AN INTEGRATED PATH FLOW ESTIMATOR METHODOLOGY FOR MULTI-MODAL TRANSPORTATION NETWORKS

GUOYUAN LI

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

Department of Civil and Environmental Engineering

An Integrated Path Flow Estimator Methodology for Multi-Modal Transportation Networks

Guoyuan LI

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

JULY 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

_____Guoyuan LI_____ (Name of student)

ABSTRACT

Multi-modal transportation systems provide multiple travel modes to create more sustainable and better-connected cities. The overall landscape of travel demand, transport infrastructure, and transport modes is rapidly changing, which highlights the need for an efficient and practical framework to model travel demand. A path flow estimator (PFE) is a single-level optimization model that serves as a flexible network analysis tool and can use various data sources to perform a range of transportation network analyses. Although PFEs have been widely explored in private car networks, a holistic PFE modeling framework for multi-modal transportation networks, particularly public transit networks, remains lacking. Hence, this thesis aims to bridge this gap by proposing an integrated PFE modeling framework for multi-modal transportation networks. Three specific research questions are considered: (i) how to model travel behavior in transit networks; (ii) how to estimate travel demand in transit networks; and (iii) how to estimate travel demand in multi-modal transportation networks.

Transit equilibrium assignment is an important aspect of travel demand planning and management by predicting the passenger flow patterns in a network. Although transit equilibrium has been extensively addressed in the literature, limited attention has been paid to the aggregate line capacity constraints and individual path constraints on the number-of-transfers. Line capacity and number-of-transfers constraints are two critical factors in transit network equilibrium because (1) transit vehicles cannot carry passengers beyond their capacity and (2) transit passengers typically avoid paths with numerous transfers. This thesis first proposes **a strategy-based transit stochastic user equilibrium model with both line capacity constraints and path constraints on number-of-transfers**. A transit path-set generation procedure is developed to generate transit paths with a limited number of transfers using a route-section-based network representation. The diagonalization method is used to solve the proposed model due to the asymmetric cost function. The diagonalized problem is solved using a path-based partial linearization algorithm embedded with an iterative balancing scheme, which is used to handle the line capacity constraints. A small network is explored to show that a standard strategy or hyperpath might contain an excessive transfer, and two

additional networks are used to demonstrate the features of the proposed model and performance of the developed algorithm.

Origin-destination (OD) travel demand is a critical input for transit equilibrium assignment models, which is rarely measured directly in practice. The rich observation data from automatic passenger counting (APC), automatic fare collection (AFC), and automatic vehicle location (AVL) can be used to estimate the transit travel demand. However, the possibility of a single-level model for OD demand estimation in urban congested transit networks remains unresolved. **A frequency-based PFE is therefore proposed for transit demand estimation**. Two kinds of core inequality constraints are considered: (1) onboard passenger counts of transit line segments from APC and AVL data, and (2) partial OD trip matrices obtained from AFC and AVL data inferred from the passenger alighting stations. Three case studies are presented: the first two illustrate the features and evaluate the performance of the proposed model, and the third one uses the Winnipeg (Canada) transit network to demonstrate the model's applicability to a real-world network.

The prevalence of public transport demonstrates the importance of multi-modal transportation network analyses, for which travel demand is the core input. Current practices for estimating multi-modal OD matrices use a four-step model in a sequential manner. The third part of this thesis therefore focuses on simultaneously considering the mode choice, route choice, vehicle interaction, and various side constraints for the OD demand estimation. **A multi-modal path flow estimator is proposed to estimate travel demand in an urban transportation network**. The model incorporates the limited available observational data as side constraints (e.g., road link traffic counts, onboard passenger counts from bus and metro line segments, mode-specific target OD demand, zonal production and attraction). The interaction of private cars and bus vehicles, route choice behavior of private cars and transit modes, and mode similarity are modeled in a congested network. The mode similarity is captured by adopting a nested logit choice model. Computational tests are performed on the proposed model and developed solution algorithm using data for a hypothetical multi-modal transportation network in Sioux Falls (USA).

This thesis proposes an alternative travel demand forecasting methodology, an integrated PFE that consistently addresses the weaknesses of traditional transport planning models, while acknowledging the difficulties of developing an activity-based travel demand model in rapidly growing urban cities.

Keywords: Multi-Modal Transportation Network; Path Flow Estimator; Transit; Route Section; Number-of-Transfers Constraint; Demand Estimation; Nested Logit

PUBLICATIONS ARISING FROM THE THESIS

Journal publications related with the thesis directly:

Li, G., Chen, A., 2021. Strategy-based transit stochastic user equilibrium model with capacity and number of transfers constraints. (Under revision in the submitted journal)

Li, G., Chen, A., 2021. Frequency-based path flow estimator for transit origin-destination trip matrices incorporating automatic passenger count and automatic fare collection data. (Under revision in the submitted journal)

Li, G., Chen, A., 2021. Multi-modal path flow estimator for origin-destination demand estimation in transportation networks. (Under revision in the submitted journal)

Other journal publications:

Chan, H.Y., Chen, A., **Li, G.**, Xu, X., Lam, W.H.K., 2021. Evaluating the value of new metro lines using route diversity measures: The case of Hong Kong's Mass Transit Railway system. *Journal of Transport Geography* 91, 102945.

Li, G., Chen, A., 2021. Sustainability evaluation of ridesharing in a multi-modal transportation system: A network equilibrium approach. Accepted by the EASTS journal.

Li, G., Chen, A., 2021. Bimodal commute demand estimation: A trip-chain-based path flow estimator. In preparation to *European Journal of Operational Research*.

Li, G., Ryu, S., Chen, A., Kitthamkesorn, S., Xu, X., 2021. Stochastic user equilibrium assignment with paired combinatorial weibit model: fixed and elastic demand formulations. In preparation to *Transportation Science.*

Kurmankhojayev, D., **Li, G.**, Chen, A., Ryu, S., Kitthamkesorn, S., 2021. To fixed or not to fixed? Impact of demand elasticity and choice set size in two logit-based stochastic user equilibrium models. In preparation to *Transportation Research Part A: Policy and Practice.*

Conference papers:

Li, G., Chen, A., Ryu, S., 2017. Solving the multi-modal traffic assignment problem with capacity constraints via the path flow estimator. Paper presented at the *22nd Hong Kong Society of Transportation Studies Conference: Transport and Society*, December 9-11, 2017, Hong Kong, P.R. China.

Chan, H.Y., Chen, A., **Li, G.**, Xu, X., 2017. Evaluating the route redundancy of Hong Kong metro network. Paper presented at the *22nd Hong Kong Society of Transportation Studies Conference: Transport and Society*, December 9-11, 2017, Hong Kong, P.R. China.

Chan, H., Chen, A., Xu, X., **Li, G.**, Lam, W.H.K., 2018. Measuring route diversity for urban rail transit networks. Paper presented at the *6th International Symposium on Reliability Engineering and Risk Management*, May 31-June 1, 2018, Singapore.

Li, G., Chen, A., Ryu, S., 2018. Hyperpath size formulation for stochastic frequency-based transit assignment. Paper presented at the 23rd *Hong Kong Society of Transportation Studies Conference: Transportation Systems in the Connected Era*, December 8-10, 2018, Hong Kong, P.R. China.

Ryu, S., **Li, G.**, Chen, A., Kitthamkesorn, S., Xu, X., 2019. Stochastic user equilibrium with paired combinatorial weibit route choice model. Paper presented at the *98th Transportation Research Board Annual Meeting*, January 13-17, 2019, Washington D.C., USA.

Ryu, S., **Li, G.**, Chen, A., Kitthamkesorn, S., Xu, X., 2019. Variable demand stochastic user equilibrium with paired combinatorial weibit route choice model. Paper presented at the *International Choice Modelling Conference*, August 19-21, 2019, Kobe, Japan.

Gu, Y., **Li, G.**, Chen, A., 2019. Modeling urban location and travel choices in path flow estimator. Paper presented at the *24th Hong Kong Society of Transportation Studies Conference: Transport and Smart Cities*, December 14-16, 2019, Hong Kong, P.R. China.

Li, G., Chen, A., 2021. Sustainability evaluation of ridesharing in a multi-modal transportation system: A network equilibrium approach. *Proceeding of the 14th International Conference of Eastern Asia Society for Transportation Studies (EASTS)*, September 12-15, 2021, Hiroshima, Japan.

Li, G., Kitthamkesorn, S., Chen, A., 2022. Path size correction factor revisited for the weibit-based route choice model. Accepted by the the *101st Transportation Research Board Annual Meeting*, January 9-13, 2022, Washington D.C., USA.

Mansoor, U., **Li, G.**, Chen, A., 2022. Modeling reliability and unreliability of safety in the network equilibrium model: An α-reliable mean-excess approach. Accepted by the the *101st Transportation Research Board Annual Meeting*, January 9-13, 2022, Washington D.C., USA.

Sun, B., **Li, G.**, Chen, A., 2022. Bounding the inefficiency of weibit-based stochastic user equilibrium and its comparison to the logit-based model. Accepted by the the *101st Transportation Research Board Annual Meeting*, January 9-13, 2022, Washington D.C., USA.

ACKNOWLEDGEMENTS

I would first thank my supervisor, Prof. Anthony Chen, for leading me into the field of transportation research. Thank you for your patience, especially because transportation network modeling was once completely foreign to me. Thank you for providing me with sufficient time and space to grasp the fundamental knowledge, which will support my future career. It has been a great fortune to have been your student. Thank you for your continuous support in times when I could not solve certain problems during the normal period. Thank you for your advice and assistance in conducting systematic research and academic writing. Apart from research, your have also provided good suggestions regarding my future career and life. Thank you very much for supporting the completion of my PhD program in Hong Kong for five years.

I would like to thank Dr. Junbiao Su for the five years of working, relaxing, and playing together at PolyU. I would like to thank Dr. Songyot Kitthamkesorn, Dr. Seungkyu Ryu, Dr. Muqing Du, and Prof. Xiangdong Xu for their assistance and useful suggestions and comments when I encountered technical issues. I thank Daniyar Kurmankhojayev for advice on computational coding issues. I would like to thank every friend I met at PolyU for making my life here colourful, and thanks to the Hong Kong Polytechnic University for providing me with a nice study environment and rich resources.

I want to express my deep thanks to my family. I would like to thank my parents for their understanding and support for all of my decisions in these 25 years of study. I would like to thank my sister and brother for taking care at home while I was in Hong Kong. I want to thank my girlfriend, Yifen Liao, for being my side and giving me encouragement and comfort when I faced challenges. I am so lucky to have your company in these three past years and future life.

Last but not least, I want to thank myself and my thirst for knowledge, which has driven me to be a student for 25 years. I am grateful for my self-encouragement, persistence, and faith when I hit the walls during my PhD research process. I believe the experience of completing a PhD program has created a new me.

TABLE OF CONTENTS

Certificate of Originality ..... **i**
Abstract ..... **ii**
Publications Arising from the Thesis ..... **iv**
Acknowledgements ..... **vii**
Table of Contents ..... **viii**
List of Figures ..... **xi**
List of Tables ..... **xiv**
List of Abbreviations ..... **xv**
Chapter 1 Introduction **1**
1.1 Background 1
1.1.1 Related literature about PFE 2
1.2 Research Problems 6
1.2.1 Travel behavior modeling in transit networks 6
1.2.2 Travel demand estimation in transit networks 7
1.2.3 Travel demand estimation in multi-modal networks 7
1.3 Objective and Contributions of the Study 7
1.3.1 Strategy-based transit SUE equilibrium 8
1.3.2 Frequency-based PFE for transit OD demand estimation 8
1.3.3 Multi-modal PFE for OD demand estimation 9
1.4 Structure of the Thesis 9
Chapter 2 Strategy-based Transit Stochastic User Equilibrium Model With Capacity and Number-of-Transfers Constraints **10**
2.1 Introduction 11
2.2 Strategy-based Transit SUE Model with Capacity Constraints 15
2.2.1 Notation 15
2.2.2 Route-section-based transit network revisit 16

2.2.3 Variational inequality formulation with capacity constraints 20
2.3 Transit Path Set Generation Procedure 22
2.4 Solution Algorithm 25
2.4.1 Diagonalization of cost functions 25
2.4.2 Overall solution algorithm framework 27
2.5 Numerical Experiments 28
2.5.1 Small network 28
2.5.2 Sioux Falls network 31
2.5.3 Winnipeg network 39
2.6 Chapter Summary 43
Chapter 3 Frequency-based Path Flow Estimator for OD Demand Estimation in an Urban Transit Network 44
3.1 Introduction 44
3.1.1 Related literature 45
3.1.2 Contribution of this work 48
3.2 Transit Network Modeling 49
3.2.1 Notation 49
3.2.2 Route-section-based transit network modeling revisited 49
3.3 Frequency-based Transit Path Flow Estimator 51
3.3.1 Framework of transit PFE with APC, AFC, and AVL data 51
3.3.2 Flow conservation in transit network 55
3.3.3 Uncongested transit PFE formulation 56
3.3.4 Congested transit PFE formulation 58
3.4 Solution Algorithm 60
3.4.1 Diagonalization of cost functions 60
3.4.2 Overall solution procedure 62
3.5 Numerical Experiments 64
3.5.1 Small network 65
3.5.2 Medium-size network 67
3.5.3 Large network 73
3.6 Chapter Summary 76
Chapter 4 Multi-Modal Path Flow Estimator for Estimating OD Demand in Urban Transportation Networks 77
4.1 Introduction 78

4.2 Multi-Modal Transportation Network Modelling 80
4.2.1 Notations 80
4.2.2 Illustration of interaction between road and bus networks 81
4.2.3 Path travel time for cars 82
4.2.4 Path travel time for bus and metro 83
4.3 Multi-Modal Path Flow Estimator Formulation 85
4.3.1 Mode and route choice modelling 85
4.3.2 Flow conservation in a multi-modal transportation network at different spatial levels 86
4.3.3 Formulated as a variational inequality problem 88
4.4 Solution Algorithm 93
4.4.1 Diagonalization of cost function 93
4.4.2 Overall solution algorithm framework 95
4.4.3 Direction finding: iterative balancing scheme 96
4.5 Numerical Experiment 103
4.5.1 Experiment setting 104
4.5.2 Result analysis 105
4.6 Chapter Summary 111
Chapter 5 Conclusions and Suggestions for Future Research 112
5.1 Conclusion 112
5.2 Further Research Directions 114
5.2.1 Extended integrated PFE 114
5.2.2 Applications of an integrated PFE 115
References 117

LIST OF FIGURES

Figure 1.1 Proposed path flow estimator versus the four-step and activity-based models.....2
Figure 1.2 Different purposes of a path flow estimator.....3
Figure 1.3 A path flow estimator as a traffic assignment with side constraints.....4
Figure 1.4 A path flow estimator for origin-destination (OD) demand estimation.....5
Figure 1.5 Scope of the research problems.....6
Figure 1.6 Thesis at a glance.....8
Figure 2.1 Overview of key static frequency-based transit equilibrium studies.....12
Figure 2.2 Network representation of a transit network (de Cea and Fernández, 1993).....17
Figure 2.3 Illustration of competing sections of a route section.....19
Figure 2.4 Framework of Yen's algorithm (Yen, 1971).....23
Figure 2.5 Selection of the subsequent transfer transit stop.....24
Figure 2.6 Solution algorithm for SUE-T-SC.....27
Figure 2.7 Illustration of a small transit network.....30
Figure 2.8 Result of optimal strategy (shortest hyperpath) with different numbers of transfers.....31
Figure 2.9 Sioux-Falls network.....32
Figure 2.10 Distribution of volume to capacity (V/C) ratio of transit line segments with different number-of-transfers constraints.....34
Figure 2.11 Volume to capacity (V/C) ratio of transit line segments with (SC) and without (No-SC) capacity constraints.....34
Figure 2.12 Flow on transit line 2 with (SC) and without (No-SC) capacity constraints.....35
Figure 2.13 Effect of increasing transit line frequency on the flow of saturated line 2 without capacity constraints.....36
Figure 2.14 Effect of increasing transit line frequency on the flow of saturated line 2 with capacity constraints.....37
Figure 2.15 Distribution of volume to capacity (V/C) ratio of transit line segments with different demand levels.....38

Figure 2.16 Winnipeg transit network: different colors denote different lines.....40
Figure 2.17 Convergence characteristics of the proposed solution algorithm.....41
Figure 2.18 Flow on two transit lines between with (SC) and without (No-SC) capacity constraints.....42
Figure 3.1 Illustration of competing sections of a route section.....51
Figure 3.2 Schematic of frequency-based transit PFE with APC, AFC, and AVL data.....52
Figure 3.3 Illustration of a method of deriving onboard passenger counts using APC data...53
Figure 3.4 Relationship between the complete OD demand matrix and transit data.....55
Figure 3.5 Framework of the solution algorithm for [VI-T-PFE].....62
Figure 3.6 Algorithm of the iterative balancing scheme.....64
Figure 3.7 Example of a small transit network.....65
Figure 3.8 Sioux-Falls road and transit networks.....68
Figure 3.9 Estimation of four scenarios on the Sioux Falls network.....70
Figure 3.10 Estimated vs. observed line segment flow of transit line 4 (both directions, Scenario I and II).....71
Figure 3.11 Estimated vs. observed line segment flow of transit line 4 (both directions, Scenario III and IV).....71
Figure 3.12 Effect of different values of the observed partial OD trip matrix.....72
Figure 3.13 Effect of the number of transfers on the estimation results.....73
Figure 3.14 Winnipeg transit network: different colors denote different lines.....74
Figure 3.15 Comparison of the observed and estimated line segment flows for the Winnipeg network.....75
Figure 3.16 Distribution of the absolute relative error of demand estimation.....75
Figure 4.1 Relationship between road and bus networks.....82
Figure 4.2 Illustration of mode and route choice.....85
Figure 4.3 Solution algorithm for MM-PFE.....96
Figure 4.4 Algorithm of the three-layer iterative balancing scheme.....103
Figure 4.5 Hypothetical multi-modal transportation network of Sioux Falls.....105
Figure 4.6 Convergence characteristics of the solution algorithm.....107
Figure 4.7 Comparison between the observed zonal productions and attractions and those calculated using the estimated mode-specific OD demand.....108
Figure 4.8 Comparison between the observed and estimated target mode-specific OD demands.....109

Figure 4.9 Comparison between the estimated mode-specific OD demands and those calculated using the estimated mode-specific path flow.....110
Figure 4.10 Comparison between the observed and estimated link (line segment) flows.....110

LIST OF TABLES

Table 1.1 Path flow estimator (PFE) applications 5
Table 2.1 Transit line data for Sioux Falls network.....32
Table 2.2 True demand matrix for Sioux Falls 33
Table 3.1 Differences between existing transit OD demand estimation models and our model 47
Table 3.2 Path information for each OD pair 65
Table 3.3 Estimated transit OD flows.....66
Table 3.4 Estimated transit line segment flows 66
Table 3.5 Transit line data for Sioux Falls network.....68
Table 3.6 True demand matrix for Sioux Falls 69
Table 4.1 Adjustment factors of dual variables 100
Table 4.2 OD pair setting in Sioux Falls.....105

LIST OF ABBREVIATIONS

AFC	Automatic Fare Collection
APC	Automatic Passenger Count
AVL	Automatic Vehicle Location
ARE	Absolute Relative Error
MM-PFE	Multi-Modal Path Flow Estimator
MNL	Multinomial Logit
NL	Nested Logit
OD	Origin-Destination
PFE	Path Flow Estimator
RMSE	Root Mean Square Error
SUE	Stochastic User Equilibrium
SUE-T-SC	Transit Stochastic User Equilibrium with Side Constraints
T-PFE	Transit Path Flow Estimator
V/C	Volume to Capacity Ratio
VI	Variational Inequality

CHAPTER 1

INTRODUCTION

Network equilibrium approaches are vital tools for modeling urban multi-modal transportation systems (e.g., policy evaluation, impact analysis, project planning development, and operation strategy optimization), for which travel demand is a critical input. Hence, this thesis aims to provide a holistic modeling framework for estimating travel demand in urban multi-modal transportation systems.

1.1 Background

Many urban cities are experiencing rapid growth due to accelerating urbanization. Multi-modal transport networks provide multiple travel modes (e.g., cars, metros, buses, and ridesharing) to build better connected cities. Governments also promote public transit modes over private car usage to achieve green and sustainable cities. Most modern urban cities have a multi-modal transport system that emphasizes public transport over private transport. For example, over 90% of daily trips in Hong Kong are made using multiple public transport modes (e.g., Mass Rapid Transit, franchised buses, public light buses, and ferries) (Transport Department, 2014). The overall landscape of travel demand, transport infrastructure, and transport modes is rapidly changing, which requires an efficient and practical framework for modeling travel demand.

Current transport planning practices involve one of two travel demand forecasting procedures: (1) the traditional four-step procedure developed in North America and the United Kingdom, which consists of trip generation, trip distribution, modal splitting, and traffic assignment in a top-down sequential process to evaluate alternative road and transit plan; (2) the state-of-the-art activity-based procedure is highly data-intensive and time consuming and

thus largely inappropriate for planning applications in growing urban cities. Moreover, neither of these approaches can use the information contained within the observational data (e.g., traffic counts of road links and transit line segments).

