parameter transforming travel time (minutes) into travel cost (dollars). The toll $\gamma_a(t)$ is collected to achieve a system-optimal flow pattern in the network while preserving DUO route choice properties. Thus, the objective of our problem is to minimize the total travel time over the entire network during time period $[0, T]$. This dynamic toll problem can be easily formulated as a route-based model. However, a route-based toll is hard to collect in practice and solving a route-based model requires explicit route enumeration, which is infeasible for a large network. Thus, we formulate a link-based model which overcomes these difficulties. We use a leader-follower game to formulate such a dynamic congestion pricing problem. The objective of the upper-level problem (or the leader of the game) is to minimize the total travel time; the decision variable is the dynamic link toll $\gamma_a(t)$. The instantaneous DUO route choice model is formulated as the lower-level problem (or the follower of the game), in which the decision variables are the link flow variables $u_a(t)$, $v_a(t)$ and $x_a(t)$. The upper level problem is a time-dependent minimization problem, whereas the lower-level problem is formulated as a variational inequality.

We first briefly describe the variational inequality formulation of the instantaneous DUO route choice problem for the lower-level problem. The formulation is a simplified version of the link-based multi-group VI model for the instantaneous DUO route choice problem in Chapter 12. For simplicity, we only consider one group in this toll model. Denote $\tilde{\psi}_p^{rs}(t)$ as the instantaneous route travel cost from origin $r$ to destination $s$ at time $t$. The instantaneous route travel cost for all allowable routes is computed using the following formula

$$\tilde{\psi}_p^{is}(t) = \sum_{a \in isp} \tilde{c}_a(t) \qquad \forall p, i, s \tag{15.124}$$

where the summation is over all links $a$ on route $p$.

Recall the definition of the link-cost-based instantaneous DUO state as follows.

> **Link-Cost-Based Instantaneous DUO State:** *If, for any departure flow from each decision node to each destination node at each instant of time, the instantaneous travel costs equal the minimal instantaneous route travel cost, the dynamic traffic flow over the network is in a link-cost-based instantaneous dynamic user-optimal state.*

Define $\tilde{\sigma}^{is}(t)$ as the minimal instantaneous route travel cost from node $i$ to destination $s$ at time $t$. The asterisk denotes that the travel time is computed using link-cost-based instantaneous DUO traffic flows. Denote $\tilde{\Theta}_a^{s^*}(t)$ as the difference between the minimal instantaneous travel cost from node $j$ to destination $s$ and the instantaneous travel cost from node $i$ to destination $s$ plus the instantaneous travel cost on link $a$ at time $t$. It follows that

$$\tilde{\Theta}_a^{s^*}(t) = \tilde{\sigma}^{js^*}(t) + \tilde{c}_a^*(t) - \tilde{\sigma}^{is^*}(t) \qquad \forall a, s; a = (i, j). \tag{15.125}$$

Thus, the link-cost-based instantaneous DUO route choice conditions can be summarized as follows:

$$\tilde{\Theta}_a^{s^*}(t) \geq 0 \qquad \forall a = (i, j), s; \tag{15.126}$$

$$u_a^{s^*}(t)\, \tilde{\Theta}_a^{s^*}(t) = 0 \qquad \forall a = (i, j), s; \tag{15.127}$$

$$u_a^s(t) \geq 0 \qquad \forall a = (i, j), s. \tag{15.128}$$

Note that the above conditions also apply to dummy link $b$ created at each origin $r$ to accomodate spillback flows at origin $r$. Then, the equivalent variational inequality formulation of the link-cost-based instantaneous DUO route choice conditions (15.126)-(15.128) may be stated as follows.

*Theorem 15.1.* The dynamic traffic flow pattern satisfying network constraint set (15.103)-(15.122) is in a link-cost-based instantaneous DUO route choice state if and only if it satisfies the variational inequality:

$$\int_0^T \sum_r \sum_a \tilde{\Theta}_a^{r*}(t) \left\{ u_a^r(t) - u_a^{r*}(t) \right\} dt \geq 0 \tag{15.129}$$

In the above variational inequality, the summation over link $a$ includes dummy link $b$. The proof of the necessity and sufficiency of the above variational inequality is similar to that in Chapter 12.

Our upper-level problem is a dynamic system-optimal problem which minimizes the total travel time on the network during period $[0, T]$. Since we have spillback constraints, we have a dummy link $b$ at each origin $r$ to accommodate the spillback vehicles. As before, the queuing delay at origin $r$ is $\tau_b(t)[x_b(t), v_b(t)]$. Thus, the bilevel program is formulated as

$$\min_{\gamma} \qquad \int_0^T \left\{ \sum_a u_a(t)\, \tau_a(t) + \sum_b f^r(t)\, \tau_b(t) \right\} dt \tag{15.130}$$

where $u$, $v$ and $x$ solve the following variational inequality:

$$\int_0^T \sum_r \sum_a \tilde{\Theta}_a^{r*}(t) \left\{ u_a^r(t) - u_a^{r*}(t) \right\} dt \geq 0 \tag{15.131}$$

In the upper-level problem, we minimize the total travel *time* instead of total travel *cost*, because the link toll policy is designed to control the total congestion level. Since the analysis on optimality conditions of the bilevel problem is very complicated, it is impossible to obtain a simple analytical expression for link tolls in terms of link travel times. Thus, a simple toll similar to the conventional marginal cost in static problems does not exist unless some relaxation methods are used.

In the bilevel program, the variational inequality has to be solved subject to the network flow constraint set defined in the previous section. In general, the above bilevel program is difficult to solve. In Chapter 12, we demonstrated that under relaxation, the lower level variational inequality can be transformed into the equivalent optimal control model presented in Chapter 5. An efficient algorithm including diagonalization and Frank-Wolfe techniques was proposed in Chapter 6 to solve this variational inequality. Thus, the lower level problem can be solved exactly. Since the upper level problem is linear, we expect that an iterative heuristic can be used to solve the bilevel program.

## 15.4.4 Tolls Based on Ideal DUO Route Choice

After a link toll $\gamma_a(t)$ is imposed, the final link travel cost on link $a$ at time $t$ is

$$\tilde{\tau}_a(t) = \alpha_a + \beta_a\, \tau_a(t) + \gamma_a(t) \tag{15.132}$$

where $\alpha_a$ is a term representing the fixed cost (dollars) on link $a$ and $\beta_a$ is a time-independent parameter to transform travel time (minutes) into travel cost (dollars). Similar to the previous instantaneous DUO, this problem is formulated in a bilevel structure. In the upper level problem, the total travel *time* is minimized. In the lower level problem, we have a variational inequality problem which equilibrates the actual route travel *costs* based on travel times and link tolls.

We now discuss how we formulate a variational inequality for the lower-level ideal DUO route choice problem. The formulation is a simplified version of the link-based multi-group VI model for the ideal DUO route choice problem in Chapter 13. For simplicity, we only consider one group in this toll model. Denote $\tilde{\eta}_p^{rs}(t)$ as the actual route travel cost from origin $r$ to destination $s$ at time $t$. Also denote $\tilde{\pi}^{rs}(t)$ as the minimal actual route travel cost from origin $r$ to destination $s$ at time $t$, and $\bar{\pi}^{rs}(t)$ as the corresponding actual route travel time from origin $r$ to destination $s$ at time $t$. We also need to use a recursive formula to compute the route travel cost $\tilde{\eta}_p^{rs}(t)$ for all allowable routes. Assume route $p$ consists of nodes $(r, 1, 2, \cdots, i, \cdots, s)$. Denote $\tilde{\eta}_p^{rj}(t)$ as the travel disutility *actually* experienced over route $p$ from origin $r$ to node $j$ by vehicles departing origin $r$ at time $t$. Then, a recursive formula for route travel cost $\tilde{\eta}_p^{rs}(t)$ is:

$$\tilde{\eta}_p^{rj}(t) = \tilde{\eta}_p^{r(j-1)}(t) + \tilde{\tau}_a[t + \eta_p^{r(j-1)}(t)] \qquad \forall p, r, j; j = 1, 2, \cdots, s;$$

where link $a = (j-1, j)$ and time $[t + \eta_p^{r(j-1)}(t)]$ is the arrival time instant at link $a$.

Recall the definition of the travel-cost-based ideal DUO state as follows.

**Travel-Cost-Based Ideal DUO State:** *If, for each group and each O-D pair at each instant of time, the actual travel* costs *for all routes that are being used equal the minimal actual route travel* cost, *the dynamic traffic flow over the network is in a travel-cost-based ideal dynamic user-optimal state.*

Denote $\tilde{\Omega}_a^{rj^*}(t)$ as the difference of the minimal travel cost from $r$ to $j$ and the travel cost from $r$ to $j$ via the minimal travel cost route from $r$ to $i$ and link $a$ for vehicles departing from origin $r$ at time $t$. It follows that

$$\tilde{\Omega}_a^{rj^*}(t) = \tilde{\pi}^{ri^*}(t) + \tilde{\tau}_a[t + \bar{\pi}^{ri^*}(t)] - \tilde{\pi}^{rj^*}(t) \qquad \forall a, r; a = (i, j). \qquad (15.133)$$

We then rewrite the link-cost-based ideal DUO route choice conditions:

$$\tilde{\Omega}_a^{rj^*}(t) \geq 0 \qquad \forall a = (i, j), r; \qquad (15.134)$$

$$u_a^{rs^*}[t + \bar{\pi}^{ri^*}(t)]\, \tilde{\Omega}_a^{rj^*}(t) = 0 \qquad \forall a = (i, j), r, s; \qquad (15.135)$$

$$u_a^{rs}[t + \bar{\pi}^{ri^*}(t)] \geq 0 \qquad \forall a = (i, j), r, s. \qquad (15.136)$$

Note that the above conditions also apply to dummy link $b$ created at each origin $r$ to accomodate spillback flows at origin $r$. The equivalent variational inequality formulation of link-cost-based ideal DUO route choice conditions (15.134)-(15.136) may be stated as follows.

*Theorem 15.2.* The dynamic traffic flow pattern satisfying constraints (15.103)-(15.122) is in a link-cost-based ideal DUO route choice state if and only if it satisfies the variational inequality:

$$\int_0^T \sum_{rs} \sum_a \tilde{\Omega}_a^{rj^*}(t) \left\{ u_a^{rs}[t + \bar{\pi}^{ri^*}(t)] - u_a^{rs^*}[t + \bar{\pi}^{ri^*}(t)] \right\} dt \geq 0 \quad (15.137)$$

In the above variational inequality, the summation over link $a$ includes dummy link $b$. The proof of the necessity and sufficiency of the variational inequality is similar to that in Chapter 13.

The dynamic link toll is designed so the total travel time on the network is minimized for period $[0, T]$. Thus, the bilevel program is formulated as

$$\min_{\gamma} \quad \int_0^T \left\{ \sum_a u_a(t)\, \tau_a(t) + \sum_b f^r(t)\, \tau_b(t) \right\} dt \qquad (15.138)$$

where $u$, $v$ and $x$ solve the following variational inequality:

$$\int_0^T \sum_{rs} \sum_a \tilde{\Omega}_a^{rj^*}(t) \left\{ u_a^{rs}[t + \bar{\pi}^{ri^*}(t)] - u_a^{rs^*}[t + \bar{\pi}^{ri^*}(t)] \right\} dt \geq 0 \quad (15.139)$$

Because the optimality conditions for this bilevel program are very complex, no simple analytical link toll can be obtained for even a two-parallel-link network. Thus, the equivalence of the toll to the difference of marginal and unit link costs in the static network model does not exist in a dynamic congestion pricing problem.

## 15.5 Notes

Merchant and Nemhauser (1978a, 1978b) presented a dynamic system-optimal (DSO) route choice model for a many-to-one network. Subsequently, Carey (1987) reformulated the Merchant-Nemhauser problem as a convex nonlinear program which has analytical and computational advantages over the original formulation. Ho (1980) solved the same model by successively optimizing a sequence of linear programs. Later on, Ho (1990) presented a nested decomposition algorithm for the same problem and implemented this algorithm on a hypercube computer.