An integrated path flow estimator (PFE) is an alternative travel demand forecasting methodology that can address the weaknesses of traditional transport planning models in a consistent manner, while acknowledging the difficulties of developing an activity-based travel demand model in rapidly-growing urban cities as shown in Figure 1.1. Transportation network models are generally twofold: traffic assignment and its reverse process (i.e. origin-destination [OD] estimation). Travel demand is the critical input for multi-modal transportation system modeling, and its quality affects the demand estimation accuracy. However, travel demand data are rarely available or only roughly known. **This motivates us to explore an integrated PFE framework to estimate travel demand in a multi-modal transportation system**. Compared with the traditional four-step model and the activity-based model mentioned above, the integrated PFE framework in this thesis has some advantages: (i) consistent structure, (ii) various side constraints related to multiple data sources, and (iii) analytical expression for path flow (mode-specific OD demand).

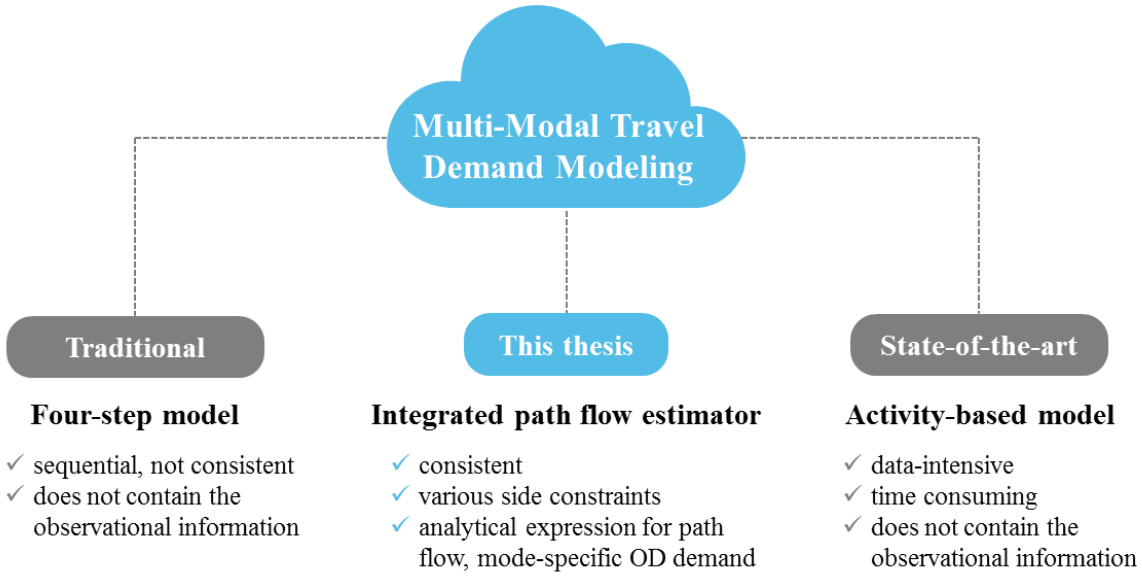


Figure 1.1 Proposed path flow estimator versus the four-step and activity-based models

1.1.1 Related literature about PFE

1.1.1.1 What is a PFE?

A PFE is a flexible network analysis tool that can use various data sources to perform a variety of transportation network analyses (Bell and Iida, 1997). A PFE generates stochastic user equilibrium (SUE) traffic flow patterns by adopting a discrete choice model (e.g., multinomial logit). The flow patterns contain information from various data sources related to the historical demand. The theoretical advantage of a PFE is its single-level mathematical formulation, which allows the flexibility of incorporating different data sources as side constraints. The optimization formulation can also obtain unique optimal solutions (i.e., path flows) because the objective function is strictly convex with respect to the decision variables, and the feasible set is also convex (i.e., all of the equality and inequality constraints are linear). Various kinds of flows can be derived at different spatial levels using the obtained path flows, including network flows, district flows, zonal flows, origin-destination (OD) flows, link flows, intersection turning movement flows, and so on.

1.1.1.2 Flexibility of a PFE

The PFE method can be used as a flexible network analysis tool for different purposes due to its flexibility for specifying different side constraints using a variety of data sources and the ability to aggregate path flows at different spatial levels (Figure 1.2). A PFE can be a traffic assignment problem without or with side constraints if the OD matrix and transportation network are denoted graphically (Bell, 1995; Chen et al., 2011; Ryu et al., 2014a). A PFE can also serve as an OD demand estimator if network measurements are available (Bell et al., 1997; Chen et al., 2005, 2009, 2010; Chootinan et al., 2005; Nie et al., 2005).

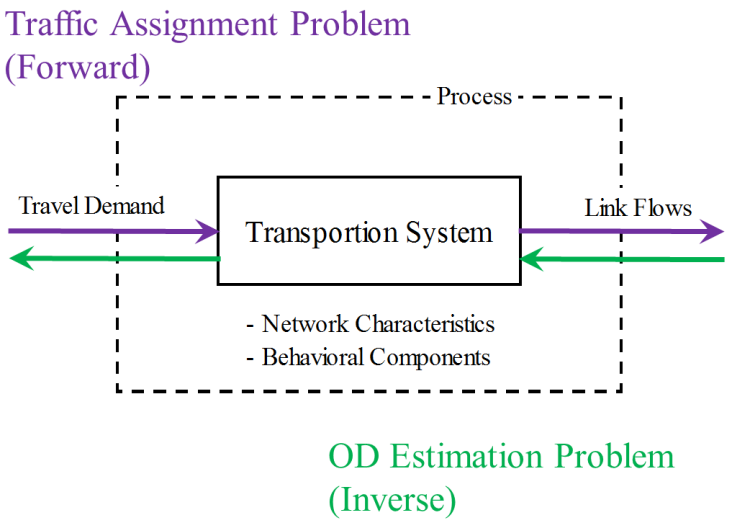


Figure 1.2 Different purposes of a path flow estimator

A PFE as a traffic assignment with side constraints

As shown in Figure 1.3, a PFE can be used a traffic assignment with side constraints using the given OD matrix and a transportation network denoted as a graph. The main procedure contains three parts: input, model, and output. The input is divided into two groups: compulsory and optional. Three main flexibilities of a PFE are emphasized: (1) improved realism of the traffic assignment with side constraints (e.g., link capacity, observed flows on major links, and environmental restrictions), which avoids the feedback step of the traditional four-step model; (2) different discrete choice models are embedded to model travel behavior (i.e., logit, extended logit [C-logit, path-size logit, cross-nested logit, and pair combinatorial logit], and weibit); and (3) the model is extended to include other choice dimensions (e.g., route and mode choices, route and destination choices, and route, mode, destination, and travel choices).

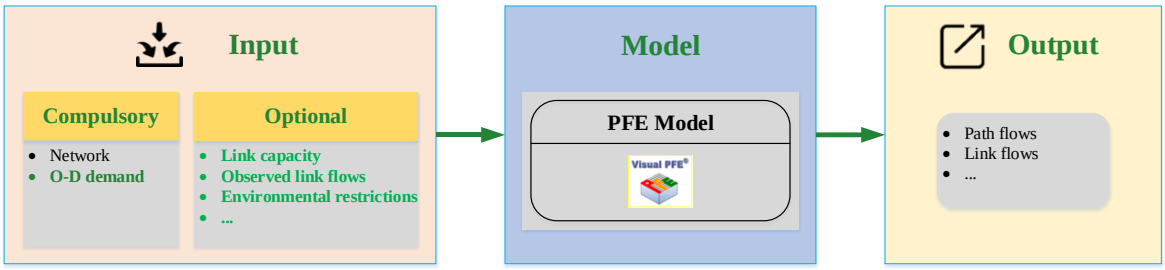


Figure 1.3 A path flow estimator as a traffic assignment with side constraints

A PFE for OD demand estimation

A PFE can serve as an OD demand estimator when network measurements are available (e.g., link traffic counts), as shown in Figure 1.4. This procedure also consists of three parts: input, model, and output. Unlike the input in the traffic assignment, the link traffic count data belong to the input part and the OD demand must be calculated. There are two main flexibilities: (1) the incorporation of different data sources, including field data (e.g., traffic counts and intersection turning movements) and planning data (e.g., land use); (2) different discrete choice models are embedded, such as logit, extended logit, and weibit, which is similar to that when using a PFE for traffic assignment).

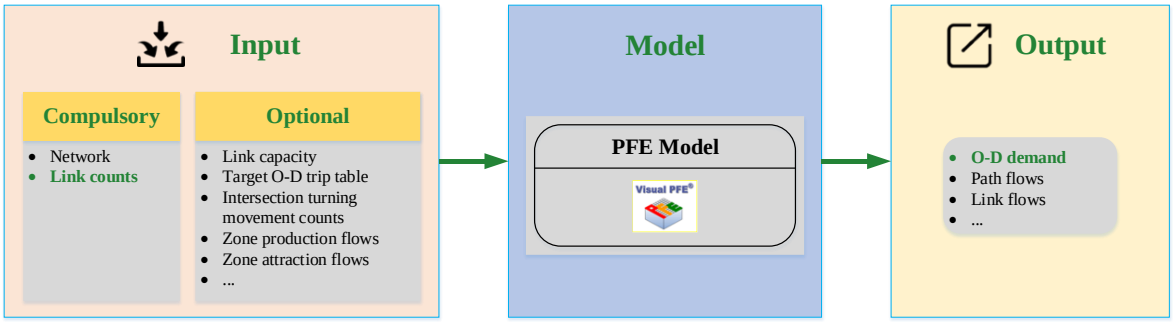


Figure 1.4 A path flow estimator for origin-destination (OD) demand estimation

1.1.1.3 Applications of a PFE

Table 1.1 Path flow estimator (PFE) applications

Research Focus	Study	Network
Multi-class PFE with different levels of network information	Bell et al. (1996)	Private car network
Intersection turning movement estimation	Chen et al. (2012)	
Statewide truck origin-destination demand estimation	Jansuwan et al. (2017)	
Network reliability assessment	Bell et al. (1999), Lam and Xu (1999), Cheng et al. (2002)	
Simplified planning tool for small communities	Jansuwan et al. (2012), Ryu et al. (2014b)	
PFE in bi-modal networks without handling the common lines issue for public transit mode	Bell and Cassir (1998)	Bi-modal network
Bicycle network analysis tool	Ryu et al. (2018)	Bicycle network
Air travel demand estimation	Li et al. (2013); Li (2016)	Air network
Maritime container assignment	Bell et al. (2011)	Maritime network

Most of the existing PFE methods focus on private transport (Table 1.1). However, the travel mode structure in Hong Kong differs substantially from that in other cities worldwide. There are many public transportation modes in Hong Kong (e.g., metro, light rail, franchised bus, non-franchised bus, minibus, taxi, ferry, and tram) and their modes are combined for travelers to choose between each OD pair. Public transportation is vital to people's quality of

life and social-economic development. The existing PFE methods therefore cannot be directly applied for the public transport mode because of the fundamental differences between private and public transport networks. This motivates us to develop PFE methods for not only public transport modes but also multiple transport modes.

1.2 Research Problems

This thesis focuses on the problem of **multi-modal travel demand estimation in an urban transportation system**. Three main research questions are considered to model a multi-modal urban transportation system (Figure 1.5): travel behavior modeling in a transit network, travel demand estimation in a transit network, and travel demand estimation in a multi-modal transportation network. The challenges and methodologies for solving these questions are presented in chapters 2-4.

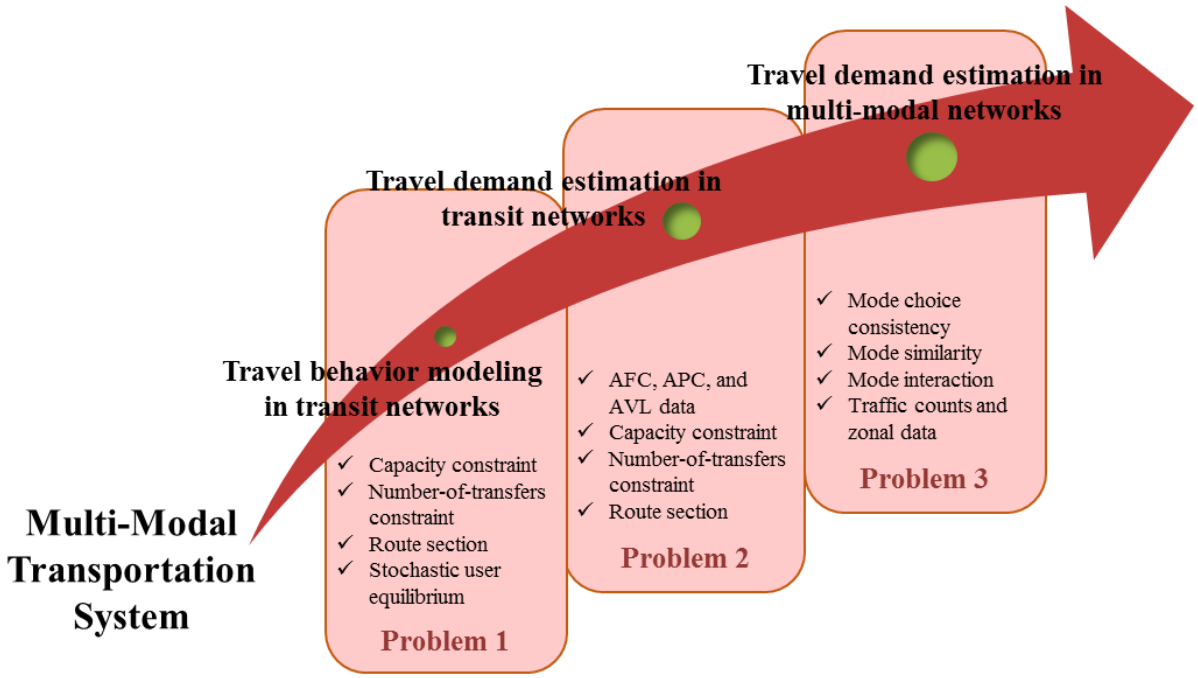


Figure 1.5 Scope of the research problems

1.2.1 Travel behavior modeling in transit networks

Transit network modeling usually consists of transit assignment and transit demand estimation. Transit choice behavior differs from that in private car networks. In reality, passengers will not choose a path with an excessive number of transfers. In contrast to the rich literature on transit equilibrium models, very little attention has been paid to the transfer issue.

In particular, transit vehicles cannot carry passengers over their capacity, thus a strict capacity constraint should be considered. Chapter 2 proposes a strategy-based transit stochastic user equilibrium model that considers capacity and number-of-transfers constraints.

1.2.2 Travel demand estimation in transit networks

The OD demand is a critical input for transit equilibrium models. Previous studies on network-based transit demand estimation have been bi-level and required heuristic algorithms, which do not guarantee that a global optimal solution can be obtained. Furthermore, data-based approaches using automatic fare collection (AFC), automatic vehicle location (AVL) and automatic passenger count (APC) usually generate an incomplete OD demand matrix. In Chapter 3, a single-level frequency-based PFE is proposed to estimate transit OD demand using AFC and APC data.

1.2.3 Travel demand estimation in multi-modal networks

Multi-modal transportation planning and management require high-quality travel demand data. There are three drawbacks to separately estimating the mode-specific OD demand: (1) poor mode choice consistency (e.g., multinomial logit [MNL] and nested logit [NL]); (2) the interaction between different vehicles in the road network cannot be modeled; and (3) the travel behavior of the different modes is not integrated. Very few studies have explored the multi-modal travel demand estimation problem except for a bi-level model by García-Ródenas and Marín (2009). Hence, Chapter 4 proposes a single-level multi-modal PFE with an NL mode choice model to estimate multi-modal travel demand in an urban transportation network.

1.3 Objective and Contributions of the Study

The objective of this thesis is to develop **an integrated PFE methodology** to estimate travel demand in an urban multi-modal transportation network. We first explore the transit stochastic user equilibrium with capacity and number-of-transfers constraints. The focus then shifts to transit OD demand estimation. An NL model is adopted to explore the multi-modal travel demand estimation problem. The details of these research components are summarized in Figure 1.6.

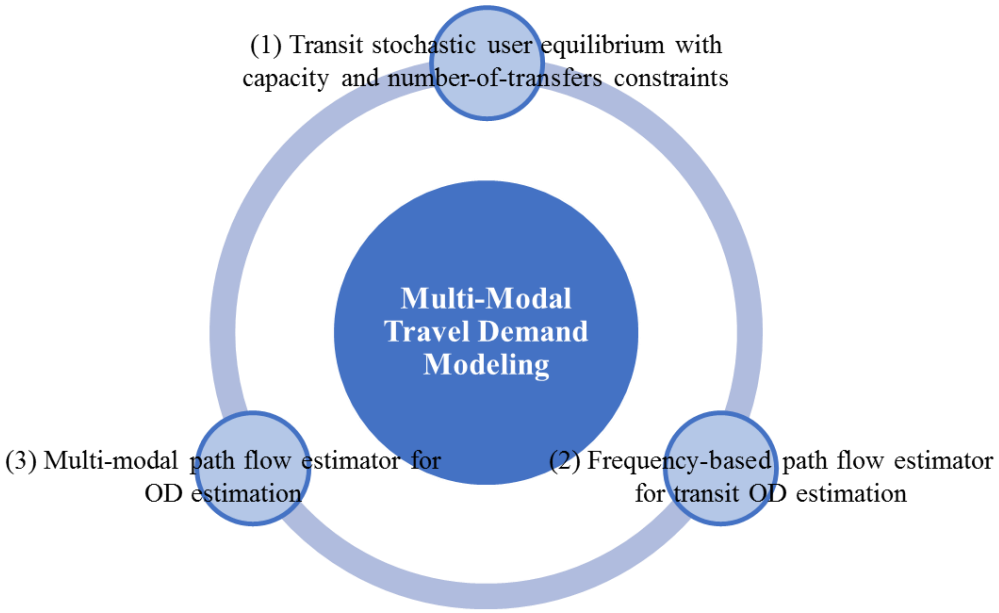


Figure 1.6 Thesis at a glance

1.3.1 Strategy-based transit SUE equilibrium

In this work, a strategy-based transit SUE model is proposed that considers capacity and number-of-transfers constraints for transit planning. The equilibrium model is formulated as a variational inequality (VI) problem. A transit path set generation procedure with number-of-transfers constraint is also proposed for a route-section-based transit network. A diagonalized method is adopted to solve the asymmetric model, and the diagonalized model is solved using a developed path-based partial linearization algorithm embedded with an iterative balancing scheme and self-regulated averaging scheme. After using a hypothetical transit network to illustrate the model features, a real-case transit network in Winnipeg, Canada, is used to demonstrate the applicability of the model and algorithm.

1.3.2 Frequency-based PFE for transit OD demand estimation

We propose a frequency-based PFE for OD demand estimation in a congested urban transit network. The key features of the proposed model are as follows: (1) a route-section-based approach is adopted to model the transit passenger choice behavior; (2) the congestion effect is taken into account; and (3) different transit data sources serve as different side constraints (e.g., onboard passenger count constraints from APC data and partial OD matrix constraints from AFC and AVL data). Overall, the proposed model is designed to address the

issue of OD matrix under-specification and overcome their partial characteristics. A path-based diagonalization algorithm embedded with an iterative balancing scheme is developed to solve the proposed model. The diagonalization framework is used to handle the issue of asymmetric cost functions, and the diagonalized convex optimization model is solved using a partial linearization algorithm. The embedded iterative balancing scheme enables the model to handle multiple inequality side constraints. A small network is used to illustrate the correctness of the proposed model and developed algorithm. A medium-sized network based on the Sioux Falls (USA) road network is chosen to demonstrate the features of the model and a large network in Winnipeg is used to show the applicability of the model and algorithm in a real-world transit network.

1.3.3 Multi-modal PFE for OD demand estimation

A multi-modal PFE model with an NL choice model, which captures the mode similarity and interaction, is proposed to estimate the OD demand in an urban multi-modal transportation network. Several kinds of available observational data are incorporated as equality or inequality side constraints, such as road link traffic counts, onboard passenger counts of bus and metro line segments, mode-specific target OD demand, and zonal production and attraction. A three-level iterative balancing scheme is developed for direction finding when solving the proposed optimization model. The proposed model and developed algorithm are tested in the hypothetical multi-modal transportation network of Sioux Falls.

1.4 Structure of the Thesis

The remainder of the thesis is organized as follows.

- In Chapter 2, a strategy-based transit SUE model with capacity and number-of-transfers constraints is proposed and a transit path set generation procedure is developed to identify a transit path with a limited number of transfers in a route-section-based transit network.
- In Chapter 3, a frequency-based path flow estimator is proposed for estimating OD demand in a congested transit network using AFC and APC data.
- In Chapter 4, a multi-modal path flow estimator with an NL choice model, which captures the mode similarity and interaction, is proposed to estimate the multi-modal travel demand using available traffic information (traffic counts, target OD demand and zonal data).
- Chapter 5 summarizes the conclusions of the thesis.

CHAPTER 2

STRATEGY-BASED TRANSIT STOCHASTIC USER EQUILIBRIUM MODEL WITH CAPACITY AND NUMBER-OF-TRANSFERS CONSTRAINTS

Line capacity and number-of-transfers constraints are critical in transit network equilibrium because (1) transit vehicles cannot carry passengers over their capacity and (2) transit passengers typically avoid paths with numerous transfers. In this chapter, we propose a strategy-based transit stochastic user equilibrium (SUE) model that considers line capacity and number-of-transfers constraints for an urban congested transit network. A route-section-based method is used for the transit network representation and the transit passengers' route choice behavior is assumed to obey the multinomial logit model. The transit line capacity and maximum number-of-transfers constraints are considered in the model. We then formulate the strategy-based transit SUE problem as a variational inequality (VI) problem. A transit path-set generation procedure is proposed to identify a transit path with a limited number of transfers using the route-section-based network representation. The diagonalization method is chosen to solve the VI problem due to the asymmetric cost function, and the diagonalized problem can be solved using a path-based partial linearization algorithm embedded with an iterative balancing scheme, which is used here to handle the numerous capacity constraints. Numerical examples are conducted to demonstrate the features of the proposed model and performance of the developed algorithm. The results show that the line capacity and number of transfers would strongly impact the passenger flow patterns.

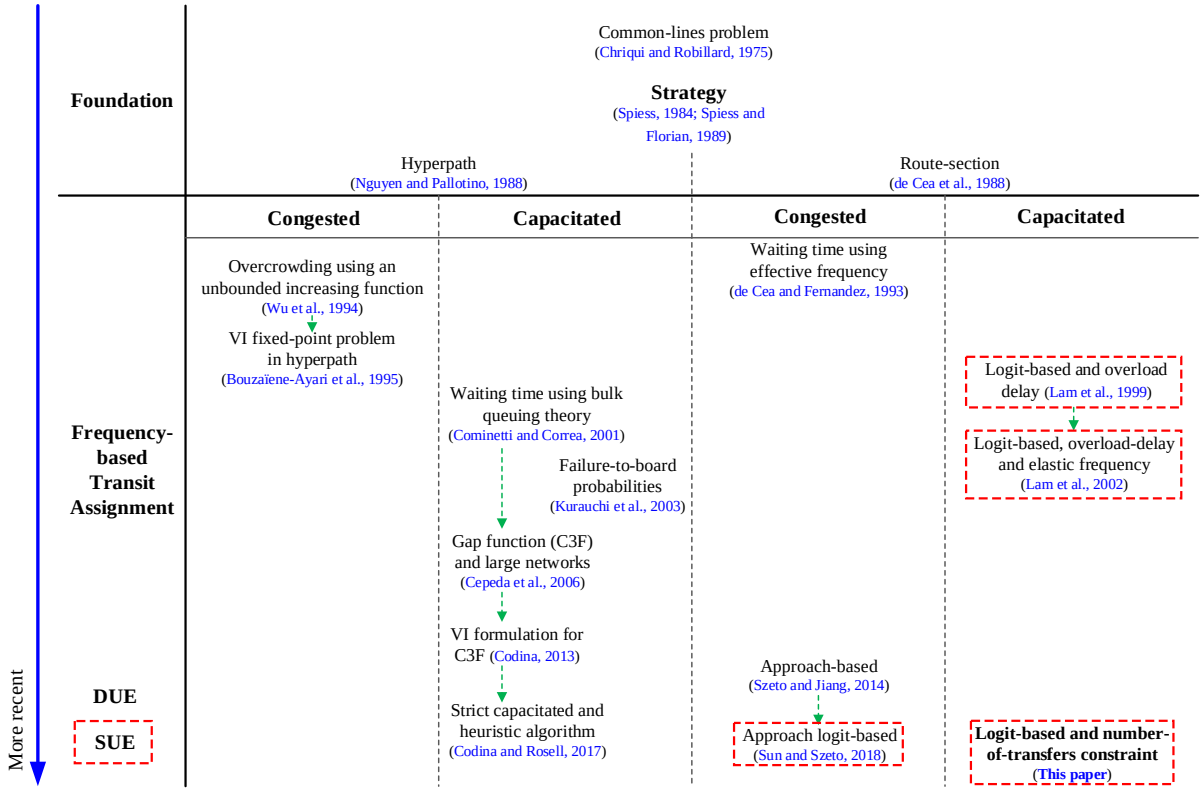
2.1 Introduction

Network equilibrium is a widely used method for transportation demand planning and management of both private cars and public transit. Traffic equilibrium (i.e., private cars) generally involves the assignment of vehicles to a road network, whereas transit equilibrium involves the assignment of passengers to a transit network. The core of network equilibrium is the path, which includes the path unit definition and path choice model. Differing from private car networks, the number of transfers is an important aspect of urban transit systems. The number of transfers, rather than the travel cost, will sometimes dictate a passenger's chosen path. This implies that passengers may choose a path with fewer or fewest transfers, not the lowest travel cost, which has been reported in numerous surveys. According to Stern (1996), approximately 58% of the transit agencies responding the survey in the United States believed that transit riders would avoid a transit path with more than one transfer for a single trip. An onboard survey of the Valley Metro (Arizona, U.S.A.) in 2015 showed that up to 96% of bus passengers chose routes with no more than one transfer (Maricopa Association of Governments and Valley Metro Transit System, 2015). Another onboard survey of Alameda-Contra Costa Transit (California, U.S.A.) also showed that more than 96% of passengers chose paths with at most one transfer for trip (Alameda-Contra Costa Transit District and Metropolitan Transportation Commission, 2018).

The number of transfers has been treated as a critical criterion in transit network design research because higher numbers of transfers discourage transit use (Farahani et al., 2013). Previous studies have classified the number-of-transfers issue into two categories: objectives and constraints relative to the number of transfers. The objectives category implies that the number of transfers is incorporated in the minimized objective function (Zhao and Ubaka, 2004; Zhao, 2006; Fan and Mumford, 2010; Szeto and Wu, 2011). The constraints category involves setting a maximum number of transfers for each origin-destination (OD) pair (trip) (Carrese and Gori, 2002; Guan et al., 2004; Mahdavi Moghaddam et al., 2019). Guihaire and Hao (2008) suggested that a passenger will switch to an alternative transport mode if the transit mode requires more than two transfers. Another perspective regarding transfers in the field of transit is a railway network equilibrium model that considers transfer reliability, which has been proposed to depict the route choice behavior of railway passengers (Shi et al., 2012).

Static transit equilibrium assignment problem can be addressed by frequency-based and schedule-based models. Frequency-based models are applied for high-frequency cases (e.g.,

urban areas), whereas schedule-based models are for low-frequency cases (e.g., suburban areas). Frequency-based models are also used for long-term planning and management, and schedule-based models are used for on-time operations. In this chapter, we focus on the frequency-based transit equilibrium assignment problem. Figure 2.1 provides an overview of the key studies in static frequency-based transit equilibrium.



Note: The green dashed lines with arrows represent direct and intrinsic connections between studies.

Figure 2.1 Overview of key static frequency-based transit equilibrium studies

Frequency-based transit network equilibrium, including deterministic user equilibrium and stochastic user equilibrium, has been well studied in the literature. To address the common lines problem proposed by Chriqui and Robillard (1975) in transit network, Spiess (1984) and Spiess and Florian (1989) proposed a strategy and mathematical optimization formulation to determine the optimal strategy. Two types of network representation were later proposed to better represent the strategy concept: hyperpath (Nguyen and Pallottino, 1988) and route section (de Cea et al., 1988). Early transit equilibrium studies focused on deterministic user equilibrium, which obeys Wardrop's equilibrium principle. In the branch of hyperpath-based network representation, Wu et al. (1994) proposed a hyperpath-based model for the congested transit equilibrium assignment problem. In their model, the cost functions of the waiting arcs

and in-vehicle arcs are asymmetric and dependent on the flows using an unbounded increasing function. Bouzaïene-Ayari et al. (1995) later extended the model of Wu et al. (1994) to the hyperpath flow space and reformulated it as a variational inequality (VI) fixed-point problem.

The route-section method was first proposed by de Cea and Fernández (1993) to handle the common-lines issue based on the concept of strategy and involved a new formulation for the transit assignment problem over a congested transit network. In their study, congestion effects were assumed to exist only at transit stops and the corresponding cost function was also asymmetric. Szeto and Jiang (2014) proposed an approach-based VI formulation that included the concept of route-section concept for the congested transit assignment problem and adopted the same cost function used in de Cea and Fernández (1993). The studies of Wu et al. (1994) and de Cea and Fernández (1993) are generally considered as milestones of the congested transit equilibrium assignment problem.

However, the congested model still generates flow results with overloaded line segments (i.e., transit vehicles). A transit vehicle cannot feasibly carry a passenger flow higher than its capacity. This issue has motivated some researchers to explore the capacitated transit equilibrium assignment problem. Cominetti and Correa (2001) proposed a frequency-based transit equilibrium model based on a congestion function obtained from bulk queuing theory. The capacity constraint issue is handled by considering effective frequency functions that vanish when the flow exceeds the line capacity. Cepeda et al. (2006) followed the work of Cominetti and Correa (2001) and proposed a computable gap function for a frequency-based assignment model with strict capacity constraints. They then developed the method of successive averages to solve the proposed model for a large-scale network. Codina et al. (2013) reformulated the model (which they called C3F) by Cepeda et al. (2006) as an equivalent variational inequality, and claimed that this reformulation could adapt the algorithm methods for VI problems to solve the C3F model. Codina and Rosell (2017) later incorporated the strict line capacities constraints into the VI reformulation of congested transit assignment problem in Codina et al. (2013) and proposed a heuristic algorithm to solve it. Specifically, they added inequality side constraints for the passenger flows of transit line segments to handle the capacity constraint. Kurauchi et al. (2003) presented a different approach for the capacity-constrained transit assignment problem using a Markov chain model. In their model, the cost function considers the passenger failure-to-board probability. The above studies of capacitated networks belong to the branch of hyperpath-based models.

The transit stochastic user equilibrium (SUE) problem has received increasing attention since the 1990s. Lam et al. (1999) proposed an SUE assignment model for congested transit

networks, where the travel cost function is independent of passenger flow and a capacity constraint is embedded to reflect the congestion. Lam et al. (2002) later extended this work by considering the elastic frequency of the capacity-constrained transit assignment problem and formulated it as a fixed-point problem. They assumed that the fleet size was fixed and the dwell time would affect the frequency. Both of these studies incorporated a hard capacity constraint for transit lines and applied an iterative balancing scheme for the solution algorithm, which could handle the problem with numerous inequality side constraints. To avoid the path enumeration and easily trace the path, link-based and approach-based methods (Sun and Szeto, 2018) later adopted the concept of the Dial's algorithm (Dial, 1971), which does not limit the number of path links. Transit SUE studies have thus mainly concentrated on route-section-based network representations.

To the best of our knowledge, the transit network equilibrium studies to date have not included the number of transfers as a consideration for path finding. Although this is reasonable for traffic equilibrium problems, transit users will avoid a path with an excessive number of transfers. For link-based or approach-based methods, it is difficult to identify the transfer configuration when calculating the split probability (or assigning the passenger flow) for each outgoing route section for a node (stop). Furthermore, the capacity issue is evidently more important in transit networks due to the number-of-transfers constraint. In fact, if the number-of-transfers constraint have been considered in previous studies, some paths that were deemed available would have become unavailable. This implies that some transit lines would have become congested or even overloaded, thus causing changes to the assigned flow pattern.

Based on the above discussion, the important task remains of exploring the transit stochastic equilibrium assignment problem with constraints on the number of transfers and capacity, especially for real-case transit networks. Thus, the objective of this chapter is to establish a strategy-based transit SUE model with a logit choice model that considers capacity and number-of-transfers constraints. The route-section-based representation is used for the transit network, and the path is set as a sequence of route sections. The paths consisting of route sections proposed here are in effect strategies (or simplified hyperpaths). Compared with existing path-based transit SUE models with capacity constraints (Lam et al. 1999; Lam et al., 2002), the path cost function in our model accounts for congestion effects at transfer transit stops via a Bureau of Public Roads (BPR)-like function, which is the same as in de Cea et al. (1993). We thus formulate the capacity-constrained transit SUE problem as a VI formulation with strict line capacity constraints enforced using asymmetric cost functions. We also consider the number-of-transfers constraint in the path finding, which is more realistic for choice

behavior in transit networks. An additional limitation of the two previous studies (Lam et al., 1999; Lam et al., 2002) is that they only conducted small numerical examples, which is insufficient for modern applications in large urban transit networks.

A path-set generation procedure with a number-of-transfers constraint is developed based on the *k*-shortest path algorithm by Yen (1971) to ensure suitable path finding for the choice behaviour in a transit network. Three rules are set in this chapter to generate a transit path. One is that the number of transfers for a transit path should have a threshold; other two are based on the route-section-based transit network representation to avoid unnecessary transfers and loop constructions.

The diagonalization method of de Cea et al. (1993) is used to solve the proposed asymmetric transit SUE model with capacity and number-of-transfers constraints. In each diagonalized iteration, the VI problem is reformulated as a convex mathematical programming formulation. The path-based linearization algorithm embedded with an iterative balancing scheme (Bell, 1995; Chen et al., 2005, 2009, 2010) and self-regulated averaging scheme (Liu et al., 2009) is then developed to solve the diagonalized problem. The iterative balancing scheme handles the issue of numerous inequality side constraints by iteratively updating the dual variables and calculating the corresponding primal variables.

2.2 Strategy-based Transit SUE Model with Capacity Constraints

2.2.1 Notation

This subchapter provides a list of the notation used in this chapter unless otherwise specified.

Sets	
N	set of transit stops
L	set of transit lines
E	set of transit line segments
S	set of route sections
A_s	set of attractive section line segments associated with route section s
OD	set of OD pairs
K^{od}	path set between OD pair od
Variables	
t_s	in-vehicle travel time of route section s
w_s	waiting time of route section s
c_s	total expected travel time of route section s
x_s^l	proportion of passengers choosing section line segment l associated with route section s
t_s^l	in-vehicle travel time of section line segment l associated with route section s
f_l	frequency of line l (vehicles/min)

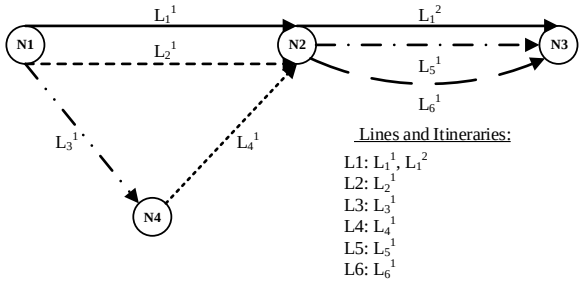
f_s	frequency of route section s (vehicles/min)
κ_l	capacity of line l (passengers/vehicle)
v_s	passenger flow on route section s
v_s^l	passenger flow on section line segment l associated with route section s
v_e	passenger flow on transit line segment e
h_k^{od}	passenger flow on path k between OD pair od
p_k^{od}	choosing probability of path k between OD pair od
q^{od}	passenger flow between OD pair od
Inputs	
θ	dispersion parameter measuring passengers' perception of travel cost
C_e	capacity on transit line segment e
q^{od}	passenger flow (demand) of OD pair od

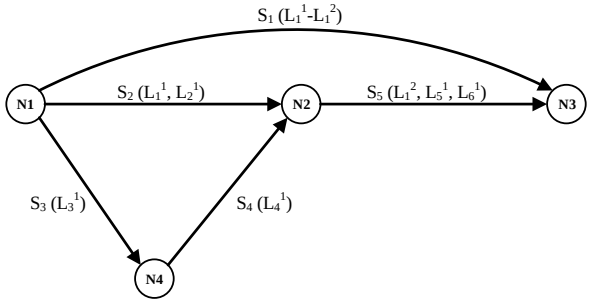
2.2.2 Route-section-based transit network revisit

Given an initial transit network $G(N, L)$, N represents the set of transit stops and L represents the set of transit lines. A line segment is any portion of a transit line between two not necessarily consecutive stops within its itinerary. Any pair of stops in the transit network might be serviced by different transit lines (i.e., common-lines problem; Chriqui and Robillard, 1975). The route-section-based transit network representation proposed by de Cea et al. (1988) and de Cea and Fernández (1993) to address the common-lines problem and a route consisting of one or more route-sections is considered a simplified version of a strategy or hyperpath. A major assumption of this method is that the passengers at a transit stop will be divided into different groups according to their following alighting stop. We denote this route-section-based network as $G(N, S)$ with a set of transit stops N and set of route sections S. A route section is the combination of some portions of all transit lines between two not necessarily consecutive nodes. Here, the portion of one transit line is named as a section line segment, which consists of one or more consecutive line segments in the transit line. With this route-section-based method, the number of links connecting any pair of stops in the transit network can only be one, which makes it easier to obtain solutions for the transit equilibrium problem. A transit path is defined as a sequence of route sections for passengers to travel between any two nodes in the transit network. It is assumed that the congestion on transit networks is concentrated at transit stops.

Here we use a small network from de Cea and Fernández (1993) to illustrate the route-section concept (Figure 2.2). The example network $G(N, L)$ consists of six transit lines: line 1 contains two transit line segments (L_1^1 and L_1^2) and the other five lines contains only one transit line segment. When using a route-section representation, the network $G(N, S)$ consists of five

route-sections. The transit line components of each route-section are shown in Figure 2.2b. We use route sections S_1 and S_2 for detailed illustration. Route section S_1 contains only one section line segment, $\{L_1^1 - L_1^2\}$, which includes two transit line segments, i.e. L_1^1 and L_1^2. Route section S_2 is a combination of some portions of transit lines L1 and L2. Route section S_2 actually consists of L_1^1 and L_2^1. Thus, there are three transit paths for the OD pair (N1, N3): transit path R1 consists of only one route section, S_1; transit path R2 contains two route sections S_2 and S_5 including one intermediate transfer node N2; and transit path R3 is composed of three route sections S_3, S_4, and S_5.





Paths	Transfer Nodes	Route Sections
R1	(N1, N3)	{L1¹-L1²} S1
R2	(N1, N2, N3)	{L1¹, L2¹}, {L1², L5¹, L6¹} S2 S5
R3	(N1, N4, N2, N3)	{L3¹}, {L4¹}, {L1², L5¹, L6¹} S3 S4 S5

(a) Transit network $G(N, L)$ using transit lines (b) Transit network $G(N, S)$ using route sections

Figure 2.2 Network representation of a transit network (de Cea and Fernández, 1993)

In an initial transit network $G(N, L)$, the information (e.g., frequency, in-vehicle travel time, capacity) is usually provided. Due to the route-section-based representation for transit network $G(N, S)$, the path is the route section sequence rather than the traditional line segment sequence. We must first specify the route section characteristics. The route section cost in this study consists of in-vehicle travel time, waiting time, and perceived congestion time.

Because the route section is a combination of attractive lines between two transfer stops, the in-vehicle travel time of the route section is the weighted summation of that of the corresponding lines. We first introduce the interim variable x_s^l, which represents the split probability of the section line segments in direct proportion to their frequencies on a route section as:

$$x_s^l = f \Big/ \sum_{l \in A_s} f_l \,, \forall l \in A_s, s \in S \tag{2.1}$$

The in-vehicle travel time of route section s can then be expressed as the weighted summation of the in-vehicle travel time of all of the attractive lines (section line segments) associated with route section s as:

$$t_s = \sum_{l \in A_s} t_s^l x_s^l, \forall s \in S \tag{2.2}$$

The waiting time of route section s is a function of its combined frequency f_s (i.e. product of parameter α and f_s). The combined frequency of a route section is the summation of the frequencies of all the section line segments. The waiting time expression is thus:

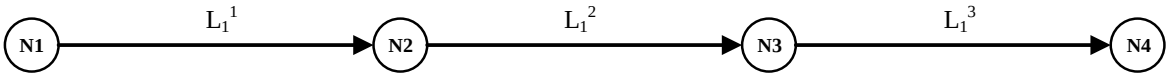
$$w_s = \frac{\alpha}{f_s} = \frac{\alpha}{\sum_{l \in A_s} f_l}, \forall s \in S \tag{2.3}$$

where α represents the vehicle headway distribution (Spiess and Florian, 1989). Specifically, $\alpha = 1$ indicates an exponential distribution and $\alpha = 0.5$ is a uniform distribution.

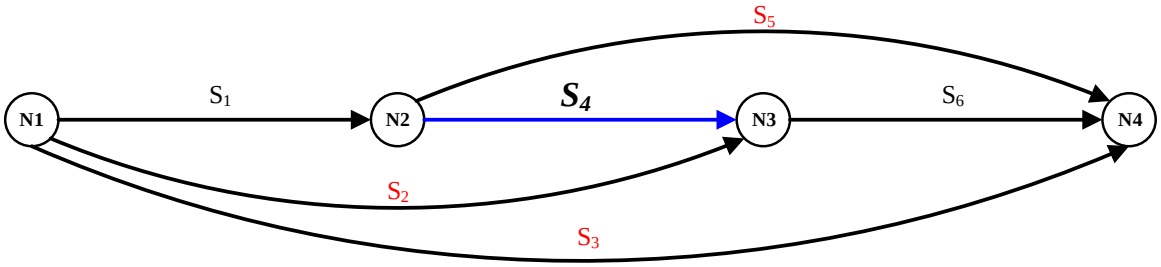
The perceived congestion time of the route section involves the additional waiting time due to vehicle congestion, which is a function of its own flow and that of its competing route sections. The flow on route section s is:

$$v_s = \sum_{od \in OD} \sum_{k \in K^{od}} a_{sk} h_k^{od}, \forall s \in S \tag{2.4}$$

The concept of competition is that one route section shares the same transit line segment with other route sections. Specifically, there are three main passenger groups competing with section s: (1) passengers boarding at $tail(s)$, and all other route sections that use transit lines are contained in route section s; (2) passengers boarding at all of the transit lines belonging to route section s at a stop prior to $tail(s)$ and alighting at $head(s)$; and (3) passengers boarding all of the transit lines belonging to route section s at a node prior to $tail(s)$ and alighting after $head(s)$. Here, $tail(s)$ and $head(s)$ represent the tail and head stops (nodes) of the route section, respectively. This also implies that a route section in $G(N, S)$ contains at most three groups of competing route sections, as illustrated in Figure 2.3. There are three kinds of competing route sections for route section S_4: the first kind is route section S_5, which shares the same tail node as S_4; the second is route section S_2, which belongs to the second group; and the last is route section S_3, which belongs to the third group.