DSO problems have also been studied systematically by Ran (1989). For DSO problems with minimal travel costs, Ran (1989) suggested several algorithms in addition to the relaxation and Frank-Wolfe algorithm. Among those algorithms, the Time Decomposition Algorithm was highly recommended. The Spatial Decomposition Algorithm and Sequential Gradient Restoration Algorithm were also discussed. Recently, many simulation-based DSO route choice models were proposed by various researchers, especially for freeway corridor problems (Mahmassani et al 1993 and Chang et al 1993). Those models provide another approach to studying DSO route choice problems.

We have formulated two types of congestion pricing models for a dynamic transportation network. Through the formulation, we find that the conventional marginal cost pricing strategy in static networks is no longer applicable to dynamic congestion pricing problems. To find appropriate congestion pricing strategies in dynamic networks requires much more computational effort. Simple analytical results are not available.

We expect that our proposed models can function as tools in the evaluation of possible congestion pricing strategies in light of evolving IVHS technologies. Eventually, they are expected to work together with traffic control schemes to combat traffic congestion in urban areas and become on-line operational tools in ATMIS systems.

# Chapter 16

# Link Travel Time Functions for Dynamic Network Models

Extensive research has occurred in recent years on dynamic transportation network models, and especially on dynamic route choice models; these models have important applications in future ATIS and ATMS systems. However, most of the existing models lack a basis in traffic engineering. A significant problem for dynamic route choice is that the traditional BPR (Bureau of Public Roads, the predecessor of the Federal Highway Administration, U.S. DOT) volume-delay function is not applicable to a time-dependent traffic network. Meanwhile, since no proper dynamic link travel time functions exist, current dynamic route choice models assume various functional forms which are either too abstract or cannot provide realistic travel time estimates, even for a small network. Thus, it is becoming increasingly urgent to develop a set of time-dependent link travel time functions for dynamic route choice problems.

In this chapter, the independent variables necessary to describe the dynamic traffic flow and estimate the corresponding time-dependent travel time over a highway link are discussed. In order to standardize the dynamic route choice formulation to be used in ATIS and ATMS applications, this chapter seeks to provide a solid foundation based on the principles of traffic engineering. For the purpose of short-term travel time forecasting, dynamic link travel time functions are also necessary to transform traffic flow data from probe vehicles or roadway detectors into travel times. The application of those functions in IVHS projects can also be expected.

Link travel time or delay functions have been extensively studied in traffic flow theory and traffic engineering research. These functions can be classified based on road types. In general, the following types of roadway links have different link travel time functions:

1. Arterial Streets

    (a) Links with Signalized Intersections

  - Fixed Signal Control
  - Actuated Signal Control

(b) Links with Unsignalized Intersections

  - Major/Minor Priority Intersections
  - All-Way-Stop Intersections

2. Freeways

  - Freeway Segment
  - Ramps
  - Weaving Sections

3. Local Streets

  - Stop/Yield Control
  - No Control

The objective of this chapter is to review currently available delay models, identify suitable functions, and develop them into dynamic link travel time functions which would be applicable to dynamic route choice models. The focus is on exploring dynamic travel time functions for signalized arterial network links and freeway segments. In Section 16.1, we discuss the classification of dynamic link travel times for various applications. In Sections 16.2 and 16.3, travel time functions for arterials with long and short time horizons are discussed separately, and two sets of functions are recommended for dynamic route choice models. The implications of those functional forms are analyzed in Section 16.4 and some modifications for dynamic models are suggested. In Section 16.5, we propose dynamic travel time functions for freeway segments.

## 16.1 Functions for Various Purposes

For an arterial link, travel time is considered to consist of two main components. The first is the travel time (or cruise time) over the uncongested portion of the link; the second is the congested travel time or queuing delay at the intersection, plus the travel time through the intersection to the downstream link. For a freeway segment, link travel time is also considered to consist of two main components. The first is the uncongested cruise time over the link; the second is the congested travel time or queuing delay on the link. Nevertheless, we use similar formulae for link travel time functions for both arterial and freeway segments.

Recall that the *instantaneous* travel time $c_a(t)$ at time $t$ is the travel time that is experienced by vehicles traversing link $a$ when prevailing traffic conditions remain unchanged. It is the sum of two components:

1. an instantaneous flow-dependent cruise time $D_{a1}(t)$ over the first part of link $a$ for an arterial segment; for a freeway segment, the uncongested cruise time over the link;

2. an instantaneous queuing delay $D_{a2}(t)$.

It follows that

$$c_a(t) = D_{a1}(t) + D_{a2}(t). \tag{16.1}$$

The instantaneous route travel time function $\psi_p^{rs}(t)$ for each route $p$ between O-D pair $rs$ is defined as the sum of the instantaneous link travel times over all links in route $p$:

$$\psi_p^{rs}(t) = \sum_{a \in rsp} c_a(t) \qquad \forall p, r, s. \tag{16.2}$$

Thus, the instantaneous route travel time is that time experienced by a vehicle, if prevailing traffic conditions do not vary until the vehicle reaches its destination. This instantaneous route travel time provides a first approximation to the time-dependent vehicle travel time.

Also recall that $\tau_a(t)$ is the actual travel time over link $a$ for vehicles entering link $a$ at time $t$. Similarly, $\eta_p^{rs}(t)$ is the actual travel time experienced over route $p$ by vehicles departing from origin $r$ toward destination $s$ at time $t$. Once the actual link travel time $\tau_a(t)$ is determined, the actual route travel time $\eta_p^{rs}(t)$ can be computed using the recursive formula discussed in Chapter 4.

Since network traffic conditions change over time, the actual route time may be significantly different from the instantaneous route travel time, especially when the route or travel time is long. Otherwise, the instantaneous route travel time provides a good estimate of the actual route travel time. The instantaneous route travel time is easily obtained or estimated compared to the actual route travel time since the prevailing traffic flow data can be obtained in real-time from a probe vehicle or a roadway detector.

The difference between the instantaneous link travel time and the actual link travel time may be insignificant since the length of links is generally short (0–1 miles), as is the travel time (several seconds to a few minutes). In conclusion, we would like to develop a temporal link travel time function which is a good representation of both the instantaneous link travel time and the actual link travel time. In the following, the link travel time function refers to the *actual* link travel time function. However, we note that when a link is extremely congested or oversaturated, these two kinds of link travel times may be quite different.

For arterials, vehicle delay at the exit from the link comprises both *deterministic* and *stochastic* components. As the analysis time interval shortens, the stochastic delay becomes less significant. The stochastic delay also depends on the exit capacity of a link. On this basis, the application of queuing and delay models may be divided into two categories. In the first category, flow and capacity information is required to predict queues and delays defined on a longer time-scale (e.g. successive 5 to 30 minute intervals). It is only possible to predict overall quantities such as the average queue length, or the average delay per vehicle evaluated over a complete traffic peak. In this case, the stochastic delay constitutes a significant part of the total delay and cannot be neglected in any travel time estimation. The travel time function in this category is suitable for off-line evaluation of ATMS and ATIS systems.

In the second category, the flow and capacity information is required to estimate queues and delays defined on a short time-scale (e.g. successive 1 to 5 minute intervals). It is possible to determine in detail the time variation of the average queue length and vehicular delay. The stochastic delay is therefore negligible, and only deterministic oversaturation delay must be considered in this case. Travel time functions in this category are suitable for real-time on-line evaluation of ATMS and ATIS systems. The functions proposed in the chapter might be calibrated for vehicle-actuated and fixed-time signals separately. We note that turning flows at an intersection are not considered in this chapter.

For freeway segments, we consider the flow and capacity variation which is required to estimate queues and delays defined only on a short time-scale (e.g., less than 5 minutes). Thus, it is possible to determine in detail the time variation of the average density and vehicular delay. The stochastic delay is therefore negligible, and only deterministic delay must be considered in this case. These travel time functions for freeway segments are suitable for real-time on-line evaluation of ATMS and ATIS systems.

## 16.2 Functions for Arterials: Longer-Time Horizons

In this section, we consider the dynamic link travel time functions for an analysis interval of 5–30 minutes or longer. We seek to apply delay formulae in the literature to derive corresponding temporal link travel time functions. In the delay formulae, the input is the average flow rate arriving at the downstream intersection. In our dynamic route choice model in Chapter 5, we use three variables ($x_a(t)$, $u_a(t)$, $v_a(t)$) to represent the dynamics of traffic on a link; however, the arrival flow rate at the downstream intersection of the link does not correspond to any of these three variables. To overcome this difficulty, we divide a physical link $a = (A, B)$ into two dummy links: link $a_1 = (A, D)$ and $a_2 = (D, B)$ (see Figure 16.1). The location of dummy node $D$ is undetermined and is movable. We assume that link $a_1 = (A, D)$ contains an uncongested vehicle stream and $a_2 = (D, B)$ contains a traffic queue. Thus, the length of $a_2$ is the length of the physical queue on link $a$. When there is no queue on link $a$, the length of dummy link $a_2$ equals zero and dummy link $a_1$ has the same length as link $a$.

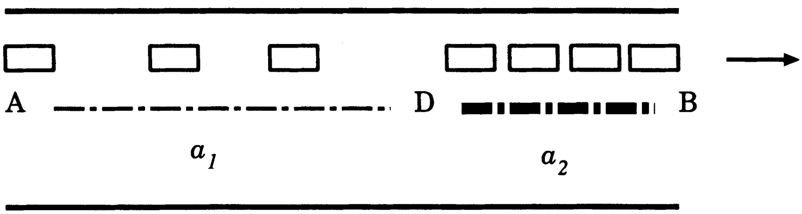



Figure 16.1: Traffic on an Arterial Road Link

The state equations for the two dummy links are:

$$\frac{dx_{a1}(t)}{dt} = u_{a1}(t) - v_{a1}(t) \qquad \forall a_1 \in a; \tag{16.3}$$

$$\frac{dx_{a2}(t)}{dt} = u_{a2}(t) - v_{a2}(t) \qquad \forall a_2 \in a. \tag{16.4}$$

Flow conservation for the two dummy links requires:

$$v_{a1}(t) = u_{a2}(t) \qquad \forall a_1, a_2 \in a. \tag{16.5}$$

Now, we use six variables, $x_{a1}(t)$,$u_{a1}(t)$,$v_{a1}(t)$,$x_{a2}(t)$,$u_{a2}(t)$ and $v_{a2}(t)$, to describe the dynamic traffic on each link $a$. Since three equations (16.3)–(16.5) are associated with each physical link $a$, only three variables are independent.