(a) Transit network using a line and itinerary description



(b) Transit network using a route section description

Figure 2.3 Illustration of competing sections of a route section

The competing flow of route section s is then defined as the summation of the flow of section line segments on the route sections competing with route section s:

$$\tilde{v}_s = \sum_{n \neq s \in S} \delta_s^n \sum_{l \in A_s \cap A_n} v_{nl}, \forall s \in S \tag{2.5}$$

where the section line segment flow v_{sl} is determined based on the split probability on route section s as:

$$v_{sl} = v_s x_s^l, \forall s \in S, l \in A_s \tag{2.6}$$

Thus, the perceived congestion time of route section s can be expressed in a BPR-like function as:

$$\phi_s(\mathbf{v}) = \varphi_s \left(\frac{\lambda v_s + \varsigma \tilde{v}_s}{\sum_{l \in A_s} f_l \kappa_l} \right)^{\zeta}, \forall s \in S \tag{2.7}$$

where calibration parameters φ_s, λ, ς, and ζ are used to model the different effects of various flows on the perceived congestion time.

For route section s, the expected total travel time is given by:

$$c_s(\mathbf{v}) = t_s + w_s + \phi_s(\mathbf{v}) \tag{2.8}$$

As discussed above, the path is a sequence of route sections. The expected travel time of path k for OD pair od is therefore:

$$c_k^{od}(\mathbf{v}) = \sum_{s \in S} a_{sk} c_s(\mathbf{v}), \forall k \in K^{od}, od \in OD \tag{2.9}$$

2.2.3 Variational inequality formulation with capacity constraints

2.2.3.1 Flow conservation in a transit network

The relationship between the OD flows and passenger path flows is expressed as:

$$\sum_{k \in K^{od}} h_k^{od} = q^{od}, \forall od \in OD \tag{2.10}$$

Let $\Gamma = (\gamma_{es})$ denote the line segment-route section incidence matrix, which equals 1 if line segment e of line l lies on route section s, otherwise 0. The line segment flow expression is:

$$v_e = \sum_{s \in S} \gamma_{es} y_l^e x_s^l v_s \,, \forall e \in E \tag{2.11}$$

where $y_l^e = 1$ indicates that transit line segment e is on transit line l.

For simple expression, let $\bar{\gamma}_{es} = \gamma_{es} y_l^e x_s^l$ denote the proportion of passengers choosing line segment e of line l associated with route section s. The line segment flow v_e can thus be further expressed as:

$$v_e = \sum_{s \in S} \bar{\gamma}_{es} v_s \,, \forall e \in E \tag{2.12}$$

The relationship between line segment flows and path flows can be obtained as:

$$v_e = \sum_{od \in OD} \sum_{k \in K^{od}} \sum_{s \in S} \bar{\gamma}_{es} \, a_{sk} h_k^{od} \,, \forall e \in E \tag{2.13}$$

Referring to Lam et al. (1999) and Codina and Rosell (2017), the strict capacity constraint for the transit line segments can be described using the following inequality equation:

$$v_e \leq C_e, \forall e \in E \tag{2.14}$$

2.2.3.2 Logit-based stochastic user equilibrium condition

Based on Lam et al. (1999), the passenger overload delay is added to the link cost and path cost to reflect the strict transit line segment capacity constraint (explicit bounds on the flows on the transit line segments). The overload delays have been verified to be equivalent to the Lagrangian multipliers related to the line segment capacity constraints. In this chapter, we also consider the same transit line segment capacity constraints, and the corresponding overload delay for path k of OD pair od is expressed using d_k^{od}.

Thus, according to the definition of the logit route choice model (Sheffi, 1985), the logit-based transit SUE condition for a congested transit network is given as:

$$p_k^{od} = \frac{\exp\left(-\theta\left(c_k^{od} + d_k^{od}\right)\right)}{\sum_{k' \in K^{od}} \exp\left(-\theta\left(c_{k'}^{od} + d_{k'}^{od}\right)\right)}, \forall k \in K^{od}, od \in OD \tag{2.15}$$

$$h_k^{od} = p_k^{od} q^{od}, \forall k \in K^{od}, od \in OD \tag{2.16}$$

2.2.3.3 Variational inequality formulation

Due to the asymmetric cost function of a route section, we could not obtain a mathematical programming formulation for the transit SUE model with capacity and number-of-transfers constraints. Here, we instead propose a path-based variational inequality formulation:

[SUE-T-SC]

$$\sum_{od \in OD} \sum_{k \in K^{od}} \left(c_k^{od}(\mathbf{h}^*) + \frac{1}{\theta} \ln h_k^{od*}\right)\left(h_k^{od} - h_k^{od*}\right) \geq 0, \forall h_k^{od} \in \Omega \tag{2.17}$$

where h_k^{od*} is the optimal solution of the problem, and the feasible region Ω is defined as:

$$\sum_{k \in K^{od}} h_k^{od} = q^{od}, \forall od \in OD \tag{2.18}$$

$$v_e \leq C_e, \forall e \in E \tag{2.19}$$

$$h_k^{od} \geq 0, \forall k \in K^{od}, od \in OD \tag{2.20}$$

$$v_e = \sum_{s \in S} \bar{\gamma}_{es} \sum_{od \in OD} \sum_{k \in K^{od}} a_{sk} h_k^{od}, \forall e \in E \tag{2.21}$$

Let $\mathbf{F}(\mathbf{h}) = c_k^{od}(\mathbf{h}) + \frac{1}{\theta} \ln h_k^{od}$, and the VI model can be simplified to a standard form:

$$\mathbf{F}(\mathbf{h}^*)^T(\mathbf{h} - \mathbf{h}^*) \geq 0, \forall \mathbf{h} \in \Omega \tag{2.22}$$

It is assumed that the demand and capacity constraints allow the model to obtain the solutions. Some properties of the VI problem are provided hereinbelow. Proposition 2.1 states that the solution of a VI problem can be found, proposition 2.2 shows that the path flow estimation of the proposed model is distributed according to a multinomial logit model, and proposition 2.3 further verifies that the solution uniqueness.

Proposition 2.1. Assume that $c_s(\mathbf{v})$ is continuous and Ω is a compact and convex set; then there is at least one solution $\mathbf{h}^*$ of [SUE-T-SC].

Proof. The expression of the path cost is $c_k^{od}(\mathbf{h}) = \sum_{s\in S} a_{sk} c_s(\mathbf{h}) + \frac{1}{\theta}\ln h_k^{od}, \forall k \in K^{od}, od \in OD$, which is a positive continuous function of $\mathbf{h}$. According to Smith (1983), it can be proved. □

Proposition 2.2. The solution $\mathbf{h}$ of [SUE-T-SC] fulfils the MNL-based route choice model.

Proof. When μ^{od} and d_e are Lagrangian multipliers associated with constraints Eqs. (18) and (19), the first-order conditions (Karush-Kuhn-Tucker conditions) with respect to h_k^{od} for [SUE-T-SC] are (Proposition 1.3.4, Facchinei and Pang, 2003):

$$h_k^{od}\left(c_k^{od} + \frac{1}{\theta}\ln h_k^{od} - \sum_{e\in E} d_e \sum_{s\in S} \bar{\gamma}_{es} a_{sk} - \mu^{od}\right) = 0, \forall k \in K^{od}, od \in OD \tag{2.23}$$

$$c_k^{od} + \frac{1}{\theta}\ln h_k^{od} - \sum_{e\in E} d_e \sum_{s\in S} \bar{\gamma}_{es} a_{sk} - \mu^{od} \geq 0, \forall k \in K^{od}, od \in OD \tag{2.24}$$

$$h_k^{od} \geq 0, \forall k \in K^{od}, od \in OD \tag{2.25}$$

The decision variables follow $h_k^{od} > 0, \forall k \in K^{od}, od \in OD$, and thus we have:

$$c_k^{od} + \frac{1}{\theta}\ln h_k^{od} - \sum_{e\in E} d_e \sum_{s\in S} \bar{\gamma}_{es} a_{sk} - \mu^{od} = 0, \forall k \in K^{od}, od \in OD \tag{2.26}$$

The path flow expression for path k between OD pair od is then given as:

$$h_k^{od} = \exp\left(\theta\left(-c_k^{od} + \sum_{e\in E} d_e \sum_{s\in S} \bar{\gamma}_{es} a_{sk} + \mu^{od}\right)\right), \forall k \in K^{od}, od \in OD \tag{2.27}$$

The choosing probability of path k for OD pair od is thus:

$$p_k^{od} = \frac{h_k^{od}}{\sum_{k'\in K^{od}} h_{k'}^{od}} = \frac{\exp\left(-\theta\left(c_k^{od} - \sum_{e\in E} d_e \sum_{s\in S} \bar{\gamma}_{es} a_{sk}\right)\right)}{\sum_{k'\in K^{od}} \exp\left(-\theta\left(c_{k'}^{od} - \sum_{e\in E} d_e \sum_{s\in S} \bar{\gamma}_{es} a_{sk'}\right)\right)}, \tag{2.28}$$

$$\forall k \in K^{od}, od \in OD$$

The overload delay d_k^{od} for path k of OD pair od equals $-\sum_{e\in E} d_e \sum_{s\in S} \bar{\gamma}_{es} a_{sk}$ due to the negative Lagrangian multiplier d_e here. This completes the proof. □

Note that [SUE-T-SC] model might exist multiple solutions due to the asymmetric and non-separable route section cost function in Eq. (2.8). This means that the uniqueness of the solution for [SUE-T-SC] model cannot be guaranteed.

2.3 Transit Path Set Generation Procedure

The *k*-shortest path algorithm has been extensively used for path-set generation in private car network equilibrium studies. One of the best-known *k*-shortest path algorithms was proposed by Yen (1971) to find the *k*-shortest loopless paths including two parts: (1) identifying the first *k*-shortest path, and (2) finding all other ones. As shown in Figure 2.4, the two parts in Yen's algorithm use the Dijkstra's shortest path algorithm.

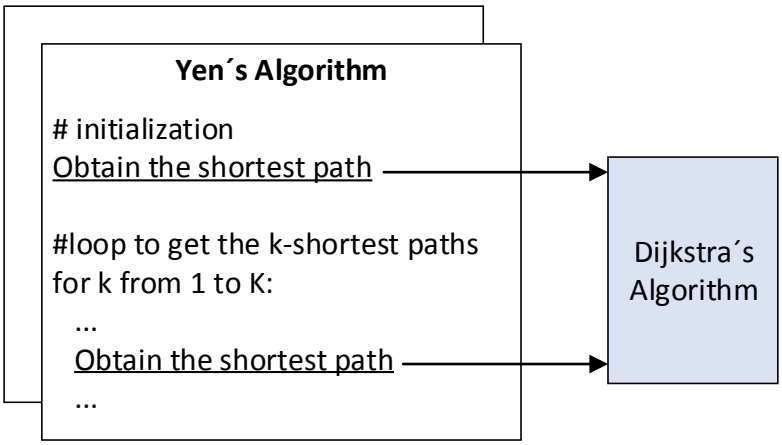


Figure 2.4 Framework of Yen's algorithm (Yen, 1971)

In this chapter, Yen's algorithm (Yen, 1971) is adopted to generate the k-shortest path for the transit path set. However, in the classical Dijkstra algorithm for path finding (Dijkstra, 1959), the selection of the subsequent node depends only on the current node itself. Realistically, in a transit network, passengers tend to avoid a path with a high number of transfers. Besides, two additional constraints exist for path finding using a route-section-based transit network representation: (1) two adjacent route-sections might share the same transit line, and (2) a potential selected subsequent stop might exist in the path which could be traced back using the current stop, its parent stop, and the origin stop. Therefore, to accommodate the transit traveler behavior and route-section-based transit network representation, three rules are embedded in the classical Dijkstra algorithm to find the suitable shortest transit path:

- Number-of-transfers constraint, which means the path has a maximum number of route sections or stops.
- Two adjacent route sections cannot contain the same transit line, which is to avoid unnecessary transfers.
- Any stop in this path can be included in any route section of this path, which is to avoid turn-back. This is necessary because a route section represents the linkage between two transfer stops, which means that the linkage might include other non-alighting stops.

However, these possible non-alighting stops would be the head nodes of other route sections.

As shown in Figure 2.5, the current stop is i and the two potential subsequent stops are j and k. For stop j, we first check the three rules above. If the number of transfers of the path constructed by $\{s, \dots, h, i, j\}$ meets the constraints and j is not the destination, stop j will not be selected. If route sections (h, i) and (i, j) contain the same transit line, stop j will be rejected. If stop j is included in any route section (as a non-alighting stop) of the path traced back from the current node i, stop j will be ignored. For stop k, all three of the rules above are satisfied and stop k will be selected as the subsequent stop of stop i.

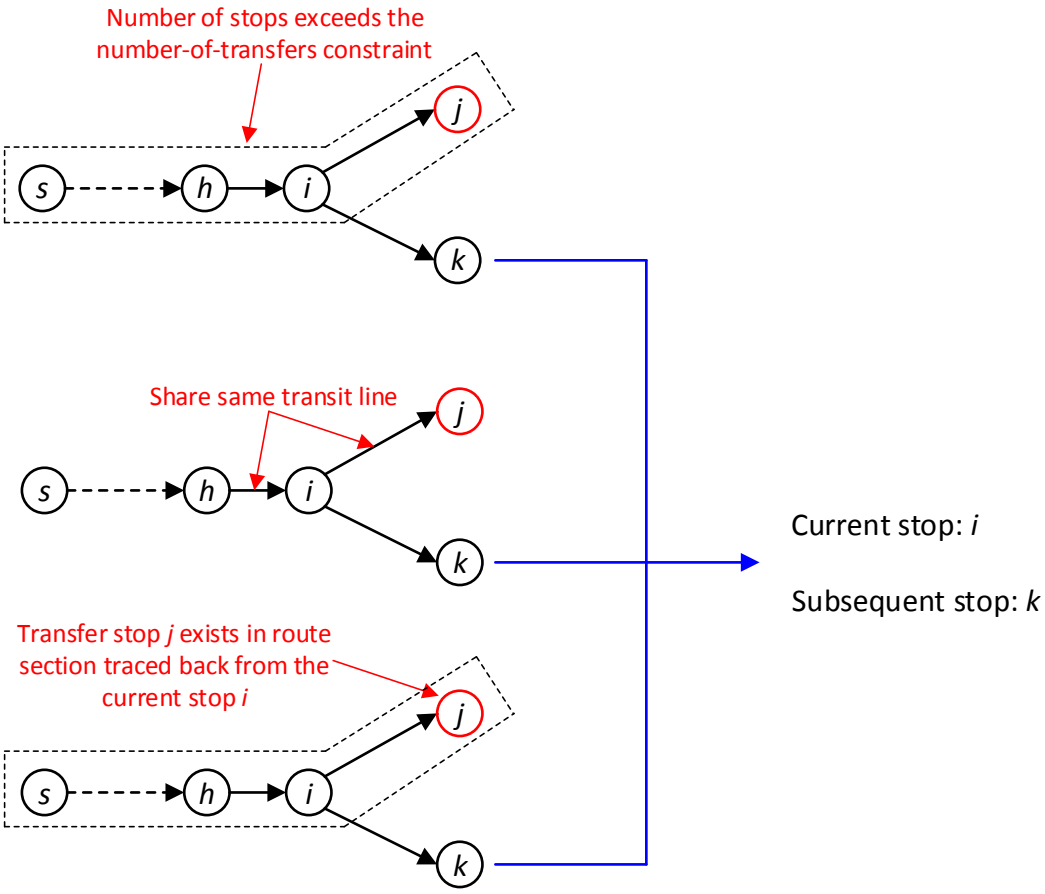


Figure 2.5 Selection of the subsequent transfer transit stop

Remarks:

(1) The predetermined number-of-transfers constraint might not suitable for all OD pairs in a real-case network, thus the number of transfers will increase until Yen's algorithm can generate the first shortest path for the OD pairs that requires more transfers.

(2) In the step of removing the links and nodes about the root path in Yen's algorithm, the nodes in the route section (link) and outgoing and incoming route sections related to these additional nodes are also removed. This is necessary because the section line segment in the route section might include more than one line segment, and the purpose is to avoid turn back.

(3) The root path must be considered in the step of calculating the spur path in Yen's algorithm using Dijkstra's algorithm. This is necessary because the route-section here represents the linkage between two transit transfer stops, which implies that the root and spur paths might contain the same transit line. Considering the root path when finding the spur path avoids an unnecessary transfer.

2.4 Solution Algorithm

Logit-based SUE has been well studied in traffic equilibrium studies and extensive efforts have been made to develop efficient algorithms to solve these models (Huang and Li, 2007; Yu et al., 2014; Zhou et al., 2014). However, the above algorithms cannot be adopted for SUE due to the asymmetric cost function. One of the most commonly used approaches to solve asymmetric network assignment problems is the diagonalization method because of its easy implementation (Florian, 1977; de Cea and Fernández, 1993).

2.4.1 Diagonalization of cost functions

Problem [SUE-T-SC] in Eq. (2.17) can be further expressed as:

$$\sum_{od\in OD}\sum_{k\in K^{od}}\left(c_k^{od}(\mathbf{h}^*)+\frac{1}{\theta}\ln h_k^{od*}\right)\left(h_k^{od}-h_k^{od*}\right)\geq 0,\forall h_k^{od}\in\Omega \tag{2.29}$$

We separate the $\left(c_k^{od}(\mathbf{h}^*)+\frac{1}{\theta}\ln h_k^{od*}\right)$ term into two parts, and multiply each part by $\left(h_k^{od}-h_k^{od*}\right)$ to obtain:

$$\sum_{od\in OD}\sum_{k\in K^{od}}c_k^{od}(\mathbf{h}^*)\left(h_k^{od}-h_k^{od*}\right)+\sum_{od\in OD}\sum_{k\in K^{od}}\frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od}-h_k^{od*}\right)\geq 0,\forall h_k^{od}\in\Omega \tag{2.30}$$

For Eq. (2.30), we substitute the route section cost function for the path cost function $c_k^{od}(\mathbf{h}^*)$. Eq. (2.30) is then be expressed as:

$$\sum_{od\in OD}\sum_{k\in K^{od}}\left(\sum_{s\in S}a_{sk}c_s(\mathbf{h}^*)\right)\left(h_k^{od}-h_k^{od*}\right)+\sum_{od\in OD}\sum_{k\in K^{od}}\frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od}-h_k^{od*}\right)\geq 0,\forall h_k^{od}\in\Omega \tag{2.31}$$

We can further calculate the route section flow based on the summation symbols for all paths for one OD pair and for all OD pairs:

$$\sum_{s\in S}\left(\sum_{od\in OD}\sum_{k\in K^{od}} a_{sk}\left(h_k^{od}-h_k^{od*}\right)\right)c_s(\mathbf{h}^*)+\sum_{od\in OD}\sum_{k\in K^{od}}\frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od}-h_k^{od*}\right) \geq 0, \forall h_k^{od}\in\Omega \tag{2.32}$$

Eq. (2.29) can then be simplified from Eq. (2.32) using the route section and path space as:

$$\sum_{s\in S}(v_s-v_s^*)c_s(\mathbf{h}^*)+\sum_{od\in OD}\sum_{k\in K^{od}}\frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od}-h_k^{od*}\right)\geq 0, \forall h_k^{od}\in\Omega \tag{2.33}$$

The route section cost in Eq. (2.33) can be expressed as a function of route-section flow, which can be obtained from the path flow. Eq. (2.29) is thus expressed as:

$$\sum_{s\in S}(v_s-v_s^*)c_s(\mathbf{v}^*)+\sum_{od\in OD}\sum_{k\in K^{od}}\frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od}-h_k^{od*}\right)\geq 0, \forall h_k^{od}\in\Omega \tag{2.34}$$

In each iteration, the $c_s(\mathbf{v})$ in Eq. (34) is diagonalized in the current solution, yielding a symmetric assignment problem:

$$\sum_{s\in S}(v_s-v_s^*)\hat{c}_s(\mathbf{v}^*)+\sum_{od\in OD}\sum_{k\in K^{od}}\frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od}-h_k^{od*}\right)\geq 0, \forall h_k^{od}\in\Omega \tag{2.35}$$

where $\hat{c}_s(\mathbf{v})$ is the diagonalized $c_s(\mathbf{v})$.

The symmetric assignment problem in Eq. (2.35) has an equivalent convex optimization formulation:

[SUE-T-SC-D]

$$\min\sum_{s\in S}\int_0^{v_s}\hat{c}_s(\omega)\,d\omega+\frac{1}{\theta}\sum_{od\in OD}\sum_{k\in K^{od}}h_k^{od}\left(\ln h_k^{od}-1\right) \tag{2.36}$$

subject to

$$\sum_{k\in K^{od}}h_k^{od}=q^{od}, \forall od\in OD \tag{2.37}$$

$$v_e\leq C_e, \forall e\in E \tag{2.38}$$

$$h_k^{od}\geq 0, \forall k\in K^{od}, od\in OD \tag{2.39}$$

$$v_e=\sum_{s\in S}\bar{Y}_{es}\sum_{od\in OD}\sum_{k\in K^{od}}a_{sk}h_k^{od}, \forall e\in E \tag{2.40}$$

2.4.2 Overall solution algorithm framework

The overall solution algorithm with a diagonalization concept for SUE-T-SC is summarized in Figure 2.6 based on the above discussion.

Step 1. Initialization.

Step 2. Diagonalization of the cost function. Diagonalize the cost function to obtain a symmetric assignment problem and proceed with the convex optimization model.

Step 3. Solve the subproblem. Develop the path-based partial linearization algorithm embedded with an iterative balancing scheme to obtain the solution.

Step 4. Stop test. If $\left\|\mathbf{h}^{i}-\mathbf{h}^{i-1}\right\| \leq \varepsilon$ (ε is a pre-set tolerance) or the maximum iteration number $Iter_{max}$ is achieved, stop and declare $(\mathbf{h}^{*},\mathbf{v}^{*}) \approx \left(\mathbf{h}^{i},\mathbf{v}^{i}\right)$. Otherwise, set $i=i+1$ and return to step 2.

Algorithm Solution algorithm for SUE-T-SC

1: Initialization. Identify an initial feasible solution $(\mathbf{h}^{1},\mathbf{v}^{1})$ and set $i=1$
2: **while** $i \leq Iter_{max}$ and $\left\|\mathbf{h}^{i}-\mathbf{h}^{i-1}\right\| > \varepsilon$, **do**
3: Diagonalize $c(\mathbf{v})$ as $\hat{c}$ at $\left(\mathbf{h}^{i-1},\mathbf{v}^{i-1}\right)$
4: Solve subproblem [SUE-T-SC-D]
$\min \sum_{s\in S}\int_{0}^{v_s}\hat{c}_s(\omega)\,d\omega+\frac{1}{\theta}\sum_{od\in OD}\sum_{k\in K^{od}}h_k^{od}\left(\ln h_k^{od}-1\right)$
subject to
$\sum_{k\in K^{od}}h_k^{od}=q^{od},\forall od\in OD,\ v_e\leq C_e,\forall e\in E,\ h_k^{od}\geq 0,\forall k\in K^{od},od\in OD,$
$v_e=\sum_{s\in S}\bar{\gamma}_{es}\sum_{od\in OD}\sum_{k\in K^{od}}a_{sk}h_k^{od},\forall e\in E.$
5: $i\leftarrow i+1,\left(\mathbf{h}^{i},\mathbf{v}^{i}\right)\leftarrow\left(\hat{\mathbf{h}}^{*},\hat{\mathbf{v}}^{*}\right)$
6: **end while**
7: **return** $(\mathbf{h}^{*},\mathbf{v}^{*}) \approx \left(\mathbf{h}^{i},\mathbf{v}^{i}\right)$

where $Iter_{max}$ is the maximum iteration number and $\varepsilon \in R_{+}$ is the tolerance.

Figure 2.6 Solution algorithm for SUE-T-SC

To solve the subproblem [SUE-T-SC-D], we develop the path-based partial linearization algorithm combined with a self-regulated averaging step-size scheme (Liu et al., 2009) embedded with an iterative balancing scheme. The iterative balancing scheme is used for direction finding with a given path set and fixed cost, and its core is to adjust the dual variables related to the side constraints and update the corresponding primal variables at each iteration, as shown in Eqs. (2.41)-(2.43):

- Adjust dual variables:

  For each transit line segment,

$$(d_e)^{j+1} = \min\left\{0, (d_e)^j + \frac{1}{\theta}\ln\frac{C_e}{(v_e)^j}\right\} \tag{2.41}$$

For each OD pair,

$$(\mu^{od})^{j+1} = (\mu^{od})^j + \frac{1}{\theta}\ln\frac{q^{od}}{\sum_{k\in K^{od}}\left(h_k^{od}\right)^j} \tag{2.42}$$

- Update primal variables:

$$\left(h_k^{od}\right)^{j+1} = \exp\left(\theta\left(-c_k^{od} + \sum_{e\in E}(d_e)^{j+1}\sum_{s\in S}\bar{\gamma}_{es}a_{sk} + (\mu^{od})^{j+1}\right)\right) \tag{2.43}$$

where d_e is the dual variable related to the capacity constraint, μ^{od} is the dual variable related to the OD demand flow conservation constraint, and j is the iteration number. The derivation of the adjustment factors for dual variables and a detailed scheme are given in Chen et al. (2009).

2.5 Numerical Experiments

This section presents three numerical examples. Example 1 is a modified transit network based on Spiess and Florian (1989) to demonstrate the issue of the number of transfers in a strategy or hyperpath applied in existing transit equilibrium studies. Example 2 is the Sioux Falls network, which is used to examine the features of the proposed model in details. Example 3 is the Winnipeg transit network in Canada (Ryu et al., 2017; INRO Consultants, 2020), which is used as a real case to demonstrate the applicability of the proposed model and solution algorithm in the real network. The parameter values in this section are set as follows: $\alpha = 1$, $\theta = 0.1$, $\varphi = 10$, and $\vartheta = \varsigma = \varpi = 1$.

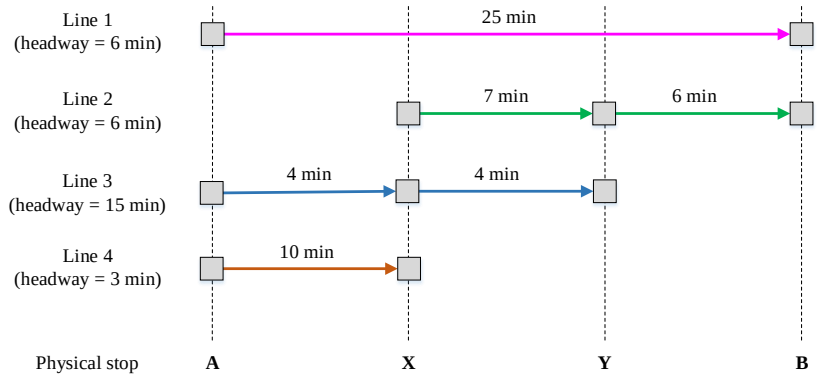
2.5.1 Small network

The example by Spiess and Florian (1989) has been widely used to illustrate the strategy-based concept to model transit passenger choice behavior. The optimal strategy in their example consists of two types of elementary paths: non-transfer and one-transfer paths. However, in a random transit network, we cannot rule out that an optimal strategy (shortest path) will contain an elementary path exceeding a maximum number of transfers. Due to the complex hyperpath structure, we use a network for illustration. In the field of route-section network representation, a transit path is a sequence of route sections. It is possible that the

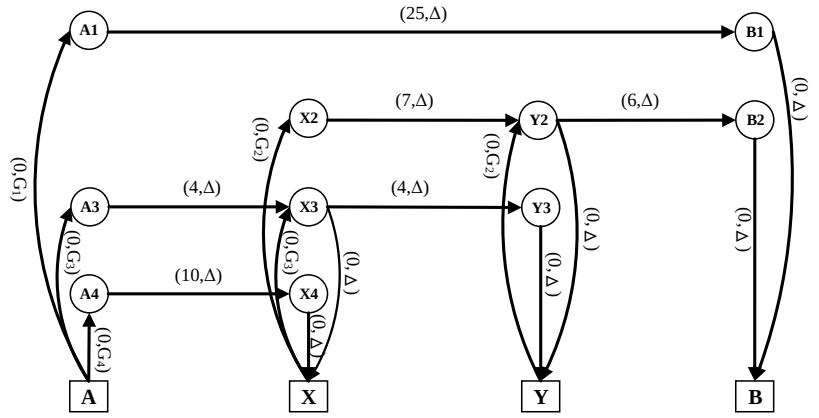
number of transfers in transit a path exceeds a predefined value, which is similar to the path in road network with a high number of crossings and therefore not illustrated here.

The illustrative network used in this study is modified from that in Spiess and Florian (1989) with modifications to change the stop sequences of lines 2, 3 and 4. Line 2 consists of stops X, Y, and B; line 3 contains stops A, X, Y; and line 4 is from A to X. This network contains only one OD pair from A to B with a travel demand as 1. Other transit line characteristics are the same as those in Spiess and Florian (1989) and shown in Figure 2.7a. The extended and simplified transit networks are also shown in Figure 2.7.

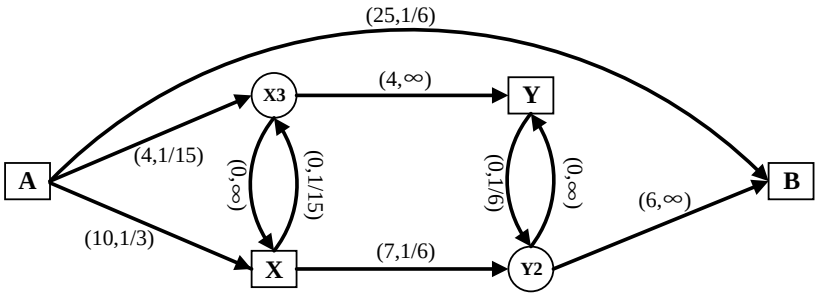
The optimal strategy from A to B is shown in Figure 2.8a when the in-vehicle travel time for line 4 is 10 ($t_{AX} = 10$). The dark black solid lines and letters represent the optimal strategy and the light-colored lines and letters are for unused parts of the transit network. The passenger flows on each transit line (segment) are also displayed. The maximum number of transfers is one here, which is the same as in Spiess and Florian (1989). When we change t_{AX} to 5, the optimal strategy (Figure 2.8b) differs substantially from that in Figure 2.8a. This optimal strategy actually contains three elementary paths: (1) only using line 1; (2) using lines 3 and 2 transferring at stop Y; and (3) using lines 4, 3, and 2 transferring at stops X and Y. The elementary path with two transfers is highlighted in red. Specifically, 0.17/0.59 of the demand alights at stop X and transfers to Line 3, whereas 0.42/0.59 takes line 2 directly to destination B. However, the former part (0.17/0.59) will transfer to line 2 after alighting from line 3 at stop Y.



(a) Initial transit network



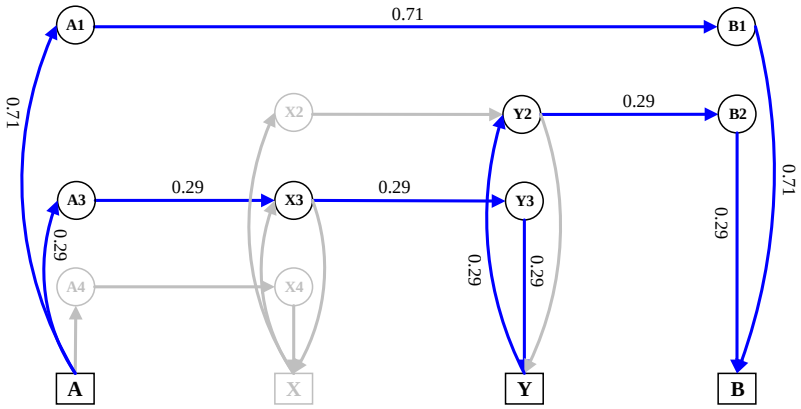
(b) Extended transit network



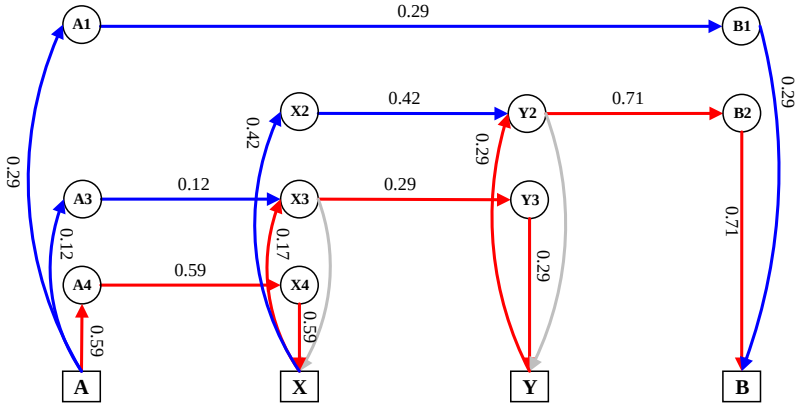
(c) Simplified transit network

Figure 2.7 Illustration of a small transit network

When conducting transit network equilibrium analysis without considering the number-of-transfers constraint, the hyperpath sets might contain one or more hyperpaths with several transfers, which would not become the choice for transit passengers in reality. For example, for the small transit network in Figure 2.7, passengers will usually not choose a path with more than one transfer. However, a hyperpath with two or more transfers might be taken into account for equilibrium analysis. This is the common limitation of existing transit equilibrium studies.



(a) Optimal strategy when $t_{AX} = 10$



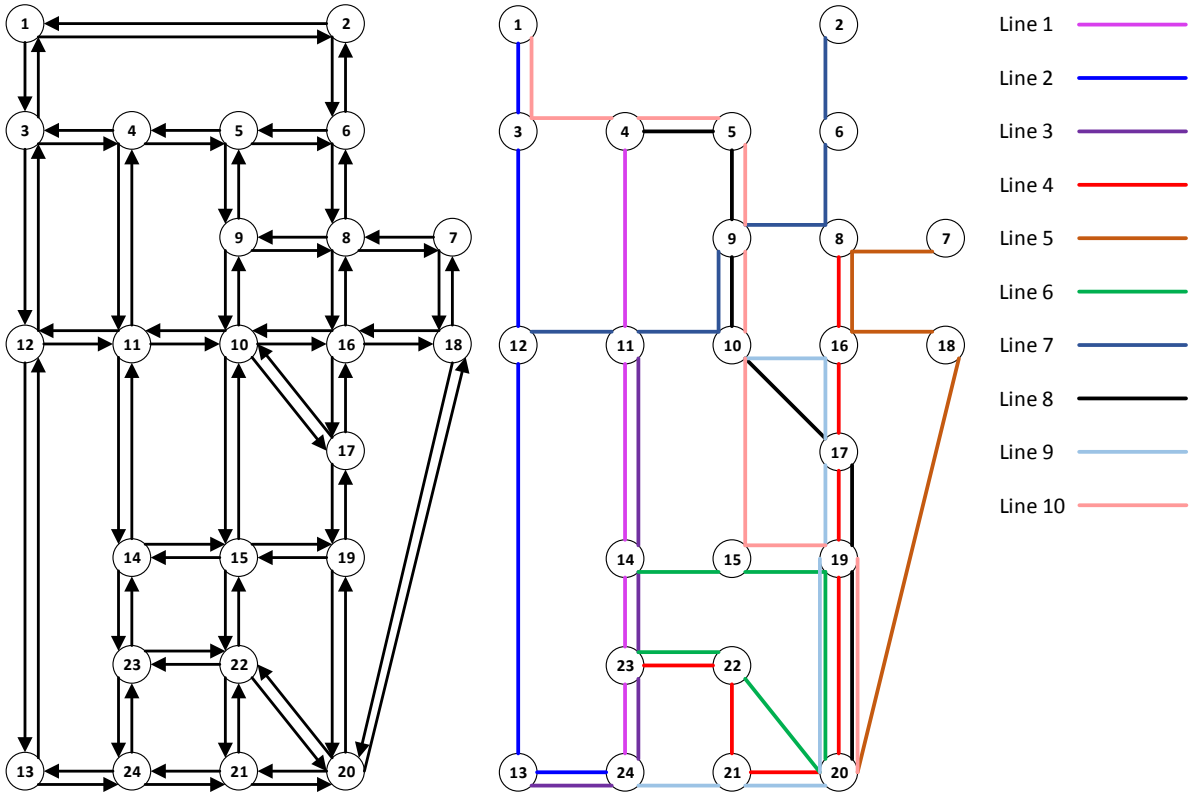
(b) Optimal strategy when $t_{AX} = 5$

Figure 2.8 Result of optimal strategy (shortest hyperpath) with different numbers of transfers

2.5.2 Sioux Falls network

2.5.2.1 Network setting

The highway network of Sioux Falls (http://www.bgu.ac.il/~bargera/tntp/) is shown in Figure 2.9 and contains 76 directed links and 24 nodes on which the itineraries of 10 transit lines (i.e., 20 itineraries) are defined modified from Sun and Szeto (2018). All of the in-vehicle movements on a given highway link are assumed to have identical travel times. Table 2.1 illustrates the frequencies, capacities and stop sequences of the lines. The in-vehicle times of transit vehicles on the highway network are assumed to be identical to those of private cars. There are 32 OD pairs (with positive demands) and the known (true) OD matrix is listed in Table 2.2. For the Sioux Falls network, the predefined path set is obtained with a large *k* value of 30 in the path set generation procedure as presented in Section 2.3.



(a) Road network (b) Transit network

Figure 2.9 Sioux-Falls network

Table 2.1 Transit line data for Sioux Falls network

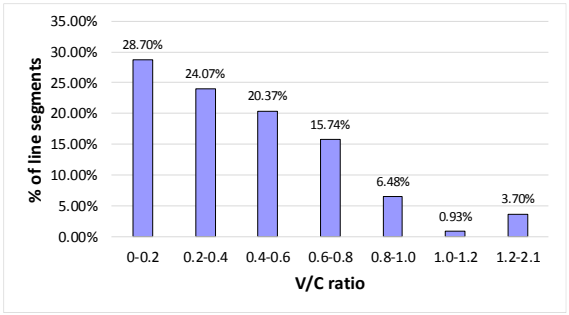
Line	Line ID	Frequency (vehicle/hour)	Capacity (pax/vehicle)	Stop sequence
1	1	10	50	4 11 14 23 24
	2			24 23 14 11 4
2	3	10	50	1 3 12 13 24
	4			24 13 12 3 1
3	5	10	50	11 14 23 24 13
	6			13 24 23 14 11
4	7	12	50	8 16 17 19 20 21 22 23
	8			23 22 21 20 19 17 16 8
5	9	10	50	7 8 16 18 20
	10			20 18 16 8 7
6	11	10	50	14 15 19 20 22 23
	12			23 22 20 19 15 14
7	13	20	50	2 6 8 9 10 11 12
	14			12 11 10 9 8 6 2
8	15	20	50	4 5 9 10 17 19 20
	16			20 19 17 10 9 5 4
9	17	20	50	10 16 17 19 20 21 24
	18			24 21 20 19 17 16 10
10	19	20	50	1 3 4 5 9 10 15 19 20
	20			20 19 15 10 9 5 4 3 1

Table 2.2 True demand matrix for Sioux Falls

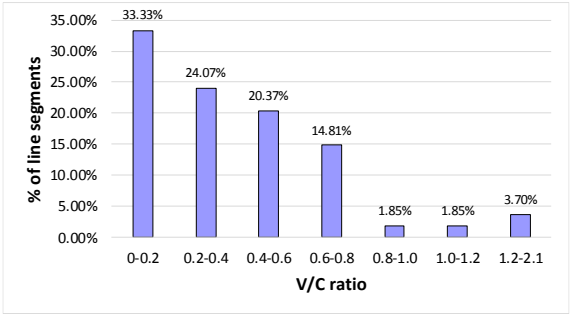
OD pair	Demand (pax/hour)	OD pair	Demand (pax/hour)	OD pair	Demand (pax/hour)	OD pair	Demand (pax/hour)
(1 - 13)	160	(3- 13)	160	(13 - 1)	160	(21 - 1)	160
(1 - 20)	160	(3 - 20)	160	(13 - 2)	160	(21 - 2)	160
(1 - 21)	160	(3 - 21)	160	(13 - 3)	160	(21 - 3)	160
(1 - 24)	160	(3 - 24)	160	(13 - 4)	160	(21 - 4)	160
(2 - 13)	160	(4 - 13)	160	(20 - 1)	160	(24 - 1)	160
(2 - 20)	160	(4 - 20)	160	(20 - 2)	160	(24 - 2)	160
(2 - 21)	160	(4 - 21)	160	(20 - 3)	160	(24 - 3)	160
(2 - 24)	160	(4 - 24)	160	(20 - 4)	160	(24 - 4)	160

2.5.2.2 Effect of number of transfers constraint

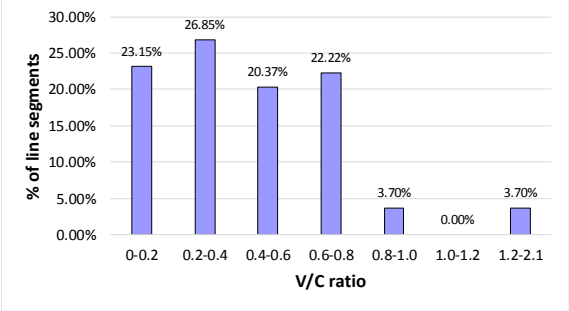
This section explores the effect of the number-of-transfers constraint on the flow patterns without capacity constraints. Here we set the maximum number of transfers to infinite (no constraint), 1, 2, and 3 for a transit path, represented by cases I, II, III, and IV. In general, we can see that the flow patterns of the transit line segments with different number-of-transfers constraints show large differences in their volume to capacity (V/C) ratios. To more clearly describe the difference, we set the case with the no-transfers constraint (case I) as the benchmark as shown in Figure 2.10a. Compared with case I, case II contains more transit line segments with less congestion (33.33% vs. 28.70%) and fewer line segments with a V/C ratio over 0.8 (7.4% vs. 11.30%). However, these two cases are similar when the V/C ratio is between 0.2 and 0.8. Compared with case II, case III differs from case I more markedly: fewer less-congested transit line segments (23.15% vs. 28.70%) and more line segments with V/C ratios between 0.2 and 0.8 (69.44% vs. 60.18%). Case IV differs slightly from case I with a V/C ratio of 0-0.8. One important observation is that cases II, III and IV contain fewer line segments with V/C ratios over 0.8 than in case I. One apparent reason is that the number-of-transfers constraint changes the components of the transit path set for these four cases. This can be further explained by the fact that without a number-of-transfers constraint, some transit line segments are repeatedly used by many paths even though some paths require more transfers. After adding the number-of transfers-constraint, these paths with over-repeated line segments and more transfers are not taken into account.