Next we consider a discrete time problem. Denote $\Delta k = [k, k+1]$ as the length of the time interval in hours. In this discrete time formulation, $x_a(k)$ represents vehicles on the link at the beginning of interval $k$; $u_a(k)$ and $v_a(k)$ represent inflow and exit flow during interval $k$. Writing the above equations in a discrete-time form, it follows that

$$x_{a1}(k+1) = x_{a1}(k) + u_{a1}(k)\Delta k - v_{a1}(k)\Delta k \qquad \forall a_1 \in a, k; \tag{16.6}$$

$$x_{a2}(k+1) = x_{a2}(k) + u_{a2}(k)\Delta k - v_{a2}(k)\Delta k \qquad \forall a_2 \in a, k; \tag{16.7}$$

$$v_{a1}(k) = u_{a2}(k) \qquad \forall a_1, a_2 \in a, k. \tag{16.8}$$

Thus, in general the average link travel time $\tau_a(k)$ per vehicle during time interval $k$ can be expressed as the sum of two components: 1) a flow-dependent cruise time $D_{a1}(k)$ over the first part of the link; and 2) a queuing delay $D_{a2}(k)$. It follows that

$$\tau_a(k) = D_{a1}(k) + D_{a2}(k) \qquad \forall a \tag{16.9}$$

Next we consider the cruise time $D_{a1}(k)$ (in seconds) over the first part of link $a$. Using Greenshields formula (Greenshields, 1933), the average cruise speed $w_{a1}(k)$ (miles/hour) for inflow entering link $a$ during time interval $k$ is

$$w_{a1}(k) = w_{ao}\,[1 - \frac{e_{a1}(k)}{e_{am}}] \tag{16.10}$$

where $w_{ao}$ is the free flow speed and $e_{am}$ (vehicles/mile) is the maximal density (jam density) of traffic on link $a$. Thus, the traffic density $e_{a1}(k)$ for inflow entering link $a$ during time interval $k$ can be expressed as

$$e_{a1}(k) = e_{am}\ [1 - \frac{w_{a1}(k)}{w_{ao}}] \tag{16.11}$$

Therefore, inflow $u_{a1}(k)$ can be expressed as

$$u_{a1}(k) = e_{a1}(k)\ w_{a1}(k) = e_{am}\ w_{a1}(k) - \frac{e_{am}}{w_{ao}}\ [w_{a1}(k)]^2 \tag{16.12}$$

We then derive the cruise speed $w_{a1}(k)$ for inflow entering link $a$ during time interval $k$ as

$$w_{a1}(k) = \frac{w_{ao}}{2}\left\{1 + \sqrt{1 - \frac{4u_{a1}(k)}{e_{am}w_{ao}}}\right\} \tag{16.13}$$

In the above derivation, we assume that the cruise traffic is in an uncongested state. Denote $l_a$ as the length of link $a$ (miles). The length of the vehicle queue at the beginning of time interval $k$ is $x_{a2}(k)/e_{am}$. Thus, the actual length of the cruise from inflow until reaching the queue is $[l_a - x_{a2}(k)/e_{am}]$. Therefore, the actual cruise time (in seconds) over link $a$ for inflow entering link $a$ during time interval $k$ is

$$D_{a1}(k) = 3600\frac{l_a - x_{a2}(k)/e_{am}}{w_{a1}(k)} \tag{16.14}$$

However, we note that for an arterial link, the dependency of cruise speed on inflow is not very significant (McShane and Roess, 1990). The cruise speed on an arterial link is mainly associated with the class of the arterial and the geometry of the link.

Before we analyze the intersection delay, the concept of link capacity needs to be clarified. In a temporal traffic network, there are different maximal discharge rates of traffic on a link. Generally, there are three roadway sections which have individual capacity constraints. The three link elements are: entry, midblock road section and exit of a link. Therefore, when we use the word *capacity* in a temporal traffic network, it is necessary to indicate whether it is an inflow capacity, a midblock flow capacity or an exit flow capacity. Since capacities per lane at entry and midblock are usually higher than at the exit point, we will refer only to the exit capacity when considering the flow capacity of a link.

It is assumed that the capacity at the exit of a link is a function of time, $\mu_a = \mu_a(k)$, for each link $a$ (in vehicles/hour). In practice, it is necessary to evaluate the exit capacity for each time interval given the exit flow $v_{a2}(k)$ in the relevant interacting link flows. Capacity calculation methods also depend on the type of junction; the three main types are traffic signals, major/minor priority and roundabouts. We only discuss signal capacity in this chapter.

Mathematical models used to estimate intersection delay are queuing models. Since we are considering a rather long time interval, the delay at the exit of the link involves deterministic and stochastic delay. Steady state queuing theory is widely used but predicts infinite queues and delays when the demand reaches the capacity available to it. However, when demand is close to capacity, or when the capacity is exceeded for short periods, the queue growth lags behind the expectations of steady state theory, and the rate of variation of demand and capacity cannot be ignored. Deterministic queuing theory, on the other hand, in which the delay is obtained as a simple integral of demand minus capacity, can sometimes be used when demand and capacity vary in time. However, this treatment ignores the random nature of traffic arrivals and departures within a rather long time interval, and leads to serious underestimates in the delay unless the capacity is exceeded by a considerable margin. When demand just reaches capacity, zero delays are predicted by the deterministic model.

Thus, the most important region for delay estimation is where demand (inflow) and capacity are approximately equal; this is the region which is inadequately represented both by the steady state and deterministic approaches. Methods are needed which adequately treat the entire range of demand and capacity, and take proper account of the random nature of traffic and of the variations in time of demand and capacity.

Here, we deal solely with queues and the corresponding approach delays. Another type of delay is geometric delay at a yield sign; such delay is suffered in the absence of queues because of the need for vehicles to slow down, negotiate the intersection, and accelerate back to normal speed. Such delays are not treated here, but must be considered in any practical implementation.

The average delay per vehicle, $D_{a2}(k)$, for vehicles arriving at the downstream intersection of link $a$ during time interval $k$ can be expressed as the sum of two delay terms:

$$D_{a2}(k) = d_{a1}(k) + d_{a2}(k) \qquad \forall a \tag{16.15}$$

where $d_{a1}(k)$ is the non-random delay (delay due to signal cycle effects calculated assuming non-random arrivals at the average inflow rate), and $d_{a2}(k)$ is the overflow delay including effects of random arrivals as well as any oversaturation delays experienced by vehicles arriving during the specified flow period.

Denote $\rho_a(k)$ as the degree of saturation at the exit from link $a$ during time interval $k$. It follows that

$$\rho_a(k) = \frac{u_{a2}(k)}{\mu_a(k)} \tag{16.16}$$

Non-random delay at the intersection is estimated by assuming that the number of vehicles which arrive during each signal cycle is fixed and equivalent to the average flow (demand) rate per cycle. Different expressions are used for the non-random delay term according to the arrival characteristics (uniform or platooned) and the signal characteristics (one or two green periods). The uniform delay formula which is valid for the case of a single green period with arrivals at a constant rate throughout the signal cycle is the first term of Webster's formula (1958):

$$d_{a1}(k) = \frac{0.5c\,[1 - g(k)/c]^2}{1 - \rho_a(k)\, g(k)/c} \tag{16.17}$$

where $c$ is the signal cycle time in seconds, and $g(k)$ is the effective green time in seconds during time interval $k$. To include the effects of traffic progression on delays at traffic signals, Fambro et al (1991) and Messer (1990) suggested converting the above uniform delay formula into a non-uniform arrival term to account for the progression effects as follows:

$$\begin{aligned} d_{a1}(k) &= \frac{0.5c\,[1 - g(k)/c]^2}{1 - \rho_a(k)\, g(k)/c} \frac{1 - P(k)}{1 - g(k)/c} \\ &= \frac{0.5c\,[1 - g(k)/c][1 - P(k)]}{1 - \rho_a(k)\, g(k)/c} \end{aligned} \tag{16.18}$$

where $P(k)$ is the proportion of traffic arriving in the green phase in time interval $k$.

Recently, a more general delay formula for this uniform delay term was proposed by Akcelik and Rouphail (1991):

$$\begin{aligned} d_{a1}(k) &= 0.5[c - g(k)] && \text{for } \rho_a(k) > 1.0 \\ &= \frac{0.5c\,[1 - g(k)/c]^2}{1 - \rho_a(k)\, g(k)/c} && \text{for } \rho_a(k) \leq 1.0 \end{aligned} \tag{16.19}$$

Note that the above formulas are not smooth (no continuous first-order derivatives). If we want to combine those functions into our framework, we need to smooth those functions, since a smooth function is necessary for the solution of dynamic route choice problems.

Overflow delay estimation has attracted extensive research. Its development was initially reported by Kimber and Hollis (1979). Later on, Hurdle (1984) further discussed the assumptions and limitations of those delay models. Several countries proposed time-dependent delay formulas in their capacity guides, including the U.S. (TRB, 1985), Canada (Teply, 1984) and Australia (Akcelik, 1981).

The 1985 Highway Capacity Manual (HCM) suggested a delay formula to account for both short-term (random or Poisson) and long-term overflows of queues to subsequent cycles due to continuous oversaturation. However, it has been widely criticized in recent years. Akcelik (1988) noted that the HCM equation predicted higher delays for oversaturated conditions than did the Australian and Canadian formulas. He recommended a general formula for the overflow delay as follows:

$$\begin{aligned} d_{a2}(k) &= 900\Delta k\, [\rho_a(k)]^n \cdot \\ &\left\{ [\rho_a(k) - 1] + \sqrt{[\rho_a(k) - 1]^2 + \frac{m[\rho_a(k) - \rho_{ao}(k)]}{\mu_a(k)\, \Delta k}} \right\} \end{aligned} \tag{16.20}$$

where $\rho_{ao}(k)$ is the degree of saturation below which the overflow delay $d_{a2}(k)$ is negligible. This can be expressed as

$$\rho_{ao}(k) = a + b\, s(k)\, g(k)$$

where $s(k)$ is the saturation flow rate in vehicles per second during time interval $k$, $(s(k)\, g(k))$ is the capacity per cycle during time interval $k$, and $a, b, m$ and $n$ are calibration parameters.

However, Akcelik's formula is not smooth at the point $\rho_{ao}(k)$. One alternative is to drop this term in the current development of link time functions for dynamic modeling purposes. Burrow (1989) proposed a generalized version of Akcelik's model for overflow delay. This model is:

$$\begin{aligned} d_{a2}(k) \quad &= 900\Delta k\, [\rho_a(k)]^n \cdot \\ &\left\{ [\rho_a(k) - 1] + \alpha + \sqrt{[\rho_a(k) - 1]^2 + \frac{m[\rho_a(k) + \beta]}{\mu_a(k)\, \Delta k}} \right\} \end{aligned} \tag{16.21}$$

where $\alpha$ is an additional term used to encompass the more general form above and $\beta$ is a term related to $\rho_a(k)$ in Akcelik's model. The above formula can be considered as an alternative to our suggestion since it is a smooth function.

In the above models, it is assumed that the initial queue $x_{a2}(k)$ is zero when the overflow period $\Delta k$ begins. However, this initial queue should be counted in our dynamic delay model and the queuing delay (in seconds) caused by initial queue $x_{a2}(k)$ can be expressed as (Akcelik and Rouphail, 1991):

$$3600 \frac{x_{a2}(k)}{\mu_a(k)} \tag{16.22}$$

The above formula is suitable for the case when the approaching flow $u_{a2}(k)$ is greater than or equal to exiting flow $v_{a2}(k)$. For other cases, it may overestimate the delays caused by the initial queue. Those cases need further study in the future. In the meantime, we suggest the following overflow delay equation for the dynamic problem:

$$\begin{aligned} d_{a2}(k) \quad = \quad & 3600 \frac{x_{a2}(k)}{\mu_a(k)} + 900\Delta k\, [\rho_a(k)]^n \cdot \\ & \left\{ [\rho_a(k) - 1] + \sqrt{[\rho_a(k) - 1]^2 + \frac{m\rho_a(k)}{\mu_a(k)\, \Delta k}} \right\} \end{aligned} \tag{16.23}$$

In summary, we propose the following dynamic link travel time function for the case of a long time interval ($\Delta k \geq 5$ minutes):

$$\tau_a(k) = D_{a1}(k) + d_{a1}(k) + d_{a2}(k) \qquad \forall a \tag{16.24}$$

where $D_{a1}(k)$ is given by (16.14), $d_{a1}(k)$ is given by (16.17) and $d_{a2}(k)$ is given by (16.23).

## 16.3 Functions for Arterials: Short-Time Horizons

Now we consider a rather short time period of 1–5 minutes in length. It is reasonable to assume that the impact of randomness of traffic arrivals at the traffic signal is negligible during such a short time. Thus, the stochastic delay is neglected and only deterministic uniform and oversaturation delays are considered. There have been some recent studies in this area such as the model presented by Takaba (1991). Furthermore, if we want to consider a time interval of less than one minute, it is necessary to know the offset for each intersection signal and determine the non-random delays.

For a short time interval problem, the delay at the exit from the link is considered deterministic in contrast with that occurring in a long time interval problem. Nevertheless, the cruise time formula (16.14) is still applicable since it is based on the average speed of the inflow. Thus, the average link travel time per vehicle during time interval $k$ is still expressed as

$$\tau_a(k) = D_{a1}(k) + D_{a2}(k) \qquad \forall a.$$

The main difference between a longer time period and a shorter time period is the second term $D_{a2}(k)$. As the time interval becomes shorter, the stochastic delay decreases. Thus, for simplicity, delay term $D_{a2}(k)$ can be developed in a deterministic manner for a short time period. The average delay per vehicle, $D_{a2}(k)$, for vehicles arriving at the exit from link $a$ during time interval $[k, k+1]$ can also be expressed as the sum of two delay terms:

$$D_{a2}(k) = d_{a1}(k) + d_{a2}(k) \qquad \forall a$$

where $d_{a1}(k)$ is the cyclic delay (delay due to signal cycle effects calculated assuming non-random arrivals at the average inflow rate in each cycle), and $d_{a2}(k)$ is the delay due to oversaturation experienced by vehicles arriving during the specified flow period. The first delay term is given by Webster's formula. For formula (16.19), smoothing is necessary for our purposes.

We next discuss the second delay term $d_{a2}(k)$. The decision variable is the queue length on physical link $a$. As discussed before, link $a = (A, B)$ is decomposed into two dummy links, $a_1 = (A, D)$ and $a_2 = (D, B)$. After the cruise time $D_{a1}(k)$, vehicles entering link $a$ during time interval $[k, k+1]$ should either reach the queue on link $(D, B)$ or proceed to downstream links. As before, the location of dummy node $D$ is movable. Thus, the length of link $a_2 = (D, B)$ is the length of queuing flow.

The deterministic (initial) queue encountered by the first vehicle arriving at the beginning of time interval $[k, k+1]$ is $x_{a2}(k)$. The deterministic queue encountered by the last vehicle arriving at the end of time interval $[k, k+1]$ is $\{[u_{a2}(k) - v_{a2}(k)]\Delta k + x_{a2}(k)\}$. The average queue for vehicles arriving during time interval $[k, k+1]$ is $\{[u_{a2}(k) - v_{a2}(k)]\Delta k/2 + x_{a2}(k)\}$. Thus, the average deterministic queuing discharge time or queuing delay $d_{a2}(k)$ per vehicle arriving during time interval $[k, k+1]$ can be expressed as

$$\begin{aligned} d_{a2}(k) &= 3600\frac{[u_{a2}(k) - v_{a2}(k)]\Delta k/2 + x_{a2}(k)}{\mu_a(k)} \\ &= 1800\frac{[u_{a2}(k) - v_{a2}(k)]\Delta k + 2x_{a2}(k)}{\mu_a(k)} \end{aligned} \tag{16.25}$$

where $d_{a2}(k)$ is in seconds.

In summary, we propose the following dynamic link travel time function for a short time interval:

$$\tau_a(k) = D_{a1}(k) + d_{a1}(k) + d_{a2}(k) \qquad \forall a \tag{16.26}$$

where $D_{a1}(k)$ is given by (16.14), $d_{a1}(k)$ is given by (16.17) and $d_{a2}(k)$ is given by (16.25). Sometimes, we use the link travel time function for vehicles entering link $a$ at the beginning of interval $[k, k+1]$. We note that most numerical examples in this book use this kind of link travel time function for computations. In this situation, the queuing delay $d_{a2}(k)$ (seconds) per vehicle arriving at the beginning of interval $[k, k+1]$ should be revised as

$$d_{a2}(k) = 3600\frac{x_{a2}(k)}{\mu_a(k)} \tag{16.27}$$

Then, queuing delay $d_{a2}(k)$ in dynamic link travel time function (16.26) should be replaced by equation (16.27).

## 16.4 Implications of Functions for Arterial Networks

### 16.4.1 Number of Link Flow Variables

The selection of suitable link travel time functions for discrete-time dynamic route choice largely depends on the length of the analysis time interval. Travel times for longer time intervals must account for stochastic delays at intersections. The intersection delays in travel time functions for shorter time intervals are predominantly deterministic.

In addition to the models presented in this text, various dynamic route choice models have been proposed by many researchers. These models use some variations of time-dependent link travel time functions. A number of models still use the static BPR volume-delay function for a time-dependent traffic network problem. However, the BPR function is based on an implicit assumption of steady state traffic flow. This assumption is invalid in a time-dependent and stochastic traffic network. Furthermore, in using time-dependent delay formulae, traffic engineering practice implies that the BPR function cannot predict intersection delays properly (HCM, 1985). Therefore, using the BPR function as the basis for dynamic route choice models would generate results that are too approximate to be realistic in a real-time environment.

Some dynamic route choice models use a single flow variable (i.e. the number of vehicles on a link or average flow rate over a link for each time interval) to describe the dynamics of traffic flow on a link. From the above analysis, it is evident that one flow variable for a link is insufficient to capture the dynamic characteristics of traffic flow on a link and cannot be used to estimate the time-dependent delays properly at the intersection. Therefore, the six link flow variables and link state equations suggested in this chapter are proposed as the basis for dynamic route choice models. Associated with these proposed dynamic link time functions, we need to modify our dynamic network models on arterial networks as well.

### 16.4.2 Notes on Functions for Arterial Links

In the derivation of dynamic link travel time functions, it is basically assumed that delays are caused by the signal control at the downstream intersection assuming isolated control. Since these travel time functions are developed to apply in dynamic network models, the interaction of upstream and downstream intersections should be taken into account.

There is a critical queue length requirement for each link (Rouphail and Akcelik, 1991). The critical queue length, $N_a$, is defined as the longest downstream queue that allows upstream platoons to accelerate to and discharge at the full saturation flow rate. It follows that

$$x_{2a}(k) \leq N_a \qquad \forall a. \tag{16.28}$$

The critical queue length is assumed fixed for a given set of platoon speeds and speed change rates of the cruise inflow. This constraint reflects the reduction of inflow capacity as a result of downstream queue interaction effects, which in turn has an impact on delays at downstream intersection. However, the functional relationship between the critical queue length and the cruise speed may complicate the formulation of dynamic network models. This difficulty needs further investigation. In the following, we mainly discuss the factors affecting the stochastic delay term in the link travel time function for longer time horizons.

In the case of longer time horizons, the stochastic delay term is derived by assuming that the downstream intersection is an isolated intersection subject to Poisson arrivals, independent of the stochastic delay at other intersections. Newell (1990) noted that under certain conditions, this assumption may grossly overestimate the stochastic delay on an arterial. This factor is especially important in dynamic network models since intersections in a network can no longer be considered isolated. Newell (1990) further pointed out that the cumulative stochastic delay on links over an arterial is mainly dependent on the stochastic delay at the critical intersection.

Also, van As (1991) noted that the randomness of arrivals in a traffic network is significantly reduced when the arrivals at upstream intersections are highly congested. Thus, delay is significantly overestimated in networks if the proposed formulae do not allow for the narrower distribution of arrivals, especially if the network is near saturation. The adjustment of dynamic link travel time functions to account for oversaturation effects in a longer time horizon case is another important issue in the application of dynamic network models.

## 16.5 Functions for Freeway Segments

Dynamic link travel time functions for freeway segments are simpler than those for arterials. In general, the average link travel time per vehicle $\tau_a(k)$ during time interval $k$ can be expressed as the sum of two components: 1) a free-flow cruise time $D_{a1}$ over the link; and 2) a flow-dependent congestion delay $D_{a2}(k)$. It follows that

$$\tau_a(k) = D_{a1} + D_{a2}(k) \qquad \forall a \tag{16.29}$$

The free-flow cruise time (seconds) on link $a$ is

$$D_{a1} = 3600 \frac{l_a}{w_{a0}} \qquad \forall a \tag{16.30}$$

where $l_a$ is the link length (miles) and $w_{a0}$ is the free-flow speed (miles/hour) on link $a$. The second delay term is caused by the traffic ahead of a vehicle when it enters link $a$. For simplicity, we assume that the average traffic density on link $a$ determines this congestion delay. We note that a non-uniform distribution of traffic over link $a$ may bring an error to this delay formula. However, as link $a$ becomes shorter, the traffic is distributed more uniformly and this formula becomes more accurate.

The deterministic queue encountered by the first vehicle arriving at the beginning of time interval $[k, k+1]$ is $x_a(k)$. The deterministic queue encountered by the last vehicle arriving at the end of time interval $[k, k+1]$ is $\{[u_a(k) - v_a(k)]\Delta k + x_a(k)\}$. The average queue encountered by vehicles arriving during time interval $[k, k+1]$ is $\{[u_a(k) - v_a(k)]\Delta k/2 + x_a(k)\}$. Thus, the average traffic density (vehicles/mile) encountered by vehicles arriving during time interval $[k, k+1]$ is

$$e_a(k) = \frac{\{[u_a(k) - v_a(k)]\Delta k/2 + x_a(k)\}}{l_a} \tag{16.31}$$

We assume that the second delay term $D_{a2}(k)$ (seconds) can be expressed as

$$D_{a2}(k) = \alpha \, [e_a(k)]^m \tag{16.32}$$

where $\alpha$ is a real-valued parameter and $m$ is an integer parameter. Both of them need to be calibrated.

In summary, we propose the following dynamic link travel time function for freeway segments.

$$\tau_a(k) = D_{a1} + D_{a2}(k) \qquad \forall a \tag{16.33}$$

where $D_{a1}$ is given by equation (16.30) and $D_{a2}(k)$ by equation (16.32). This formula is different from the traditional BPR function because the second delay term depends on the average traffic density encountered by vehicles arriving during time interval $[k, k+1]$ instead of average traffic flow on the link. Sometimes, we use the link travel time function for vehicles entering link $a$ at the beginning of interval $[k, k+1]$. In this situation, the average traffic density (vehicles/mile) encountered by vehicles arriving at the beginning of time interval $[k, k+1]$ is

$$e_a(k) = \frac{x_a(k)}{l_a} \tag{16.34}$$

Then, queuing delay $D_{a2}(k)$ in dynamic link travel time function (16.33) should be computed using equations (16.32) and (16.34) instead of equations (16.32) and (16.31).

## 16.6 Notes

This chapter has investigated different aspects of time-dependent link travel time functions for signal-controlled arterial and freeway links. For an arterial link, the following conclusions are emphasized.

1. Two sets of dynamic link travel time functions are proposed depending on the analysis time horizon. These functions can be applied to discrete-time dynamic network models.

2. For each link, six variables (three of which are independent) are necessary to describe the dynamics of traffic flow and calculate temporal link travel time on an arterial link with a signal controlled intersection.

3. Each physical link is decomposed into two dummy links in order to identify the queue length on the physical link. Thus, the link travel time functions for each physical link depend on the flow variables of two dummy links so that link interaction enters the dynamic link travel time functions. This interaction should be considered in dynamic network formulations.

There are many critical assumptions underlying the delay equations for arterial links. Those assumptions are especially important for the longer time interval delay model. The generalization of those assumptions will make the travel time functions more realistic. Among those assumptions, the most critical factors are the impact of signal coordination and the case when the flow is not zero after the peak ends. For the short time interval problem, we also need to investigate more realistic inflow arrival patterns at the exit of a link.

Validation of the proposed dynamic link travel time functions is a major task for future research.

A dynamic link travel time function for a freeway segment link is also proposed. The major delay in the travel time function is dependent on the average traffic density on the freeway link. However, this function needs to be calibrated and validated.

The proposed link travel time functions for arterial and freeway segments are still subject to future challenges and validations from many sources. The first challenge is from traffic simulation models. Microscopic simulation models might provide detailed answers to many remaining questions for dynamic link travel time functions and also serve as validation tools.

The second challenge is from hydrodynamics theory (Newell, 1993) and its discrete form, the highway cell transmission model (Daganzo, 1993). The second challenge is more theory-oriented. Hydrodynamics theory and its discrete form may provide a basis for the derivation of dynamic link travel time functions, especially for freeway segments. We believe that with all these efforts plus a large amount of realistic data generated from many IVHS operational tests, we can produce appropriate link travel time functions for both freeway and arterial links. Consequently, those functions will ensure a good representation of travel times and traffic propagation on arterial and freeway links.