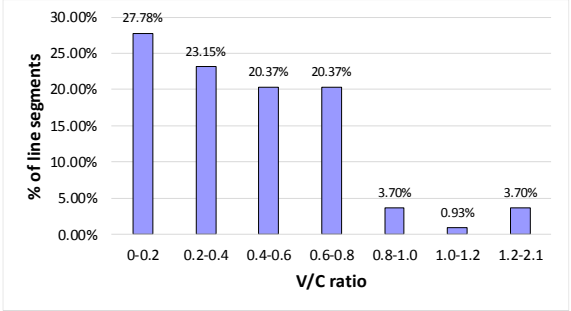
(a) No number-of-transfers constraint (case I)



(b) Number of transfers: 1 (case II)



(c) Number of transfers: 2 (case III)



(d) Number of transfers: 3 (case IV)

Figure 2.10 Distribution of volume to capacity (V/C) ratio of transit line segments with different number-of-transfers constraints

2.5.2.3 Effect of capacity constraint

This section investigates the effect of capacity constraints on flow patterns. Without a loss of generality, the maximum number of transfers is set as 2. We assume that the travel demand is suitable for obtaining a feasible solution given the available capacity constraints on all transit line segments.

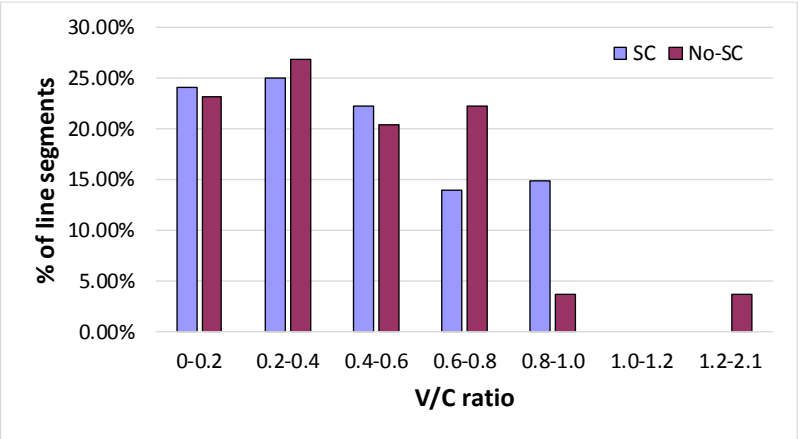
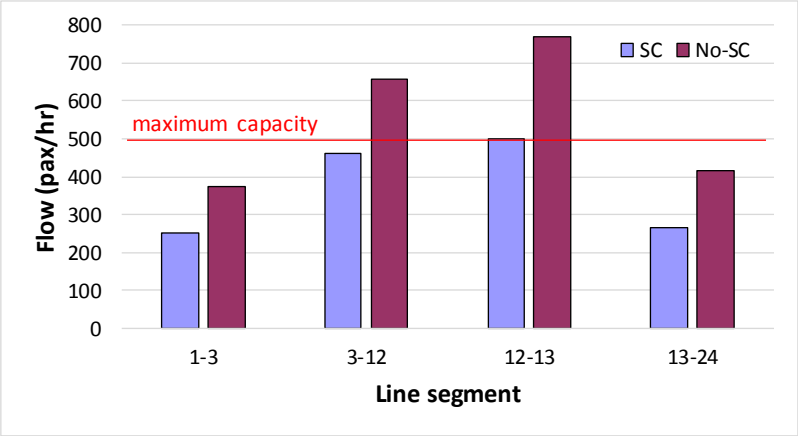
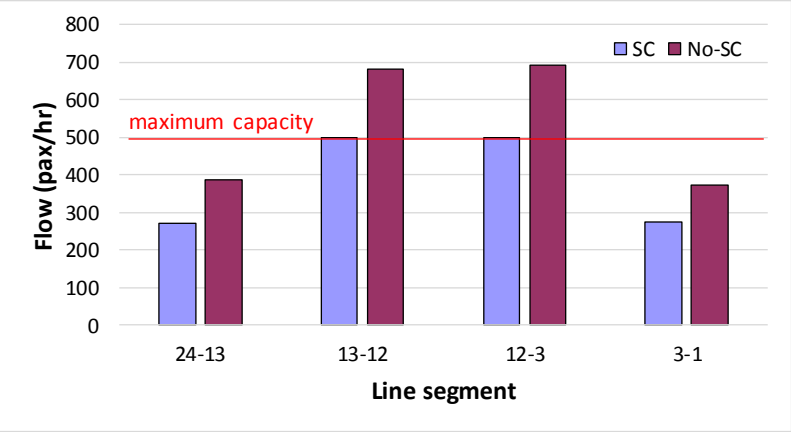


Figure 2.11 Volume to capacity (V/C) ratio of transit line segments with (SC) and without (No-SC) capacity constraints



(a) Downward direction: stop 1 to 24



(b) Upward direction: stop 24 to 1

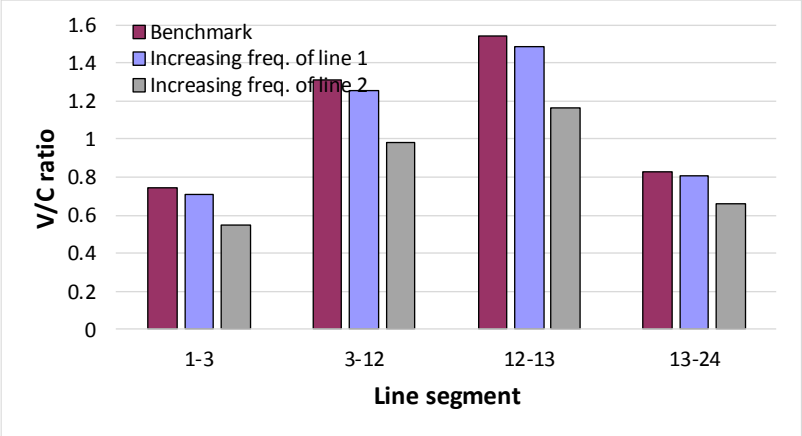
Figure 2.12 Flow on transit line 2 with (SC) and without (No-SC) capacity constraints

Figure 2.11 compares the V/C ratio of transit line segments with and without capacity constraints (referred to as SC and No-SC, respectively). Both scenarios generate similar results in terms of V/C ratio sections of [0,0.2], (0.2,0.4], and (0.4,0.6]. However, different results are obtained for line segments with high V/C ratios. The SC-case contains only 12.96% of line segments with a ratio of (0.6,0.8], whereas the No-SC-case contains 21.30%. In contrast, there are 15.74% of line segments with a ratio of (0.8,1.0] for the SC-case compared with only 3.70% for the No-SC-case. However, the No-SC-case generates overflow line segments (3.7%), whereas the SC-case at most contains line segments reaching the capacity.

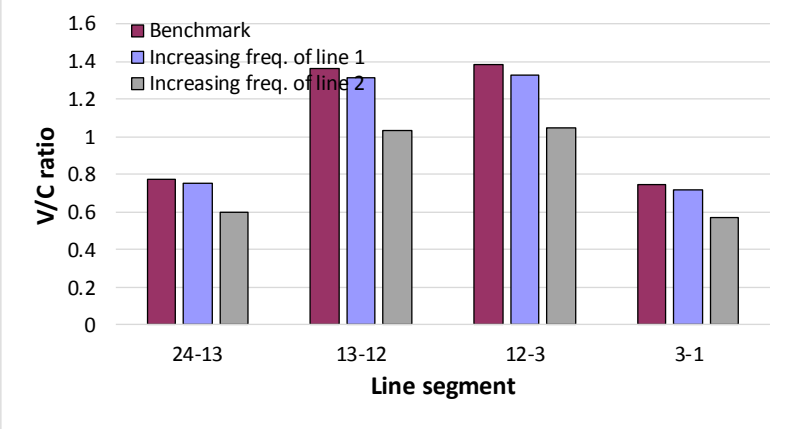
Figure 2.12 illustrates the flows on different line segments of transit line 2 with and without capacity constraints. For both directions of transit line 2, the segments between stops 3 and 12 and stops 12 and 13 carry higher passenger flows than those without line capacity constraints. After embedding the capacity constraints, the equilibrium flow of all line segments

of transit line 2 does not exceed their capacity. Interestingly, the flow of line segment 3-12 decreases to below its capacity with the implementation of a capacity constraint, whereas the other overloaded line segments continue to reach their own capacities. The flow patterns between stops 1 and 3 and stops 13 and 24 are nearly the same with and without capacity constraints. This might be explained by the fact that lower passenger flow is assigned to the transit paths with overloaded line segments due to the capacity constraints. These transit paths do not use other uncongested line segments of transit line 2.

2.5.2.4 Effect of increasing transit line frequency



(a) Downward direction: stop 1 to 24

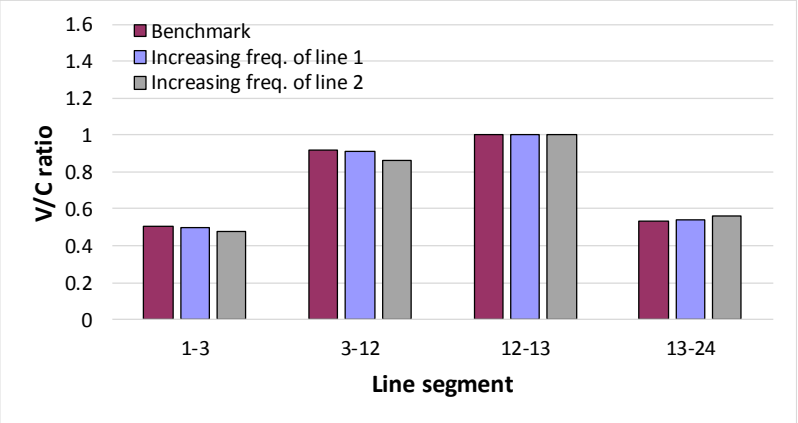


(b) Upward direction: stop 24 to 1

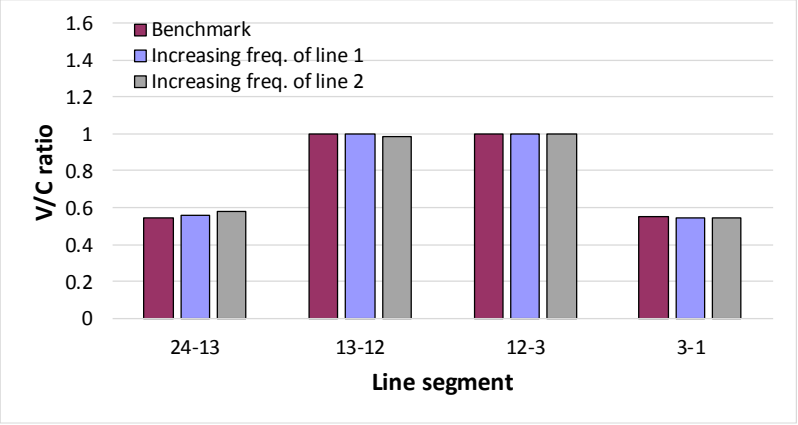
Figure 2.13 Effect of increasing transit line frequency on the flow of saturated line 2 without capacity constraints

This section examines the effect of increasing transit line frequency on flow patterns based on the overflow observation. Without loss of generality, the maximum number of transfers is

set to 2. Increasing transit line frequency essentially increases the transit line capacity and relieves the transit line congestion. Two strategies are conducted in this section: (1) increase the frequency of the saturated transit line, i.e. line 2, and (2) increase the frequency of the unsaturated transit line, i.e. line 1. The purpose of strategy (1) is to make the saturated transit line carry more flow, and that of strategy (2) is to improve the service of the other transit lines and help share the flow.



(a) Downward direction: stop 1 to 24



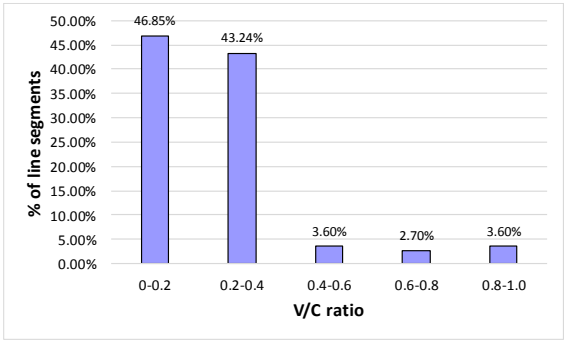
(b) Upward direction: stop 24 to 1

Figure 2.14 Effect of increasing transit line frequency on the flow of saturated line 2 with capacity constraints

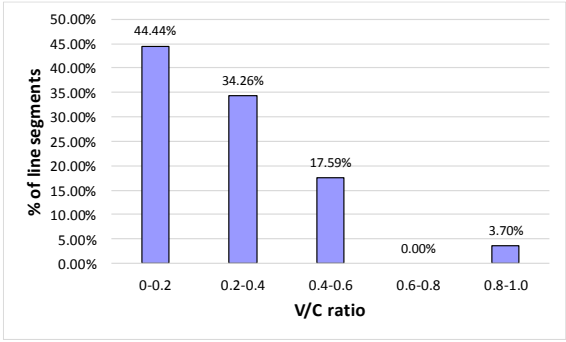
Figure 2.13 shows that even though both strategies improve the level of service of transit line 2 without capacity constraints, the overflow line segments still maintain overflow. However, Figure 2.14 shows that neither strategy has an effect on the service level, which means that the results of both strategies are essentially the same as that of initial frequency. This is meaningful for practical transit planning. If capacity constraints are not incorporated in strategy evaluation, the results might appear good but in fact be unreasonable because a transit

vehicle cannot carry a passenger flow above its capacity. Overall, the flow results without capacity constraints might be overestimated compared with those with capacity constraints.

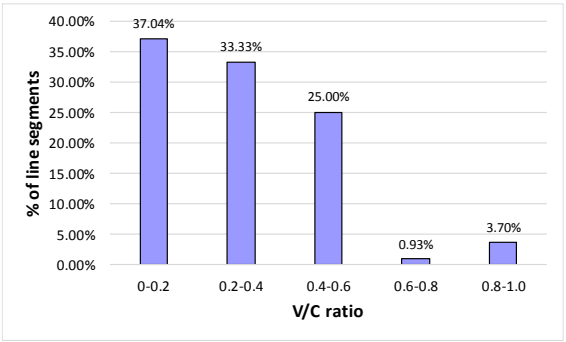
2.5.2.5 Effect of demand level



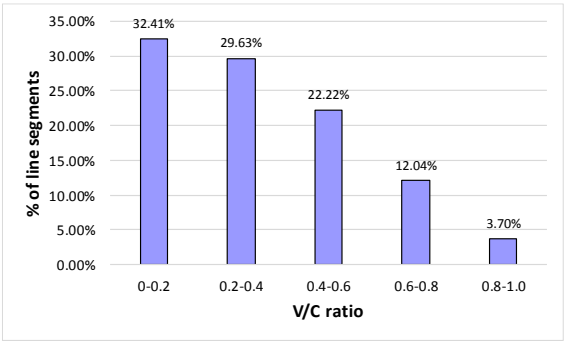
(a) demand: 0.5 × q



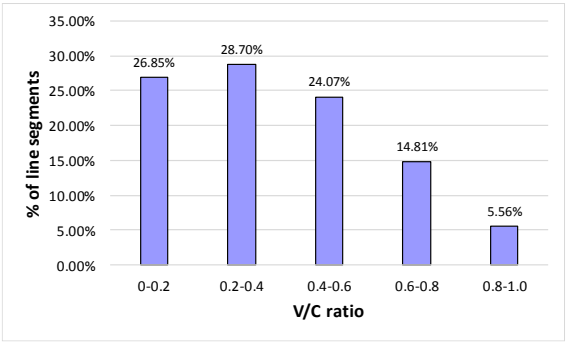
(b) demand: 0.6 × q



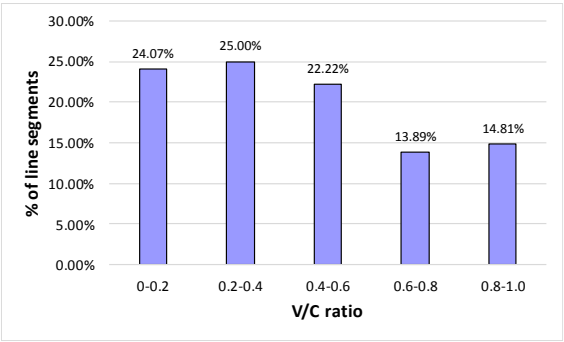
(c) demand: 0.7 × q



(d) demand: 0.8 × q



(e) demand: 0.9 × q



(f) demand: 1.0 × q

Figure 2.15 Distribution of volume to capacity (V/C) ratio of transit line segments with different demand levels

This section mainly explores how the demand levels affect the flow patterns with capacity constraints. To illustrate the effect, a demand adjustment factor, from 0.5 to 1.0 with the interval of 0.1, is used. The reference demand q is the same at that in Table 2.2. Without loss of

generality, the maximum number of transfers is set to 2. The result of flow patterns is shown in Figure 2.15.

As expected, no matter how demand levels change, the V/C ratio of all transit line segments is less than or equal to one due to the strict capacity constraints. Besides, with the demand increases (i.e. larger demand adjustment factor), the number of transit line segments with V/C ratio larger than 0.4 increases. This is because with more demand, more flow will be assigned to the transit network, and the network will become more congested. Moreover, the number of transit line segments with larger V/C ratio (i.e. 0.8-1.0) also increases as the demand increases. This can be explained from two perspectives: (i) more demand means more congested, and (ii) the strict capacity constraints make the saturated line segments not carry more flow, which will make other unsaturated line segments become more congested.

2.5.3 Winnipeg network

The Winnipeg transit network in Figure 2.16 is used to illustrate the features of the proposed model and performance of the developed solution algorithm. The network is extracted from Emme V4.3.2 and consists of 130 transit lines, 4187 transit line segments, and 924 transit stops. The number of origins is 106 and that of OD pairs is 5303. To connect the origins and transit stops, the network also contains 803 walking segments. After conducting the route-section network construction, the network consists of 44,408 route sections, in which each walking segment is interpreted as a special case of a transit route section with a cost of 0. The solution algorithm is implemented in Microsoft Visual Studio 2015 and run on a 2.7-GHz processor with 20.00 GB of RAM. Without loss of generality, the maximum number of transfers here is set to 3, and the k value in the transit path-set generation procedure is 20.

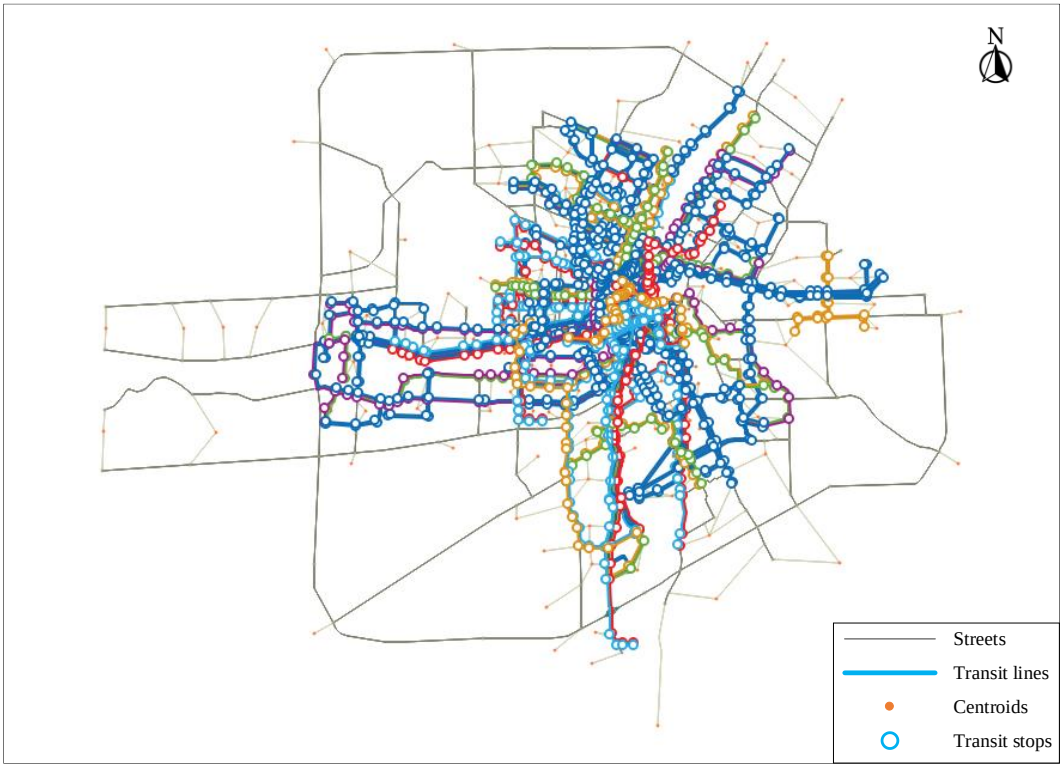


Figure 2.16 Winnipeg transit network: different colors denote different lines

2.5.3.1 Convergence characteristics

The root mean square error of the transit path flows between two adjacent iterations is adopted to represent the convergence, which is:

$$RMSE = \sqrt{\frac{\sum_{od \in OD} \sum_{k \in K^{od}} \left(h_k^{od,i} - h_k^{od,i-1}\right)^2}{|K|}} \tag{2.54}$$

Figure 2.17 shows the convergence of the proposed solution algorithm, including each outer iteration and each inner iteration of outer iterations 1 and 12. Both the outer and inner iterations efficiently reach a value below 10^{-8}, which means the proposed solution algorithm is capable of solving real-case transit network scenarios. The convergence of the dual variables of the two selected transit line segments is also shown in Figure 2.17.