# Chapter 17

# Implementation in IVHS

The rapid evolution of IVHS technologies presents more and more specific requirements for dynamic network modeling. Conversely, implementation of dynamic models is becoming more and more important for the design and evaluation of IVHS. In Section 17.1, several applications of dynamic models to IVHS components are discussed. We mainly investigate the technical aspects of applying these models. Subsequently, we discuss various data requirements for implementing these dynamic models in Section 17.2.

## 17.1 Implementation Issues

Dynamic transportation network models describe the basic operating functions as well as providing evaluation tools for IVHS. To simplify our discussion on the application of dynamic models, we focus on the following items:

1. traffic prediction;
2. traffic control;
3. incident management;
4. congestion pricing;
5. operations and control for automated highway systems (AHS);
6. transportation planning.

In the following, we investigate various issues for dynamic network models in serving each of the above applications.

### 17.1.1 Traffic Prediction

Dynamic transportation network models function as predictive models for many ATMIS systems. Travelers' choice behavior determines which dynamic route choice model best fits in the predictive module of a realistic ATMIS. In general, no single-group dynamic route choice model can represent the travel choices of the entire population. The most plausible model is a reasonable combination of several dynamic route choice models, including both deterministic and stochastic models. As discussed in Chapter 12, travelers can be stratified into different groups based on the following route diversion behavior: 1) prespecified routes; 2) a few alternative routes; 3) many alternative routes. The population and characteristics of each group can be determined with surveys and updated periodically. For travelers with prespecified routes, the route must be first generated exogenously. In general, this route includes a freeway segment and some surface streets. For networks with fewer alternative routes, such as the San Francisco Bay Area, the population with prespecified routes may be large. On the other hand, for networks with many alternative routes, such as the Chicago Area, the population with prespecified routes is relatively small. Similar arguments apply to travelers with few and many alternative routes.

For recurrent and non-recurrent congestion, both instantaneous and ideal DUO route choice criteria may apply. Note that in this application, the DUO state is defined using a general definition of travel disutility including fuel consumption, auto operating cost, etc., in addition to travel time. The population of travelers using either the instantaneous or ideal DUO route choice criterion could be determined by survey. Senior citizens or cautious travelers may prefer to choose routes based on traffic information from their past experience. In other words, they may choose routes using the ideal DUO criterion. On the other hand, young or aggressive travelers may prefer to choose routes based on current traffic information. That is, they may choose routes using the instantaneous DUO criterion.

For a traffic network, the multimodal problem should be handled explicitly. Conventional modes include HOV, bus, truck and passenger car. HOV lanes should be designated as separate links and may be subject to possible pricing or toll charges during congested periods. These will alter the travel cost and subsequently change the flow pattern of HOV lanes. Bus constitutes a special mode which should be handled carefully in the modeling. In general, buses move slowly and cause additional delays to other vehicles. For a road link with bus traffic, the travel time function needs to be adjusted for through traffic in the right lane and right turning traffic, since the impact of bus traffic is significant. For the left turn lane, buses only affect traffic flow at the time when making a left turn. Truck traffic flows also need to be transformed into equivalent passenger car flows. However, possible revisions of the travel time functions may be necessary, especially for turning movements because trucks make wide turns and take longer times for turning.

Since bus routes are fixed, there is no route choice for bus traffic itself. Thus, it belongs to the group with prespecified routes. Truck traffic can be classified with the group with fewer alternative routes because some roads are closed to truck traffic. The trip chaining problem is not explicitly considered in this framework. However, a similar modeling framework is applicable to the situation of trip chaining when time-dependent trip information is available.

Concerning the route choice criterion, a more plausible model for dynamic traffic prediction may be a combination of instantaneous/ideal DUO/SDUO route choice models. For travelers without route guidance devices or not complying with guidance information, instantaneous/ideal SDUO route choice models can be used to model this group of travelers. The population with prespecified routes may be quite large so that the dispersion parameter for this group is small. Table 17.1 presents route choice criteria for these two groups of travelers. Table 17.2 summarizes the stratification of travelers and their corresponding travel choice criteria. A general VI model for the above multi-group, multi-criteria route choice problem can be formulated by generalizing the VI models of Chapters 12 and 13.

Table 17.1: Route Choice Criteria for Different Travelers

| Travelers | Route Choice Criteria | |
|---|---|---|
| Guided Travelers | Instantaneous DUO | Ideal DUO |
| Travelers with No Guidance | Instantaneous SDUO | Ideal SDUO |

Table 17.2: Groups of Travelers and Travel Choice Criteria

| Modes | | Passenger Car | Trucks | HOV | Bus |
|---|---|---|---|---|---|
| Groups by Diversion | No | Y | Y | Y | Y |
| | Few | Y | Y | Y | N |
| | Many | Y | Y | N | N |
| Travel Choice Criterion | Instantaneous DUO | Y | Y | Y | N |
| | Ideal DUO | Y | Y | Y | N |
| | Instantaneous SDUO | Y | Y | Y | N |
| | Ideal SDUO | Y | Y | Y | N |

Y – Yes; N – No.

Furthermore, dynamic traffic prediction problems can be classified into several types based on time periods and travel purposes. Basically, we have:

1. morning home-to-work period (6:30 am - 9:30 am);
2. midday non-commuting period (9:30 am - 3:30 pm);
3. work-to-home period (3:30 pm - 7:30 pm);
4. evening period (7:30 pm - 10:30 pm);

5. night period (10:30 pm - 6:30 am).

The times indicated are illustrative and need to be determined by survey. For dynamic traffic prediction during the morning home-to-work period, the major concern is the travel time and the arrival time. The joint mode/departure time/route choice program might be revised as a joint mode/arrival time/route choice program. The arrival times for commuting travelers need to be specified and a large penalty charged for late arrival. Among the four modes specified in Table 17.2, the priority of bus and HOV should be guaranteed in terms of travel time reliability. For dynamic traffic prediction during other periods, the models should be revised accordingly to reflect characteristics of the period of interest.

### 17.1.2 Traffic Control

Dynamic network models can be extended to incorporate traffic control measures, such as signal control and congestion pricing. Congestion pricing is discussed in Section 17.1.4. We now investigate how the dynamic models are used in dynamic traffic control and coordination problems. In general, we can construct a bilevel program for the combined dynamic traffic prediction and control problem (Figure 17.1). In the upper level, we have a dynamic traffic control model, which can be formulated using various objective functions as follows:

1. minimize total travel time;
2. minimize total travel cost or disutility;
3. minimize total number of vehicles during time period $[0, T]$;
4. minimize average congestion level during time period $[0, T]$;
5. minimize the length of the congested time period $[0, T]$;
6. minimize total emissions during time period $[0, T]$.

In the lower level, we have a dynamic traffic prediction model, which determines dynamic traffic flows in the network by considering the control strategies provided by the upper level model. The above bilevel program is solved so that an optimal control strategy can be found while the traffic flow follows the desired DUO/SDUO route choice criteria. This bilevel program can be either hierarchical or non-hierarchical, depending on the nature of the problem. The control strategies can also be classified as centralized or decentralized, which can generate different versions of the objective function for the upper level control model.

| | |
|---|---|
| Upper Level: Dynamic Traffic Controller: | Determine optimal traffic control strategies based on the predicted dynamic flows. |

where the dynamic traffic flows are given by:

| | |
|---|---|
| Lower Level: Dynamic Traffic Predictor: | Determine DUO/SDUO dynamic traffic flows using the provided dynamic traffic control strategy. |

Figure 17.1: A Bilevel Program for Dynamic Traffic Prediction and Control

### 17.1.3 Incident Management

Incident management is another area in which dynamic network models can play an important role. Traditional incident management strategy minimizes incident congestion by clearing incidents as quickly as possible and diverting traffic before vehicles are trapped in the incident queue. However, such simple incident management plans cannot solve most incident congestion problems. Instead, a systematic and comprehensive incident management strategy is needed to tackle the incident congestion problem, as offered by advanced ATMIS/APTS technologies. Such an incident management strategy could use travelers' information on origins, destinations and departure times to develop a coordinated strategy to advise each person regarding a best mode and route to their destination on a real time basis. In this way, we can achieve either user-optimal or system-optimal objectives by appropriately integrating available information and control measures.

In the context of such an incident management strategy, dynamic transportation network models are necessary for incident-related routing and travel time prediction. These models are both information and control oriented. The prediction procedure assumes either of two types of routing of vehicles: normal conditions and incident conditions. Routing for normal conditions provides time-dependent traffic flows and travel times under normal traffic conditions. Routing for incident conditions provides forecasts and suggestions following an incident. Both instantaneous and ideal DUO/SDUO route choice models can be used for this procedure. The system-optimal objective can also be applied to incident rerouting in conjunction with dynamic congestion pricing. Oversaturation and spillback can be serious problems when incidents happen. In the use of dynamic network models in response to incidents, corresponding constraints should be formulated to reflect these phenomena.

An incident can occur anywhere on a link. For simplicity, in the models we can specify that an incident is always located at the exit point of a link since a link with an incident can always be partitioned into two shorter links by placing a dummy node at the location of the incident. Thus, the location of the incident can be designated at the exit point of some shorter link. The models require that the start time and duration of an incident be reported by some source, such as the highway patrol. The duration of incident is defined as the period from the reported time through the clearance of the incident.

When incidents occur in a highway network, there are several ways for travelers to avoid congestion. Travelers already en route can shift to less congested routes which are less affected by the incident. If an incident happens in a central business district, some travelers may choose different parking places (destinations) and use other modes such as walking or buses to their final destinations. Travelers who have not yet departed may choose to delay their departures or shift to rapid transit. Thus, travelers' choices under incident conditions can be summarized as: route, departure time, mode and destination. Instead of discussing all these choices, we focus on route choice in the following.

A major difficulty of incident-related dynamic network models is that the continuity properties of most dynamic network models are destroyed. To overcome this difficulty, we consider a multi-period dynamic travel choice procedure in which each incident-related dynamic travel choice is treated separately and the continuity of traffic flow at the time and location of an incident instant is preserved. A detailed procedure is now described.

We consider two types of dynamic routing strategies under two route choice behavioral assumptions. The first type of route choice behavior is based on current or instantaneous travel time information. The second is based on projected or actual travel time information.

Instantaneous DUO/SDUO routing strategies pertain to drivers under incidents based on current traffic conditions. This framework is helpful in handling many unpredictable events that occur in traffic flow, such as accidents and illegal double parking on streets, etc. In case of such unforeseen events, optimal control models based on instantaneous travel times can provide improved results when feedback is taken into account and drivers adjust their routes enroute using updated traffic information.

Ideal DUO/SDUO routing strategies are appropriate for drivers' diversion under incidents based on projected travel time information. These strategies are useful when some future disturbance (incidents and other future events) are predictable, such as the increased traffic flow from a baseball stadium after a game. In general, if the future disturbance is more predictable, we can use the ideal DUO/SDUO route choice model. If the future disturbance is less predictable, we may prefer the instantaneous DUO/SDUO route choice model. Thus, the group stratification discussed in the dynamic traffic prediction module should be adjusted to the nature of the incident. An expert system can be designed to achieve this goal.

Under unpredictable incidents, our strategy is designed as a multi-period routing procedure. For simplicity, we first discuss the case of one incident only. We advise motorists of minimal travel cost routes based on current traffic information. Now, assume that an incident begins on link $b$ at time $t_1$ and will be cleared at time $T_1$. We divide the period $[0, T]$ into three periods: $[0, t_1]$, $[t_1, T_1]$ and $[T_1, T]$. At the transition points $t_1$ and $T_1$, there are sudden capacity reductions and increases, respectively, resulting in discontinuities of traffic flows on incident link $b$. In our routing strategy, we assume that motorists choose routes based on current traffic information. Thus, for period $[0, t_1]$ (before the incident occurs), motorists' route choices are based on current travel cost information generated using an instantaneous DUO/SDUO route choice model over the whole time span $[0, T]$, as if no incident would occur. It covers the entire time period $[0, T]$. During the initial period $[0, t_1]$, rerouting is provided by a normal instantaneous DUO/SDUO route choice procedure.

During incident period $[t_1, T_1]$, rerouting advice must be rearranged. We consider the impact of the incident on the reduction of capacity. We assume that the capacity reduction is uniform during the incident duration in order to simplify the problem. During incident period $[t_1, T_1]$, rerouting is provided from time $t_1$ through T, as if the incident would not be cleared until time $T_1$. This strategy is consistent with the route choice criterion of current traffic information, because the clearance time of incident is not known at the initial moment of the incident. This process is called *incident-based routing*. During incident conditions, we reroute vehicle flows which are already on the network and route O-D departures for period $[t_1, T]$ at the same time. Regarding the vehicles already on the network, we consider each link as a dummy origin from which flows may have different destinations. Thus, those vehicle flows on each link a at time instant $t_1$ can be considered as dummy O-D flows. A routing procedure similar to the initial routing can be implemented for the same network with a reduced capacity on the incident link for period $[t_1, T]$.

We now consider the last period $[T_1, T]$, which can be called the *recovery routing* period. During period $[T_1, T]$ when the incident is cleared, the capacity of the incident link is recovered. Routing is provided as normal instantaneous DUO/SDUO route choice. Still, the vehicle flows which are already on links at time instant $T_1$ must be considered as dummy O-D flows during the incident period. The above approach can be generalized to a general situation where there are multiple incidents occurring in the network during period $[0, T]$.

### 17.1.4 Congestion Pricing

Dynamic tolls can be collected with the application of automatic vehicle identification (AVI) technology. In conjunction with a dynamic route guidance system, effective traffic controls and congestion pricing can be implemented together to influence the routing strategies employed by either a central controller or individual drivers so that congestion levels in the transportation network can be controlled or adjusted.

It is anticipated that real-time congestion pricing can be evaluated using dynamic network models. Our focus is on technical aspects instead of policy issues. Traffic controls, including both surface street signal control and freeway ramp control, are assumed to be fixed in the current models. Their interaction with dynamic congestion pricing will be modeled as future extensions. In considering the impact of congestion pricing, therefore, we focus our attention on routing strategies instead of traffic signal controls. Various toll strategies and their modeling issues are discussed in the following.

**Bottleneck Tolls** A typical example of a bottleneck toll is a bridge toll. The relationship between the congestion level and the toll level at the bottleneck can be represented using a time-dependent function. Thus, dynamic tolls can be combined with dynamic link travel cost functions, and dynamic prediction models can be applied to this situation.

**Link and Route Tolls** As discussed in Chapter 15, a bilevel program can be constructed to determine dynamic link tolls. We use a leader-follower game to formulate such a dynamic congestion pricing problem. The objective of the upper-level problem (or the leader of the game) is to minimize the total travel time; the decision variable is the dynamic link toll. In the upper-level problem, we minimize the total travel *time* instead of total travel *cost*, because the link toll policy is designed to control the total congestion level. The DUO route choice problem (or the follower of the game) is formulated as the lower-level problem, in which the decision variables are the link flow variables. The upper level problem is a time-dependent minimization problem, whereas the lower-level problem is formulated as a variational inequality.

A bilevel program can also be constructed to determine dynamic *route* tolls. This program is similar to that for dynamic link tolls. However, solving this model requires explicit route enumeration.

**Area Congestion Tolls** An alternative congestion pricing strategy, area tolls or congestion zone tolls, may be easier to implement than the link or route toll policy. This dynamic congestion pricing strategy is designed to control the congestion level within the central business district (CBD) area, such as downtown Los Angeles. We charge a uniform toll which may be time-varying and adjustable depending on the congestion level within the CBD area. In our continuous time model, we assume that the toll is continuous in time. This toll applies to any vehicle traveling within the CBD area during peak hour periods.

As with link tolls, this problem can also be formulated as a bilevel model. The upper level is the congestion control or toll policy decision level and the lower level is the description of responsive traffic status. The control authority keeps track of the congestion level in terms of average saturation degree in the CBD area and charges a congestion toll for each vehicle present in CBD areas during the toll period. However, such a bilevel model is computationally difficult for a large network. Thus, we need to formulate a single level model which endogenously considers the interaction of dynamic congestion toll policy and travelers' choice responses. This formulation is similar to that for the bottleneck toll problem.

Dynamic toll models can function as tools in the evaluation of possible congestion pricing strategies in light of the advance of IVHS technologies. Eventually, they are expected to work together with traffic control schemes to combat traffic congestion in urban areas and become on-line operational tools in ATMIS systems.

### 17.1.5 Operations and Control for AHS

Dynamic transportation network models can be directly used as operations and control models in advanced IVHS systems. One example is an Automated Highway System (AHS). Some proposed AHS have only automated lanes. Thus, lane-changing itself is a matter of route choice which can be controlled by the central controller. A one-lane segment between two barriers can be considered as a link as shown in Figure 17.2. Thus, the dynamic prediction models can be adopted directly in the control of AHS. Furthermore, some proposed AHS have both automated lanes and non-automated lanes. The routing model for this case is more complicated than the previous one.

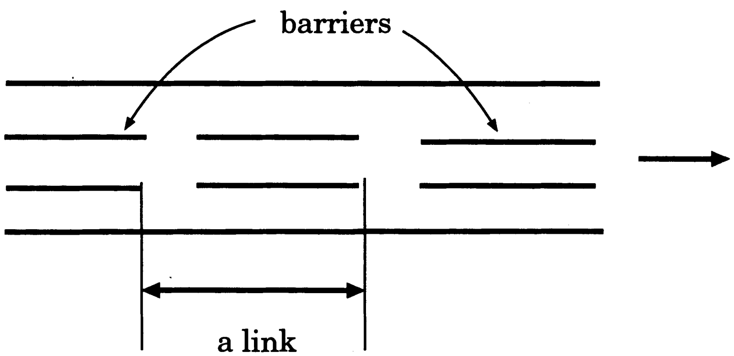



Figure 17.2: A Segregated Automated Highway System

### 17.1.6 Transportation Planning

Dynamic transportation network models can be generalized for application to long term planning purposes. Since dynamic models can predict traffic variations more accurately, they provide a good alternative to traditional planning models which are based on static equilibrium concepts. Furthermore, dynamic models can investigate disequilibrium aspects of traffic from day-to-day traffic variations.

With the rapid evolution of IVHS systems, the concepts and principles of IVHS are affecting the long-term transportation planning process. Thus, dynamic transportation network models can be extended as dynamic planning tools for a metropolitan area with various IVHS field projects. Combining dynamic location choices into dynamic planning models could become another area of interest in order to evaluate the long-term impact of IVHS projects on facility and residential locations.

## 17.2 Data Requirements

Inputs and parameter calibration for dynamic network models imply an extensive data collection effort. In general, most data required for dynamic models does not presently exist. Proposed and ongoing IVHS field tests provide an excellent opportunity to collect such data. Associated with our dynamic transportation network models, the following data requirements are identified:

1. time-dependent O-D matrices;

2. network geometry and intersection/ramp control data;

3. traffic flow data for calibrating various dynamic link travel time functions;

4. traveler information for stratifying travelers into multi-groups and calibrating travel disutility functions.

### 17.2.1 Time-Dependent O-D Matrices

A time-dependent O-D matrix is considered as given in dynamic route choice models. For joint departure time/route choice models, the total number of departures between each O-D pair during a certain period of time has to be given. However, such a time-dependent O-D matrix does not exist in general. To overcome this difficulty, a tentative approach is to transform a 24 hour O-D matrix into an approximate time-dependent O-D matrix by using a prespecified transformation function.

In estimating a time-dependent O-D matrix, the traditional maximum likelihood method can be used. Data collected from probe vehicles and traffic counts will serve this purpose. In addition, a time-dependent O-D matrix has to be estimated for different modes, such as passenger car, truck, HOV and bus.

The zone size for the time-dependent O-D matrix must be appropriate to the scale of the model. In general, it is smaller than the zones defined for static planning models. The time intervals for time-dependent O-D estimation can be 5, 10, 15 or 30 minutes, depending on the problem requirements. We note that zone size also depends on the length of a time interval. If a time interval is smaller, the zone can be larger and vice versa. During a time interval, the O-D departure flow rate is assumed to be constant. Since the O-D departure flow rate does not change substantially, as compared with the traffic flow during a short time interval, the interval for estimating a time-dependent O-D matrix can be larger than the interval used in dynamic route choice problems.

### 17.2.2 Network Geometry and Control Data

Network geometry and control data for dynamic problems must be more detailed than for static problems. For freeways, data are required for freeway segments, ramps and weaving areas. In general, network geometry and control data for freeways include:

1. link length;
2. free flow speed and speed limit on freeway segments and ramps;
3. length of weaving areas;
4. types of ramp control.

If a freeway has HOV lanes, these lanes should be treated as a set of special links. The length of HOV lanes and charges for single occupancy vehicles are additional data inputs. If there is main line metering on the freeway, a control algorithm is needed.

For arterials, the network geometry and control data include:

1. link length;
2. length and number of lanes for turning movements;
3. number of midblock lanes;
4. number of bus stops on each link;
5. pedestrian activity;
6. number of bus/HOV lanes;
7. free flow speed and speed limit;
8. intersection control data.

The intersection controls can be classified more specifically as follows:

1. links with signalized intersections:
   (a) fixed signal control;
   (b) actuated signal control;
2. links with unsignalized intersections:
   (a) major/minor priority intersections;
   (b) all-way-stop intersections.

For local streets, the intersection controls can be classified as:

1. stop/yield control;
2. no control.

Intersection control data need to be collected for each of the above control types. In addition, offset data are necessary for signalized intersections.

### 17.2.3 Traffic Flow Data

Traffic flow data are needed to calibrate dynamic link travel time functions. Ideally, both detector data and probe data are desired. For a freeway segment, speed/occupancy and travel time data are required. When validating dynamic link travel time functions under highly congested conditions, information on queue length is also necessary.

For an arterial link, the following traffic flow data are required:

1. inflow/exit flow rates;
2. queue length;
3. link travel time;
4. intersection control parameter.

These traffic flow data can also be used to validate the FIFO constraints and oversaturation constraints.

### 17.2.4 Traveler Information

Since more realistic dynamic network models are based on travel costs or disutilities instead of travel times, it is necessary to stratify travelers into multiple groups according to their socio-economic characteristics. As discussed in Chapters 11-13 and in the dynamic prediction module in Section 17.1, traveler information should be collected based on: income and age; route diversion willingness; driving behavior; compliance degree with guidance information; compliance with current or predicted information.

In the following, we present some parameters which need to be calibrated and some travel characteristics data which need to be collected:

1. fixed travel disutilities (fuel consumption, automobile operating costs, etc.);
2. parameters to transform travel times into disutilities;
3. dispersion parameter for SDUO route choice models;
4. desired arrival time interval;
5. parameter for early arrival bonus;
6. parameter for late arrival penalty;
7. dispersion parameter for mode choice.

The above data must be collected for each group of travelers and parameters must be calibrated for each group of travelers as well.

## 17.3 Notes

The implementation issues discussed in Section 17.1 involve only a portion of the potential applications of dynamic network models. As the backbone of ATMS and ATIS, dynamic network models can be applied in different traffic control systems and various user services for travelers. Since IVHS will definitely change travelers' behavior and the location of facilities, dynamic network models can be useful in both aspects. Another area is the integration with telecommunication systems, because transportation and telecommunication are more and more closely related in IVHS. The data requirements in Section 17.2 are generated based on the present level of development of dynamic network modeling. Further requirements are possible as the state of the art of dynamic network modeling evolves.

Successful deployment of dynamic network models in ATMS and ATIS requires extensive effort in the future. According to Yagar and Santiago (1993) and Solanki and Rathi (1993), ongoing research on dynamic network models should address the following issues:

1. integration of models of freeways and surface streets;
2. modeling dynamic route selection;
3. emulation of adaptive signal systems;
4. replication of surveillance and communication functions;
5. representation of driver behavior in ATMS/ATIS implementation;
6. incorporation of safety, energy and environmental impacts;
7. common databases and interfaces with other models;
8. interface with data reduction software at a Traffic Control Center;
9. real time capability;
10. simulation of Automated Vehicle Control Systems and Automated Highway Systems.