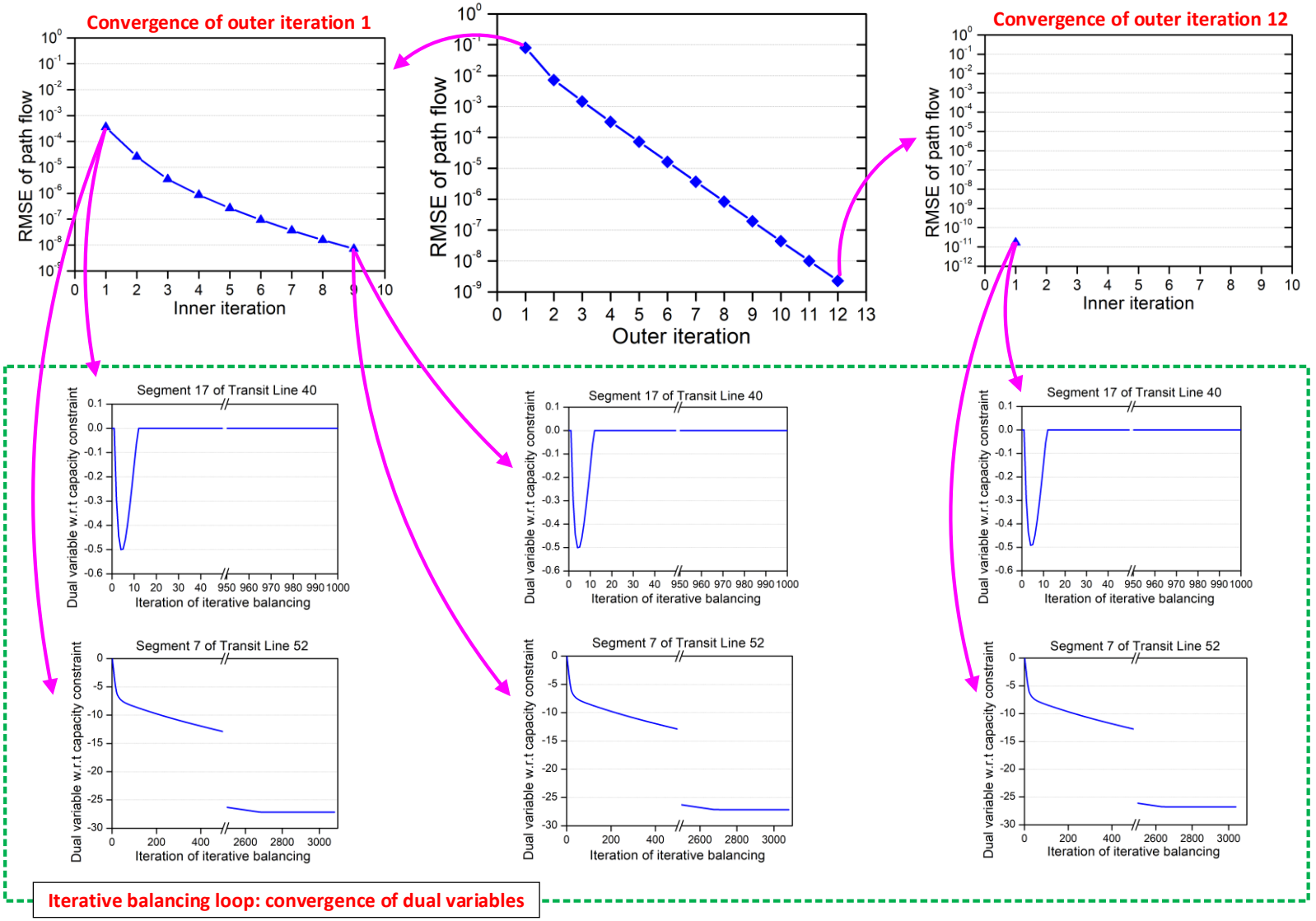
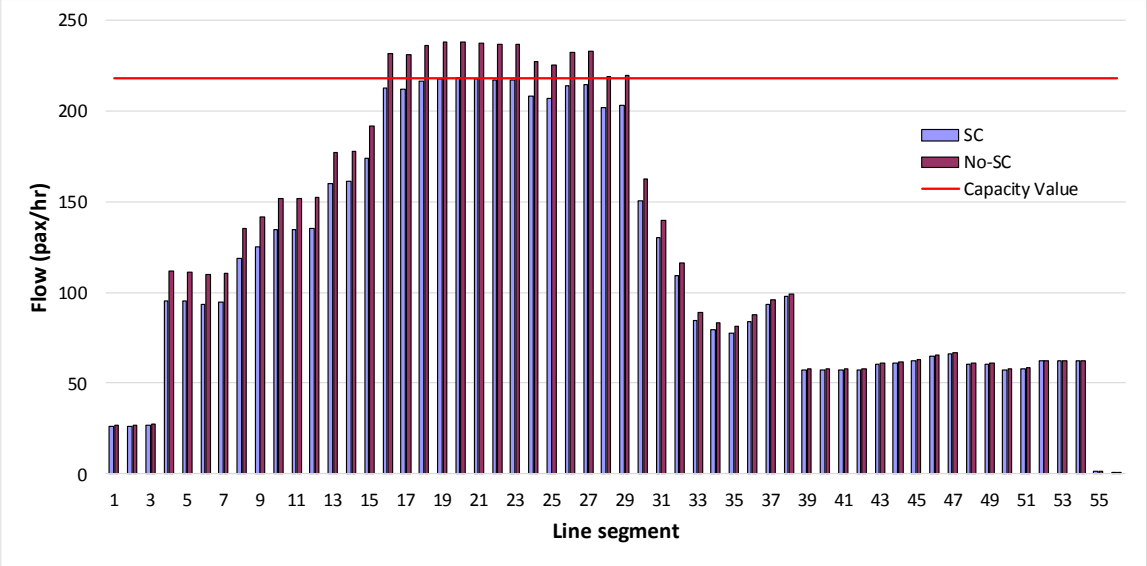
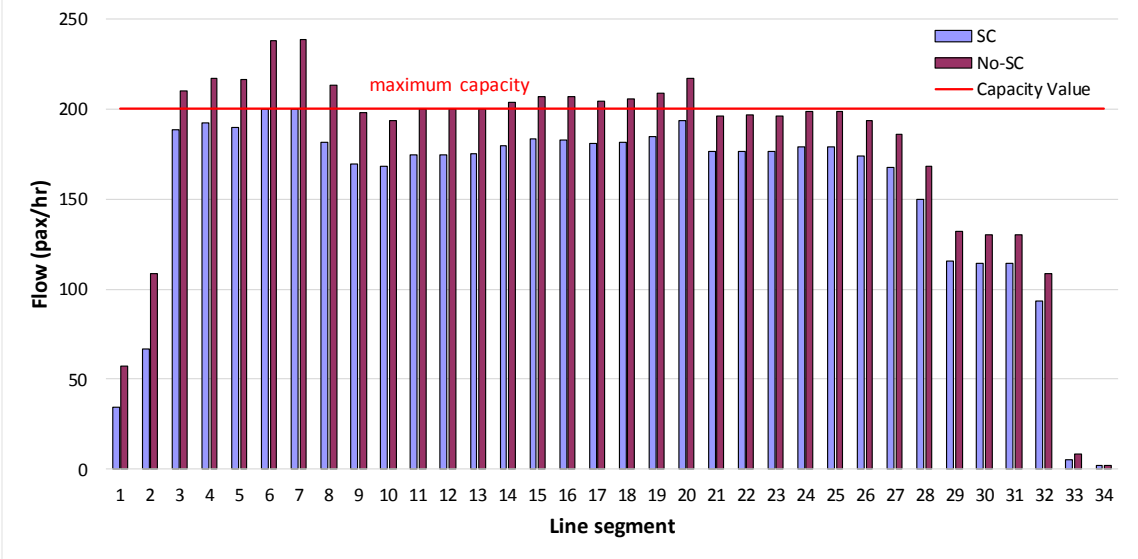


Figure 2.17 Convergence characteristics of the proposed solution algorithm

2.5.3.2 Flow patterns



(a) Line 40



(b) Line 52

Figure 2.18 Flow on two transit lines between with (SC) and without (No-SC) capacity constraints

Figure 2.18 shows that the flow of some line segments of lines 40 and 52 exceeds the maximum capacity when incorporating capacity constraints. After adding the capacity constraints, the flow of the transit line segments only reaches the capacity. A comparison of Figure 2.18a and 2.18b show that not all of the overflow line segments reach capacity when

capacity constraints are incorporated. Some overflow line segments in Figure 2.18a contain less flow than capacity after incorporating capacity constraints. This illustrates that without capacity constraints in transit assignment problem, not only are unreasonable results obtained for some line segments with overflow but the flow for some line segments that do not reach capacity is overestimated. This is consistent with the dual variables shown in Figure 2.17, i.e. the dual variable with respect to the capacity constraint of segment 17 on transit line 40 becomes zero, and that of segment 7 on transit line 52 is -20.7678, which means that the passenger flow has reached the capacity.

2.6 Chapter Summary

In this chapter, a strategy-based transit stochastic user equilibrium model with capacity and number-of-transfers constraints is proposed. Specifically, the logit-based stochastic path choice behavior and in-vehicle congestion cost are taken into account, a strict capacity constraint of transit line segments is added to handle the overload problem, and a number-of-transfers constraint is considered for transit path finding. This transit equilibrium problem is formulated as a logit-based VI problem. A transit path-set generation procedure based on the *k*-shortest path algorithm is introduced, which also considers the features of a route-section-based transit network and number-of-transfers constraint. The diagonalization method is adopted to solve the proposed model, and the diagonalized subproblem can be solved using the path-based partial linearization solution algorithm embedded with an iterative balancing scheme to handle the capacity constraints.

Numerical examples are provided to demonstrate the features of the proposed model and evaluate the performance of the developed solution algorithm. The results indicate that the number of transfers constraint changes the components of the transit path set, which strongly impacts the passenger flow patterns. The results also show that the capacity constraint affects flow patterns, which revises the evaluation of some transit management strategies. The results of a real-case transit network further verify the applicability of the developed solution algorithm. Overall, the numerical examples depict the importance of capacity and number-of-transfers constraints in transit equilibrium problems.

CHAPTER 3

FREQUENCY-BASED PATH FLOW ESTIMATOR FOR OD DEMAND ESTIMATION IN AN URBAN TRANSIT NETWORK

A frequency-based path flow estimator is proposed to estimate the transit passenger origin-destination (OD) matrix in an urban congested transit network. The proposed model not only considers the effect of congestion under an equilibrium framework, but also benefits from being formulated as a single-level model with the route-section-based network representation. Multiple transit data sources are incorporated, including automatic passenger count, automatic fare collection and automatic vehicle location data. A path-based diagonalization approach embedded with an iterative balancing scheme is developed to solve the model. Case studies are conducted to demonstrate the features and applicability of the proposed model and algorithm.

3.1 Introduction

In transportation (including both traffic and transit) planning and management studies, a fundamental input is the origin-destination (OD) trip matrix, which expresses the traffic demand between each OD pair (Li et al., 2012; Huang et al., 2016; Chen et al., 2018; Canca et al., 2019; An et al., 2020; Xu et al., 2020). However, it is rarely possible to obtain the "true" OD matrices directly in practice. A dedicated survey to collect this information would be highly labor- and resource-intensive. Many researchers instead focus their efforts on estimating the OD matrices based on the limited observations of traffic conditions on the network.

The OD demand estimation problem is well defined in the case of road (or traffic) networks, in which researchers make use of limited observations of traffic network conditions, i.e. link traffic counts and historical (or targeted) OD matrices (Carey et al., 1981; Cascetta and Nguyen, 1988; Yang, 1995; Yang et al., 2001; Maher et al., 2001). In transit networks, the traditional methods of estimating OD matrices are labor-intensive and difficult to implement, such as onboard survey and passenger counting at bus stops. With the development of electronic technology, observed data on passenger flows in transit networks can increasingly be obtained directly from the transit systems. These data principally consist of the onboard passenger counts of transit line segments and the smartcard payment data (transactions). Data of the first kind are obtained from automatic passenger count (APC) systems, which provide information including the boarding and alighting counts at each stop on the route, the time label, stop location and so on. Data of the second kind come from both automatic fare collection (AFC) and automatic vehicle location (AVL) systems, which provide detailed information on individual passengers, e.g. which stop they board at, which line they board and at what time.

A lot of studies have explored the transit OD demand estimation with the usage of APC, AFC and AVL data, including data-based approaches (Barry et al., 2002; Zhao et al., 2007; Trépanier et al., 2007; Munizaga and Palma, 2012), uncongested network approaches (Nguyen et al., 1988; Wong and Tong, 1998; Nuzzolo and Crisalli, 2001), and bi-level approaches with congested choice behavior (Lam et al., 2003; Wu and Lam, 2006). However, it is still unsolved that whether there exists a single-level optimization model incorporating APC, AFC, and AVL data for OD demand estimation problem in congested transit network. Therefore, this paper aims to answer this unsolved question.

3.1.1 Related literature

The implementation of these automatic transit-data collection systems in numerous urban networks has sparked a surge of interest in estimating transit OD matrices using observed data. At first, most studies of transit passenger OD flow estimation were conducted at the route level, and the observed data were principally the boarding and alighting counts at each stop acquired from APC systems. The most widely used method of OD matrix estimation, iterative proportional fitting (IPF), was proposed by Ben-Akiva et al. (1985). The inputs for this method are the boarding and alighting counts at every stop along a bus route and a seed OD flow matrix. Li and Cassidy (2007) presented an algorithm to estimate not only an OD matrix but also the passenger alighting probabilities at every stop on the route. The method's main advantages are

that it does not need a seed matrix and is more computationally efficient than the balancing method (i.e., IPF). Ji et al. (2015) proposed a computationally tractable method using both APC data and labor-intensive onboard survey data to estimate transit route passenger OD flow matrices. Later, Cui (2006) reported that route-level OD matrices could also serve as inputs to network-level OD demand estimation methods. Most of the above methods depend on a seed OD matrix to obtain high-quality estimates, but a good seed matrix is highly challenging to construct.

Elsewhere, researchers have explored the possibility of obtaining a network-level OD matrix directly using the data from AFC and AVL systems (Barry et al., 2002; Zhao et al., 2007; Trépanier et al., 2007). Munizaga and Palma (2012) presented a method of how to use the smartcard and GPS data to estimate a multimodal public transport OD matrix for Santiago, Chile. Their method centers on reconstructing passengers' trip chains from smartcard data by estimating the destination points from the information available. To apply methods of this kind, some assumptions are needed: (1) after a trip, passengers will return to the same stop from which they began that trip; (2) at the end of the day, passengers will return to the stop from which they began their first trip of that day; and (3) a criterion of maximum walking distance to the next boarding bus stop is used to define the alighting bus stop. AFC and AVL data can accurately illustrate individual passengers' boarding information. However, these methods face two drawbacks: (1) not all passengers use a smart card to pay the fare: for example, the smartcard penetration rate is approximately 90% in Chicago (Zhao et al., 2007); and (2) not all transaction records can be used to infer the alighting stations due to various practical reasons, e.g. only a single transaction is recorded, there is a data error or the trip is wrongly estimated to begin and end at the same location. The success rates of alighting-stop inference in previous studies include 66% by Trépanier et al. (2007), 71% by Zhao et al. (2007), and over 80% by Munizaga and Palma (2012). Therefore, the OD trip matrices obtained using AFC and AVL data are usually partial (incomplete).

In contrast to the above methods, which can be classified as data-based, other researchers have followed a different approach, using network-based models for transit OD trip matrix estimation. Due to the difficulties in modeling passenger route choice behavior in congested transit networks, the initial studies of transit OD matrix estimation focused on uncongested cases. Nguyen et al. (1988) developed a maximum entropy model for passenger OD matrix estimation in frequency-based transit systems, taking into account time information contained in the passenger counts. Wong and Tong (1998) also presented a maximum entropy model, but theirs was used to estimate the time-dependent passenger matrix in a schedule-based transit

network. Nuzzolo and Crisalli (2001) proposed a least-squares model using a schedule-based approach.

Later, with the advent of transit route choice behavior studies, researchers started to explore passenger OD matrix estimation in congested transit networks using a bi-level programming approach (Lam et al., 2003; Wu and Lam, 2006; Babazadeh et al., 2010). In such an approach, the upper level is the conventional OD matrix estimation problem while the lower level is the frequency-based transit equilibrium assignment problem. Lam et al. (2003) and Wu and Lam (2006) explored the transit OD demand estimation problem using a new frequency-based transit assignment model with elastic line frequencies (Lam et al., 2002). In their approach, it is assumed that the updated passenger counts and historical OD matrices are available. Babazadeh et al. (2010) proposed a bi-level model for transit OD matrix estimation with a path-based formulation of a strategy-based transit equilibrium assignment problem (Babazadeh and Aashtiani, 2005). The critical input to these bi-level methods is a seed (targeted) OD matrix, which usually cannot be directly obtained in practice. Moreover, the partial OD matrix from a data-based model could not be used in the upper level for these bi-level methods because the upper level is usually a generalized least squares problem, which will generate a result (i.e., an output matrix) that is as close as possible to the input partial OD matrix.

Table 3.1 Differences between existing transit OD demand estimation models and our model

Models			**Input**			**Output**		**Math. model**	
Type	**Transit choice behavior**	**References**	**APC data**	**AFC and AVL data**	**Historical OD matrices**	**Route level**	**Network level**	**Single level**	**Bi-level**
Data-based	-	Ben-Akiva et al., 1985; Li and Cassidy, 2007; Ji et al., 2015	✓	-	-	✓	-	-	-
		Barry et al., 2002; Zhao et al., 2007; Trépanier et al., 2007; Munizaga and Palma, 2012	-	✓	-	-	✓	-	-
Network -based	Uncongested	Nguyen et al., 1988; Wong and Tong, 1998	✓	-	✓	-	✓	✓	-
		Nuzzolo and Crisalli, 2001	✓	✓	-	-	✓	✓	-
	Congested	Lam et al., 2003; Wu and Lam, 2006; Babazadeh et al., 2010	✓	-	✓	-	✓	-	✓
		This study	✓	✓	✓	✓	✓	✓	-

Based on the above literature review, Table 1 summarizes the existing transit OD demand estimation models and their differences compared from the model proposed in this study. As can be seen, our model makes good use of APC, AFC and AVL data to solve the transit OD matrix estimation problem. A flexible network analysis tool, the path flow estimator (PFE), was developed by Bell et al. (1997) to conduct OD matrix estimation in road networks with the assumption of logit-based stochastic equilibrium assignment. A major advantage of PFE is that it allows different data sources to be incorporated into the model. Moreover, the PFE not only considers the effect of congestion but also the benefits from being formulated as a single-level mathematical problem (Bell and Iida, 1997; Chen et al., 2005, 2009, 2010).

3.1.2 Contribution of this work

This chapter proposes a frequency-based PFE for passenger OD matrix in congested urban transit networks via incorporating transit passenger route choice behavior and multiple transit data sources, i.e. APC, AFC and AVL data. However, we cannot apply the PFE approach to transit network directly due to the inconsistency in route choice units. Specifically, in a road network, the route choice unit is the elementary path, whereas in our problem, the route choice unit is the combination of a set of elementary paths. As described by Spiess and Florian (1989), passengers are assumed to follow their individual optimal strategies when making travel choices. Based on this assumption, there are two approaches for modeling the strategy: the hyperpath-based approach (Nguyen and Pallottino, 1988) and the route-section-based approach (de Cea and Fernández, 1993). In this chapter, we apply the route-section-based approach to model the passenger choice behavior. The route section is used to address the common lines issue, and a route is a sequence of route sections (i.e., a simplified strategy or hyperpath). The core of this approach is to apply the concept of transit logit-based stochastic equilibrium assignment to account for the effect of congestion at stops along the route section when modeling route choice behavior. The observed transit data serve as side constraints in our proposed approach. Two types of constraints are included: the onboard passenger counts, which can be obtained from APC data, and the partial OD matrix, which can be calculated from AFC and AVL data using data-based models (Barry et al., 2002; Zhao et al., 2007; Trépanier et al., 2007; Munizaga and Palma, 2012). Subsequently, the diagonalization method is adopted to solve the proposed frequency-based PFE for congested transit networks. In each diagonalized iteration, the proposed model is reformulated as a convex mathematical programming problem, and a path-based partial linearization algorithm embedded with an

iterative balancing scheme and a self-regulated averaging (SRA) scheme is developed to solve this convex optimization model.

3.2 Transit Network Modeling

3.2.1 Notation

This subchapter provides a list of the notation used in this chapter unless specified otherwise.