# References


Abkowitz M. 1981. Understanding the Effect of Transit Service Reliability on Work Travel Behavior. *Transportation Research Record*, **794**, 33-41.

Akcelik R. 1981. Traffic Signals: Capacity and Timing Analysis. *Australian Road Research Board Research Report*, **123**.

Akcelik R. 1988. The Highway Capacity Manual Formula for Signalized Intersections. *ITE Journal*, **58(3)**, 23-27.

Akcelik R. and Rouphail N.M. 1991. Estimation of Delays at Traffic Signals for Variable Demand Conditions. *Australian Road Research Board Report* **WD TE91/005**.

Alfa A.S. and Minh D.L. 1979. A Stochastic Model for the Temporal Distribution of Traffic Demand—the Peak Hour Problem. *Transportation Science*, **13**, 315-324.

Bard J.F. 1984. Optimality Conditions for the Bilevel Programming Problem. *Naval Research Logistics Quarterly*, **31**, 13-26.

Bazaraa M.S., Sherali H.D. and Shetty C.M. 1993. *Nonlinear Programming Theory and Algorithms*, 2nd edition, John Wiley & Sons, New York.

Beckmann M., McGuire C.B. and Winsten C.B. 1956. *Studies in the Economics of Transportation.* Yale University Press, New Haven.

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

Ben-Akiva M., Cyna M. and de Palma A. 1984. Dynamic Model of Peak Period Congestion. *Transportation Research*, **18B**, 339-355.

Ben-Akiva M., de Palma A. and Kanaroglu P. 1986. Dynamic Model of Peak Period Traffic Congestion with Elastic Arrival Rates. *Transportation Science*, **20**, 164-181.

Ben-Akiva M. and Lerman S.R. 1985. *Discrete Choice Analysis*, MIT Press, Cambridge.

Boyce D.E. 1978. Equilibrium Solution to Combined Urban Residential Location, Mode Choice and Trip Assignment Models. in *Competition Among Small Regions*, W. Buhr and P. Friedrich (eds), Nomos Verlagsgesellschaft, Baden-Baden, Germany, 246-264.

Boyce D.E. 1980. A Framework for Constructing Network Equilibrium Models of Urban Location. *Transportation Science*, **14**, 77-96.

Boyce D.E. 1989. Route Guidance Systems for Managing Urban Transportation Networks: Review and Prospects. Presented at *Tenth Italian Regional Science Conference*, Rome, Italy.

Boyce D.E., Chon K.S., Lee Y.J. and Lin K. T. 1983. Implementation and Computational Issues for Combined Models of Location, Mode, and Route Choice, *Environment and Planning A*, **15**, 1219-1230.

Boyce D.E., Day N.D. and McDonald C. 1970. *Metropolitan Plan Making.* Monograph Series No.4, Regional Science Research Institute, Philadelphia.

Boyce D.E., Kirson A.M. and Schofer J.L. 1994. ADVANCE - The Illinois Dynamic Navigation and Route Guidance Demonstration Program. in *Advanced Technology for Road Transport*, I. Catling (ed.), Artech House, Boston, MA, 247-270.

Boyce D.E., LeBlanc L.J. and Chon. K.S. 1988. Network Equilibrium Models of Urban Location and Travel Choices: A Retrospective Survey. *Journal of Regional Science*, **28**, 159-183.

Boyce D.E., Ran B. and LeBlanc L.J. 1991. Dynamic User-Optimal Traffic Assignment: A New Model and Solution Techniques. Presented at First Triennial Symposium on Transportation Analysis, Montreal, Canada.

Bryson, Jr. A.E. and Ho Y.C. 1975. *Applied Optimal Control.* Hemisphere Publishing Corp., Washington, DC.

Burrow I.A. 1989. A Note on Traffic Delay Formulas. *ITE Journal*, **59(10)**, 29-32.

Canon M.D., Cullum, Jr., C.D. and Polak E. 1970. *Theory of Optimal Control and Mathematical Programming.* McGraw-Hill, New York.

Carey M. 1987. Optimal Time-Varying Flows on Congested Networks. *Operations Research*, **35**, 58-69.

Carey M. 1992. Nonconvexity of the Dynamic Traffic Assignment Problem. *Transportation Research*, **26B**, 127-133.

Cascetta E. 1991. A Day-to-Day and Within-Day Dynamic Stochastic Assignment Model. *Transportation Research*, **25A**, 277-291.

Cascetta E., Inaudi D. and Marquis G. 1993. Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts. *Transportation Science*, **27**, 363-373.

Cesari L. 1983. *Optimization - Theory and Applications.* Springer-Verlag, New York.

Chang G.L., Ho P.K. and Wei C.H. 1993. A Dynamic System-Optimum Control Model for Commuting Traffic Corridors. *Transportation Research*, **1C**, 3-22.

Chen M. and Alfa A.S. 1991. Algorithms for Solving Fisk's Stochastic Traffic Assignment Model. *Transportation Research*, **25B**, 405-412.

Codelli D.K., Niedringhaus W.P., Wang P.T.R. and Young B.S. 1993. *User's Manual for Traffic and Highway Object for REsearch Analysis, and Understanding (THOREAU)*, The MITRE Corporation, McLean, VA.

Codina E. and Barcelo J. 1991. An Algorithm for Dynamic Traffic Assignment Model Based on Optimal Control Approach. Presented at First Triennial Symposium on Transportation Analysis, Montreal, Canada.

Courage K.G., Wallace C.E. and Alqasem, R. 1988. Modeling the Effect of Traffic Signal Progression on Delay. *Transportation Research Record* **1194**, 139–146.

Cruz Jr. J.B. 1978. Leader-Follower Strategies for Multilevel Systems. *IEEE Transactions on Automatic Control* **AC-23**, **2**, 244-254.

Dafermos S.C. 1972. The Traffic Assignment Problem for Multiclass-User Transportation Networks. *Transportation Science*, **6**, 73-87.

Dafermos S.C. 1980. Traffic Equilibrium and Variational Inequalities. *Transportation Science*, **14**, 42-54.

Dafermos S.C. and Nagurney A. 1984a. On Some Traffic Equilibrium Theory Paradoxes. *Transportation Research*, **18B**, 101-110.

Dafermos S.C. and Nagurney A. 1984b. Stability and Sensitivity Analysis for the General Network Equilibrium-Travel Choice Model. *Proceedings of 9th International Symposium on Transportation and Traffic Theory*, VNU Science Press, Utrecht, The Netherlands, 217-234.

Dafermos S.C. and Sparrow F.T. 1969. The Traffic Assignment Problem for a General Network. *Journal of Research of the National Bureau of Standards*, **72B**, 91-118.

Daganzo C.F. 1985. The Uniqueness of a Time-Dependent Equilibrium Distribution of Arrivals at a Single Bottleneck. *Transportation Science*, **19**, 29-37.

Daganzo C.F. 1993. The Cell Transmission Model. Part I: A Simple Dynamic Representation of Highway Traffic. *PATH Report*, **93-0409-3**, University of California at Berkeley.

Daganzo C.F. and Sheffi Y. 1977. On Stochastic Models of Traffic Assignment, *Transportation Science*, **11**, 253-274.

de Palma A., Ben-Akiva M., Lefevre C. and Litinas N. 1983. Stochastic Equilibrium Model of Peak Period Traffic Congestion. *Transportation Science*, **17**, 430-453.

de Palma A. and Lindsey R. 1992. The Potential Benefits of A Combined Route Guidance and Road Pricing System: An Economic Analysis. Presented at the Second Capri Seminar on Traffic Networks, Capri, Italy.

Dial R.B. 1971. Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration. *Transportation Research*, **5**, 83-111.

Evans S.P. 1973. *Some Applications of Optimisation Theories in Transport Planning*. Ph.D. dissertation, University College London, London.

Evans S.P. 1976. Derivation and Analysis of Some Models for Combining Trip Distribution and Assignment. *Transportation Research*, **10**, 37-57.

Fambro, D.B., Chang, E.C.P., and Messer, C.J. 1991. Effects of the Quality of Traffic Signal Progression on Delay, *National Cooperative Highway Research Program Report*, **339**, Transportation Research Board, Washington DC.

Fambro D.B., and Messer, C.J. 1991. Estimating Delay at Coordinated Signalized Intersections, *Proceedings, International Symposium on Highway Capacity*, Karlsruhe, A.A. Balkema, Rotterdam, 127–143.

Fisk C. 1980. Some Developments in Equilibrium Traffic Assignment. *Transportation Research*, **14B**, 243-255.

Fisk C. and Boyce D.E. 1983. Alternative Variational Inequality Formulations of Network Equilibrium Travel Choice Problem. *Transportation Science*, **17**, 454-463.

Florian M. 1986. Nonlinear Cost Network Models in Transportation Analysis. *Mathematical Programming Study*, **26**, 167-196.

Florian M. and Fox B. 1976. On the Probabilistic Origin of Dial's Multipath Traffic Assignment Model, *Transportation Research*, **10**, 339-341.

Florian M. and Nguyen S. 1978. A Combined Trip Distribution Modal Split and Trip Assignment Model. *Transportation Research*, **12**, 241-246.

Florian M. and Spiess H. 1982. The Convergence of Diagonalization Algorithms for Asymmetric Network Equilibrium Problems. *Transportation Research*, **16B**, 447-483.

Frank M. and Wolfe P. 1956. An Algorithm for Quadratic Programming. *Naval Research Logistics Quarterly*, **3(1-2)**, 95-110.

Friesz T.L. 1985. Transportation Network Equilibrium, Design and Aggregation: Key Developments and Research Opportunities. *Transportation Research*, **19A**, 413-428.

Friesz T.L., Bernstein D., Smith T.E., Tobin R.L. and Wie B.-W. 1993. A Variational Inequality Formulation of the Dynamic Network User Equilibrium Problem. *Operations Research*, **41**, 179-191.

Friesz T.L., Luque F.J., Tobin R.L. and Wie B.-W. 1989. Dynamic Network Traffic Assignment Considered as a Continuous Time Optimal Control Problem. *Operations Research*, **37**, 893-901.

Gallo G. and Pallottino S. 1984. Shortest Path Methods in Transportation Models. *Transportation Planning Models*, M. Florian (ed.), North-Holland, Amsterdam, 227-256.

Ghali M. and Smith M.J. 1993. Traffic Assignment, Traffic Control and Road Pricing. *Proceedings of 12th International Symposium on Transportation and Traffic Theory*, Elsevier Science, Amsterdam, 147-170.

Greenshields B. 1933. A Study of Traffic Capacity. *Highway Research Board Proceedings*, **14**, 468-477.

Hendrickson C. and Kocur G. 1981. Schedule Delay and Departure Time Decisions in a Deterministic Model. *Transportation Science*, **15**, 62-77.

Hendrickson C. and Plank E. 1984. The Flexibility of Departure Times for Work Trips. *Transportation Research*, **18A**, 25-36.

Ho J.K. 1980. A Successive Linear Optimization Approach to the Dynamic Traffic Assignment Problem. *Transportation Science*, **14**, 295-305.

Ho J.K. 1990. Solving the Dynamic Traffic Assignment Problem on a Hypercube Multicomputer. *Transportation Research*, **24B**, 443-451.

Hurdle V.F. 1974. The Effect of Queuing on Traffic Assignment in a Simple Road Network. *Proceedings of 6th International Symposium on Transportation and Traffic Theory*, Elsevier Science, Amsterdam, 519-540.

Hurdle V.F. 1984. Signalized Intersection Delay Models – A Primer for the Uninitiated. *Transportation Research Record*, **971**, 96-105.

Inouye H. 1986. Equilibrium Traffic Assignment on a Network with Congested Flows. *Proceedings of Infrastructure Planning of JSCE*, 177-184.

Janson B.N. 1991. Dynamic Traffic Assignment for Urban Road Networks, *Transportation Research*, **25B**, 143-161.

Janson B.N. 1993. Dynamic Traffic Assignment with Arrival Time Costs. *Proceedings of 12th International Symposium on Transportation and Traffic Theory*, Elsevier Science, Amsterdam, 127-146.

Kamien M.L. and Schwartz N.L. 1981. *Dynamic Optimization: the Calculus of Variations and Optimal Control.* Elsevier North Holland, New York.

Kaufman D.E., Nonis J. and Smith R.L. 1992. A Mixed Integer Linear Programming Formulation of the Dynamic Traffic Assignment Problem. *Proceedings of 1992 IEEE International Conference on Systems Man and Cybernetics*, **1**, 232-235.

Kaysi I., Ben-Akiva M. and Koutsopoulos H. 1993 An Integrated Approach to Vehicle Routing and Congestion Prediction for Real-Time Driver Guidance. Presented at the 72nd Annual Transportation Research Board Meeting, Washington D.C.

Kimber R.M. and Hollis E.M. 1979. Traffic Queues and Delays at Road Junctions. *TRRL Laboratory Report*, **909**, U.K.

Kinderlehrer D. and Stampacchia G. 1980. *An Introduction to Variational Inequalities and Their Applications.* Academic Press, New York.

Kirk D.E. 1970. *Optimal Control Theory: An Introduction.* Prentice-Hall, Englewood Cliffs, NJ.

Kuhn H. and Tucker A.W. 1951. Nonlinear Programming. *Proceedings of Second Berkeley Symposium of Mathematical Statistics and Probability*, University of California Press, Berkeley, CA.

Lafortune S., Sengupta R., Kaufman D.E. and Smith R.L. 1991. A Dynamic System Model for Traffic Assignment in Networks. *Vehicle Navigation and Information Systems Conference Proceedings*, Society of Automotive Engineers, 701-708.

LeBlanc L.J. 1973. *Mathematical Programming Algorithms for Large Scale Network Equilibrium and Network Design Problems.* Ph.D. dissertation, Northwestern University, Evanston, IL.

LeBlanc L.J. and Abdulaal M. 1982. Combined Mode Split-Assignment and Distribution-Model Split-Assignment Models with Multiple Groups of Travelers. *Transportation Science*, **16**, 430-442.

LeBlanc L.J., Morlok E.K. and Pierskalla W.P. 1975. An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem. *Transportation Research*, **9**, 309-318.

LeBlanc L.J., Ran B. and Boyce D.E. 1992. Dynamic Travel Choice Models for Urban Transportation Networks. *Proceedings of 1992 IEEE International Conference on Systems Man and Cybernetics*, **1**, 225-231.

Lee E.B. and Markus L. 1967. *Foundations of Optimal Control Theory.* John Wiley & Sons, New York.

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

Lieberman E.B. 1993. An Advanced Approach to Meeting Saturated Flow Requirements. Presented at the 72nd Annual Transportation Research Board Meeting, Washington, DC.

Lin F. 1989. Application of 1985 Highway Capacity Manual for Estimating Delays at Signalized Intersection, *Transportation Research Record*, **1225**, 18–23.

Lo H.K., Chatterjee A., Wegmann F., Richards S. and Rathi A. 1994. An Evaluation Framework for IVHS. To appear in *Journal of Transportation Engineering*, ASCE.

Luenberger D.G. 1984. *Linear and Nonlinear Programming.* Second Edition, Addison Wesley, Menlo Park, CA.

Mahmassani H.S. and Chang G.L. 1987. On Boundedly Rational User Equilibrium in Transportation Systems. *Transportation Science*, **21**, 89-99.

Mahmassani H.S. and Herman R. 1984. Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials. *Transportation Science*, **18**, 362-384.

Mahmassani H.S., Hu T. and Jayakrishnan R. 1992. Dynamic Traffic Assignment and Simulation for Advanced Network Informatics (DYNASMART). Presented at The 2nd International Capri Seminar on Urban Traffic Networks, Capri, Italy.

Mahmassani H.S. and Mouskos K. 1988. Some Numerical Results on the Diagonalization Network Assignment Algorithm with Asymmetric Interactions Between Cars and Trucks. *Transportation Research*, **22B**, 275-290.

Mahmassani H.S., Peeta S., Hu T.Y. and Ziliaskopoulos A. 1993. Dynamic Traffic Assignment with Multiple User Classes for Real-Time ATIS/ATMS Applications. in *Large Urban Systems, Proceedings of the Advanced Traffic Management Conference*, Federal Highway Administration, US Department of Transportation, Washington, DC, 91-114.

Matsui H. 1987. A Model of Dynamic Traffic Assignment. *Text of Infrastructure Planning Lectures*, JSCE, **18**, 84-96.

McShane W.R. and Roess R.P. 1990. *Traffic Engineering.* Prentice Hall, Englewood Cliffs, NJ.

Merchant D.K. and Nemhauser G.L. 1978a. A Model and an Algorithm for the Dynamic Traffic Assignment Problems. *Transportation Science*, **12**, 183-199.

Merchant D.K. and Nemhauser G.L. 1978b. Optimality Conditions for a Dynamic Traffic Assignment Model. *Transportation Science*, **12**, 200-207.

Messer C.J. 1990. The 1985 Highway Capacity Manual Delay Equation. *ITE Compendium of Technical Papers*, 60th Annual Meeting, Orlando, FL, 205-209.

Murchland J.D. 1966. Some Remarks on the Gravity Model of Traffic Distribution and an Equivalent Maximisation Formulation. Transport Network Theory Unit, London School of Economics, London.

Nagurney A. 1986. Computational Comparisons of Algorithms for General Asymmetric Traffic Equilibrium Problems with Fixed and Elastic Demands. *Transportation Research*, **20B**, 78-84.

Nagurney A. 1993. *Network Economics: A Variational Inequality Approach.* Kluwer Academic Publishers, Norwell, MA.

Newell G.F. 1990. Stochastic Delays on Signalized Arterial Highways. *Proceedings of 11th International Symposium on Transportation and Traffic Theory*, Elsevier Science, Amsterdam, 589-598.

Nguyen S. 1976. A Unified Approach to Equilibrium Methods for Traffic Assignment, in *Traffic Equilibrium Methods*, M. Florian (ed), *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, Berlin.

Olszewski P.S. 1990. Traffic Signal Delay for Non-Uniform Arrivals, *Transportation Research Record*, **1287**, 42–53.

Pontryagin L.S., Boltyanskii V.G., Gamkrelidze R.V. and Mischenko E.F. 1962. *The Mathematical Theory of Optimal Processes.* Interscience Publishers, New York.

Powell W.B. and Sheffi Y. 1982. The Convergence of Equilibrium Algorithms with Predetermined Step Sizes. *Transportation Science*, **16**, 45-55.

Ran B. 1989. *Dynamic Travel Assignment and Optimum Control for Urban Transportation Networks.* M.S. Thesis, University of Tokyo, Tokyo.

Ran B., Boyce D.E. and LeBlanc L.J. 1992. Dynamic User-Optimal Route Choice Models Based on Stochastic Route Travel Times. Presented at The 2nd International Capri Seminar on Urban Traffic Networks, Capri, Italy.

Ran B., Boyce D.E. and LeBlanc L.J. 1993. A New Class of Instantaneous Dynamic User-Optimal Traffic Assignment Models. *Operations Research*, **41**, 192-202.

Ran B. and Shimazaki T. 1989a. A General Model and Algorithm for the Dynamic Traffic Assignment Problems. *Transport Policy, Management and Technology Towards 2001, Proceedings of Fifth World Conference on Transport Research*, Yokohama, Japan, 463-477.

Ran B. and Shimazaki T. 1989b. Dynamic User Equilibrium Traffic Assignment for Congested Transportation Networks. Presented at the Fifth World Conference on Transport Research, Yokohama, Japan.

Rouphail N.M. 1989. Progression Adjustment Factors at Signalized Intersections, *Transportation Research Record*, **1225**, 8–17.

Rouphail N.M. and Akcelik R. 1991. Paired Intersections: Initial Development of Platooned Arrival and Queue Interaction Models. *Australian Road Research Board Report* **WD TE91/010.**

Russak B.I. 1970. On General Problems with Bounded State Variables. *Journal of Optimization Theory and Applications*, **6**, 424-452.

Sage A.P. and White, III C.C. 1977. *Optimum Systems Control.* Second Edition. Prentice-Hall, Englewood Cliffs, NJ.

Sheffi Y. 1985. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods.* Prentice-Hall, Englewood Cliffs, NJ.

Sheffi Y. and Powell W.B. 1982. An Algorithm for the Equilibrium Assignment Problem with Random Link Times. *Networks*, **12**, 191-207.

Singh M.G. and Titli A. 1978. *Systems: Decomposition, Optimization and Control.* Pergamon Press, Oxford, England.

Small K.A. Guest Editorial. 1992. Special Issue: Congestion Pricing. *Transportation*, **19**, 287-291.

Smith M.J. 1979. The Existence, Uniqueness and Stability of Traffic Equilibria. *Transportation Research*, **13B**, 295-304.

Smith M.J. 1993. A New Dynamic Traffic Model and the Existence and Calculation of Dynamic User Equilibria on Congested Capacity-Constrained Road Networks. *Transportation Research*, **27B**, 49-63.

Solanki R.S. and Rathi A.K. 1993. A Conceptual Framework for a Dynamic Traffic Flow Model. in *Large Urban Systems, Proceedings of the Advanced Traffic Management Conference*, Federal Highway Administration, US Department of Transportation, Washington, DC, 59-70.

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

Tabak D. and Kuo B.C. 1971. *Optimal Control and Mathematical Programming.* Prentice-Hall, Englewood Cliffs, NJ.

Takaba S. 1991. Estimation and Measurement of Travel Time by Vehicle Detectors and License Plate Readers. *Vehicle Navigation and Information Systems Conference Proceedings*, Society of Automotive Engineers, 257-267.

Teply S. (ed.) 1984. *Canadian Capacity Guide for Signalized Intersections.* Institute of Transportation Engineers—District 7 and the University of Alberta, Edmonton, Alberta, Canada.

Transportation Research Board. 1985. Highway Capacity Manual. *Special Report*, **209**, Washington, DC.

Van Aerde M. 1992. INTEGRATION: A Dynamic Traffic Simulation/Assignment Model. Presented at the IVHS Dynamic Traffic Assignment and Simulation Workshop, Federal Highway Administration, McLean, VA.

Van As S.C. 1991. Overflow Delay at Signalized Networks. *Transportation Research*, **25A**, 1-7.

Vythoulkas P.C. 1990. Two Models for Predicting Dynamic Stochastic Equilibria in Urban Transportation Networks. *Proceedings of 11th International Symposium on Transportation and Traffic Theory*, Elsevier Science, Amsterdam, 253-272.

Wardrop J.G. 1952. Some Theoretical Aspects of Road Traffic Research, *Proceedings of the Institution of Civil Engineers, Part II*, **1**, 325-378.

Webster F.V. 1958. Traffic Signal Settings. *Road Research Laboratory Technical Paper*, **39**, HMSO, London.

Wie B.-W. 1989. Dynamic System Optimal Traffic Assignment on Congested Multidestination Networks. Presented at the Fifth World Conference on Transport Research, Yokohama, Japan.

Wie B.-W., Friesz T.L. and Tobin R.L. 1990. Dynamic User Optimal Traffic Assignment on Congested Multidestination Networks. *Transportation Research*, **24B**, 431-442.

Wilson A.G. 1969. The Use of Entropy Maximizing Models in the Theory of Trip Distribution, Mode Split and Route Split. *Journal of Transport Economics and Policy*, **3**, 108-126.

Yagar S. 1971. Dynamic Traffic Assignment by Individual Path Minimisation and Queuing. *Transportation Research*, **5**, 179-196.

Yagar S. and Santiago A.J. 1993. *Large Urban Systems. Proceedings of the Advanced Traffic Management Conference*, Federal Highway Administration, US Department of Transportation, Washington, DC.

# Subject Index

# Author Index

# List of Figures

# List of Tables