Sets	
N	set of transit stops
L	set of transit lines
E	set of transit line segments
E_U	set of transit line segments without observed data
E_M	set of transit line segments with observed data
S	set of route sections
A_s	set of attractive section line segments associated with route section s
OD	set of OD pairs
K^{od}	path set between OD pair od
Intermediate variables	
t_s	in-vehicle travel time of route section s
w_s	waiting time of route section s
c_s	total expected travel time of route section s
x_s^l	proportion of passengers choosing section line segment l associated with route section s
t_s^l	in-vehicle travel time of section line segment l associated with route section s
f_l	frequency of line l (vehicles/min)
κ_l	capacity of line l (passengers/vehicle)
v_s	passenger flow on route section s
v_s^l	passenger flow on section line segment l associated with route section s
v_e	passenger flow on transit line segment e
q^{od}	passenger flow between OD pair od
Decision variable	
h_k^{od}	passenger flow on path k between OD pair od
Inputs	
θ	dispersion parameter measuring passengers' perception of transit path cost
C_e	capacity on transit line segment e
$\bar{v}_e$	observed passenger count on transit line segment e
ϵ_e	percentage of measurement error allowed for the passenger count on transit line segment e
$\bar{q}^{od}$	observed partial trip demand of OD pair od

3.2.2 Route-section-based transit network modeling revisited

With the route-section-based transit network representation in Chapter 2.2.2, the transit cost is modeled here. To formulate the path cost, we need an expression for the route section cost, which consists of the in-vehicle travel time, the waiting time and the perceived congestion

time. As shown below, we first need to introduce an interim variable, x_s^l. Based on the network representation above, the route section flows are assigned to section line segments in direct proportion to their frequencies on route sections, i.e.

$$x_s^l = \frac{f_l}{\sum_{j \in A_s} f_j}, \forall l \in A_s, s \in S \tag{3.1}$$

The in-vehicle travel time of route section s can then be expressed as the weighted summation of the in-vehicle travel time of all of the attractive lines (section line segments) associated with route section s as:

$$t_s = \sum_{l \in A_s} x_s^l t_l, \forall s \in S \tag{3.2}$$

The waiting time for passengers boarding route section s can be expressed as:

$$w_s = \frac{\alpha}{f_s} = \frac{\alpha}{\sum_{l \in A_s} f_l}, \forall s \in S \tag{3.3}$$

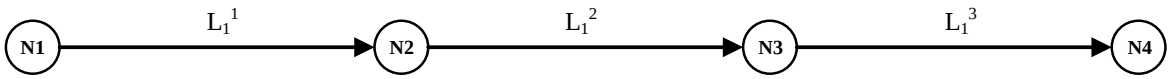
where the parameter α may be chosen to approximate the distribution assumed for the vehicle headway (Spiess and Florian, 1989). The value $\alpha = 1$ corresponds to an exponential distribution assumed for the vehicle headway and $\alpha = 0.5$ represents a uniform distribution.

The perceived congestion time of the route section involves the additional waiting time due to vehicle congestion, which is a function of its own flow and that of its competing route sections. The flow on section s is:

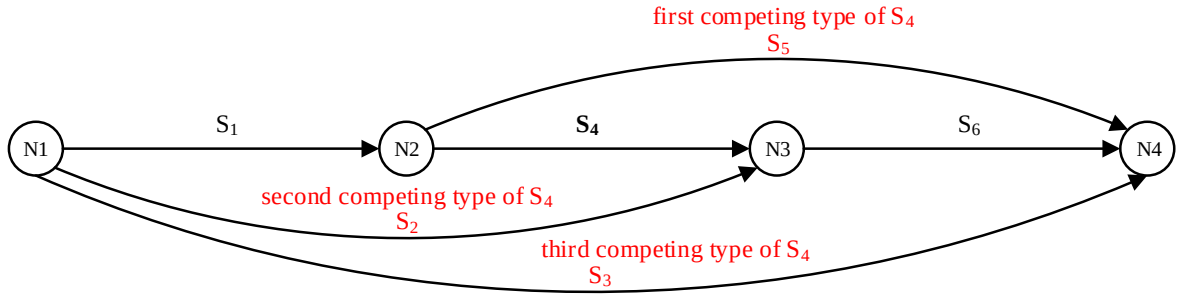
$$v_s = \sum_{od \in OD} \sum_{k \in K^{od}} a_{sk} h_k^{od}, \forall s \in S \tag{3.4}$$

where the path-section incidence a_{sk} equals 1 if section s lies on path k, otherwise 0.

There are three main groups of passengers competing with section s: (1) passengers boarding at $tail(s)$ of all other route sections (sharing the same tail node) that use lines contained in route section s, (2) passengers boarding any of the lines belonging to route section s at a node before $tail(s)$ and alighting at $head(s)$ and (3) passengers boarding any of the lines belonging to route section s at a node before $tail(s)$ and alighting after $head(s)$. Figure 3.1 shows an example of each type of competing section. Specifically, for route section S_4, route section S_5 is of the first competing type, as it shares the same tail node with S_4; route section S_2 belongs to the second type; and route section S_3 belongs to the third type.



(a) Transit network using the line and itinerary description



(b) Transit network using the route-section description

Figure 3.1 Illustration of competing sections of a route section

Then, the competing section flow of section s is:

$$\tilde{v}_s = \sum_{\hat{s} \neq s \in S} \delta_s^{\hat{s}} \sum_{l \in A_s \cap A_{\hat{s}}} v_{\hat{s}l}, \forall s \in S \tag{3.5}$$

where $\delta_s^{\hat{s}}$ is the competing section indicator, such that $\delta_s^{\hat{s}} = 1$ means that section $\hat{s}$ is a competing section of section s, and otherwise $\delta_s^{\hat{s}} = 0$.

The section line segment flow v_s^l is determined by:

$$v_s^l = v_s x_s^l, \forall l \in A_s, s \in S \tag{3.6}$$

Then the perceived congestion time function for route section s is expressed as:

$$\phi_s(\mathbf{v}) = \varphi_s \left(\frac{\vartheta v_s + \varsigma \tilde{v}_s}{\sum_{l \in A_s} f_l \kappa_l} \right)^{\varpi}, \forall s \in S \tag{3.7}$$

where calibration parameters ϑ, ς, φ_s and ϖ are used to model different effects of various flows on the perceived congestion time.

For route section s, the expected total travel time is given by:

$$c_s(\mathbf{v}) = t_s + w_s + \phi_s(\mathbf{v}), \forall s \in S \tag{3.8}$$

With the expected route section time function, the expected travel time associated with path k between OD pair od can be expressed as:

$$c_k^{od}(\mathbf{v}) = \sum_{s \in S} a_{sk} c_s(\mathbf{v}), \forall k \in K^{od}, od \in OD \tag{3.9}$$

3.3 Frequency-based Transit Path Flow Estimator

3.3.1 Framework of transit PFE with APC, AFC, and AVL data

This section mainly presents modeling process of the frequency-based path flow estimator. After introducing the framework of transit PFE with APC, AFC, and AVL data, the flow conservations are introduced. Then, the uncongested and congested transit PFE models are proposed, which help to illustrate the modeling process.

3.3.1.1 Overall schematic

The onboard passenger counts and the observed partial OD trip matrix can serve as two kinds of supplementary data for transit OD demand estimation. A schematic of frequency-based transit PFE with APC, AFC, and AVL data is shown in Figure 3.2. The properties of these three kinds of data from automatic data collection systems could be summarized as

- APC data: line no., bus no., time, boarding count, and alight count;
- AFC data: smart card ID, tap-in time, line no., bus no., line name;
- AVL data: line no., bus no., bus stop, arrival/departure time (AD-Time), arrival/departure flag (AD-Flag).

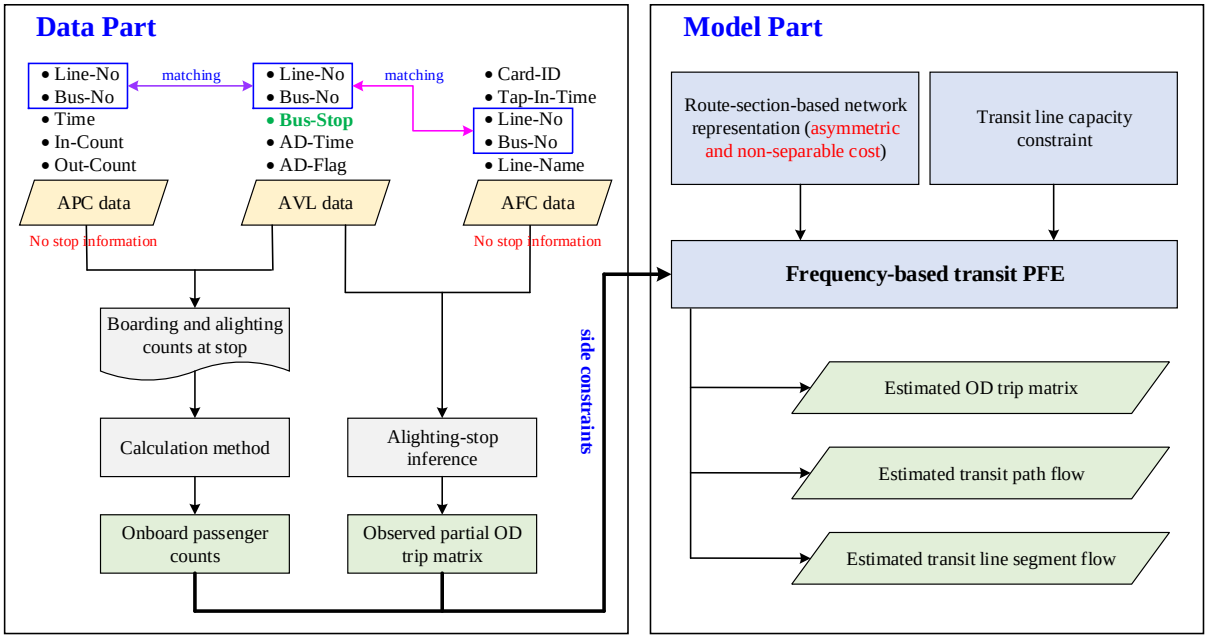


Figure 3.2 Schematic of frequency-based transit PFE with APC, AFC, and AVL data

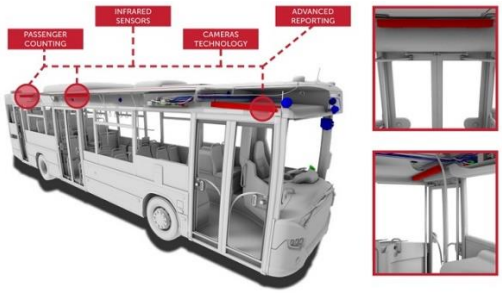
Note that neither APC nor AFC data provides the stop information (i.e., location), which requires the AVL data to help locate the stop. First, the onboard passenger counts of transit line segments are obtained from APC and AVL data, and the observed partial OD trip matrix will also be generated from AFC and AVL data based on inference of the passengers' alighting

stops. Based on the route-section-based transit network representation, a frequency-based transit PFE for OD estimation is proposed, together with three kinds of side constraints: an onboard passenger count constraint for measured line segments, a capacity constraint for unmeasured line segments and a partial OD flow matrix constraint. The proposed frequency-based transit PFE approach yields the following outputs: estimated OD flow matrix, estimated transit path flow and estimated transit line segment flow. By combining with the onboard passenger counts and the observed partial OD trip matrix, the estimated result improves the underspecified and partial issues of the estimated matrix as discussed in Section 3.1.

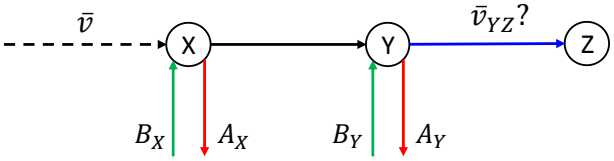
3.3.1.2 Onboard passenger counts

Onboard passenger counts of each transit line segment can be obtained from the boarding and alighting counts at each stop on the line. Figure 3.3a shows a representative method for the detection of boarding and alighting counts at a bus stop using infrared light. Infrared sensors are installed at both the entry and exit doors of buses and register whenever a passenger passes through either door by detecting the temporary blockage of the light beam. From these boarding and alighting counts, the number of passengers on each line segment (or onboard passenger count) can be calculated. As illustrated in Figure 3.3b for sequential stops on a bus line, the flow on the bus line segment before stop X is $\bar{v}$. The boarding and alighting flow at stop X is B_X and A_X, while that at stop Y is B_Y and A_Y. Thus, the observed onboard passenger count of line segment YZ calculated via the boarding and alighting counts is:

$$\bar{v}_{YZ} = \bar{v} + B_X - A_X + B_Y - A_Y \tag{3.10}$$



(a) Automatic passenger count (APC) system using infrared sensors



(b) Number of boarding, alighting and onboard passengers

Figure 3.3 Illustration of a method of deriving onboard passenger counts using APC data

Specifically, when stop X is the starting stop of the transit line (i.e. $\bar{v} = 0$), the observed onboard passenger count of line segment XY equals the boarding flow at stop X.

3.3.1.3 Observed partial OD trip matrix

Many AFC systems, notably those for bus transit, are access-based (i.e. swipe-on or tap-on) only and thus only record the stops where passengers board, not record where they alight. A well-analyzed form of the transit passenger OD matrix estimation problem is based on estimating a passenger's alighting stop following a sequence of smartcard transactions and assuming that the next transaction occurs after alighting (Barry et al., 2002; Zhao et al., 2007; Trépanier et al., 2007; Munizaga and Palma, 2012). The inputs for these existing methods, which are used to infer passengers' alighting stops, are obtained from three main types of databases: transactions (boarding) from an AFC system, vehicle positions from an AVL system and a geocoded representation of a public transport network. Unfortunately, the estimated OD matrix is usually partial compared for two reasons:

(1) Alighting stops cannot be estimated with 100% accuracy due to several potential issues, e.g. only a single transaction is recorded, there is a data error or a trip is wrongly estimated to begin and end at the same location. The success rates of previous rates include 66% by Trépanier et al. (2007), 71% by Zhao et al. (2007) and over 80% by Munizaga and Palma (2012). In such cases, not all of the transaction data can be utilized due to missing information on the chain linkages. We use the variable ϱ to represent the success rate with which the transaction data are used to infer the OD matrix, and $\bar{q}^{od}$ for the OD matrix estimated via inference of the alighting stops.

(2) Although smartcard data can be used to estimate the passenger OD matrix, the matrix will still be partial even if the success rate ϱ reaches 100% because not all passengers use the a smartcard for payment. For example, the penetration rate of smartcards in Chicago is close to 90% (Zhao et al., 2007), while that in Santiago, Chile is approximately 97% (Beltrán et al., 2011). We can calculate that the estimated partial matrices in Zhao et al. (2007) and Beltrán et al. (2011) capture approximately only 63.9% and 77.6% of the complete matrices, respectively.

3.3.1.4 Relationship between transit observations and complete OD matrix

Figure 3.4 shows the relationships between the complete OD demand matrix and the various transit data sources:

(1) Onboard passenger counts from APC data (e.g. infrared sensors) usually reflect the characteristics of the complete OD matrix. However, this does not guarantee that we can obtain a high-quality OD matrix using only onboard passenger counts to formulate the OD estimation

problem. Such a problem would be underspecified and would have multiple solutions because the number of observations (i.e., transit line segments) is generally less than the number of variables (i.e., OD pairs).

(2) Not all passengers use smartcard payment systems, while some smartcard transaction records are unusable for various practical reasons. The estimated OD matrix from data-based models using AFC and AVL data is usually only a part of the complete matrix. Even if all of the smartcard records could be used for estimation, this estimated (observed) OD matrix would still be partial due to the incomplete penetration rate of smartcards in urban networks.

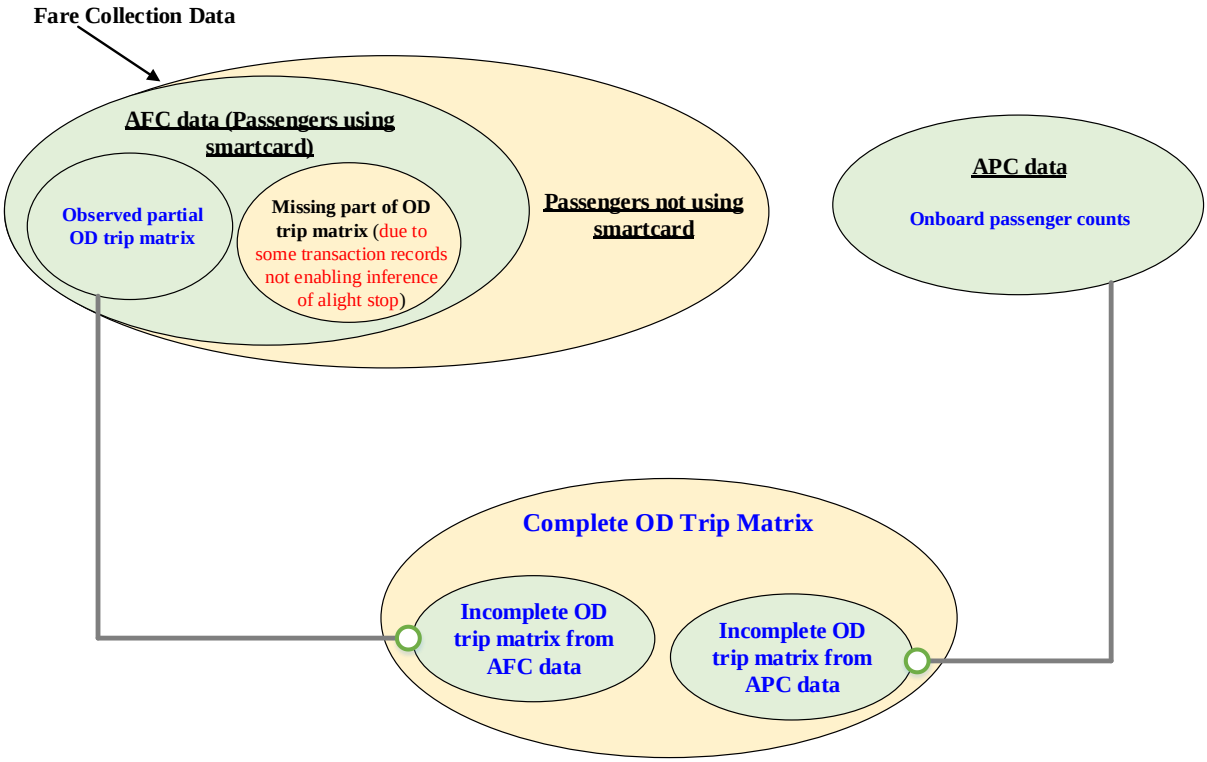


Figure 3.4 Relationship between the complete OD demand matrix and transit data

3.3.2 Flow conservation in transit network

The relationship between the OD flows and passenger path flows is expressed as follows:

$$\sum_{k \in K^{od}} h_k^{od} = q^{od}, \forall od \in OD \tag{3.11}$$

Let $\Gamma = (\gamma_{es})$ denote the line segment-route section incidence matrix, which equals 1 if line segment e of line l lies on route section s, otherwise 0. The line segment flow expression is:

$$v_e = \sum_{s \in S} \gamma_{es} y_l^e x_s^l v_s \,, \forall e \in E \tag{3.12}$$

where $y_l^e = 1$ means that transit line segment e is on transit line l.

For simplicity of expression, let $\bar{\gamma}_{es} = \gamma_{es} x_s^l$ denote the proportion of passengers choosing line segment e of line l associated with route section s. The line segment flow v_e can thus be further expressed as:

$$v_e = \sum_{s \in S} \bar{\gamma}_{es} v_s \,, \forall e \in E \tag{3.13}$$

The relationship between line segment flows and path flows can be obtained:

$$v_e = \sum_{od \in OD} \sum_{k \in K^{od}} \sum_{s \in S} \bar{\gamma}_{es}\, a_{sk} h_k^{od} \,, \forall e \in E \tag{3.14}$$

The line capacity constraint for the line segments without observed data can be described by:

$$v_e \le C_e, \forall e \in E_U \tag{3.15}$$

The constraint for the observed passenger count on a line segment e is:

$$(1 - \epsilon_e)\bar{v}_e \le v_e \le (1 + \epsilon_e)\bar{v}_e, \forall e \in E_M \tag{3.16}$$

The observed partial OD demand matrix $\bar{q}^{od}$ in Section 2.3.2 can serve as supplementary (observed) data to estimate the OD trip matrix. In other words, the true OD demand matrix should be larger than or equal to this observed partial matrix. Using the success rate ϱ and penetration rate Υ, the upper bound of the estimated trip matrix is $\bar{q}^{od}/(\varrho\Upsilon)$. Thus, the constraint of the OD trip matrix is:

$$\bar{q}^{od} \le q^{od} \le \bar{q}^{od}/(\varrho\Upsilon) \,, \forall od \in OD \tag{3.17}$$

3.3.3 Uncongested transit PFE formulation

The initial PFE formulation given by Bell and Iida (1997) was for an uncongested private car network. Before proposing the frequency-based PFE formulation for a congested transit network, we present the uncongested case first.

When ignoring congestion, the term quantifying the perceived congestion time, $\phi_s(\mathbf{v})$, becomes 0. Then, the travel cost of route section s is expressed as the summation of only two terms:

$$c_s = t_s + w_s, \forall s \in S \tag{3.18}$$

Further, the path cost of path k between OD pair od can be expressed as:

$$c_k^{od} = \sum_{s \in S} a_{sk} c_s \,, \forall k \in K^{od}, od \in OD \tag{3.19}$$

Thus, the frequency-based PFE formulation for an uncongested transit network is as follows:

[T-PFE]

$$\min \frac{1}{\theta} \sum_{od \in OD} \sum_{k \in K^{od}} h_k^{od} \left(\ln h_k^{od} - 1 \right) + \sum_{od \in OD} \sum_{k \in K^{od}} c_k^{od} h_k^{od} \tag{3.20}$$

s.t.

$$v_e \le C_e, \forall e \in E_U \tag{3.21}$$

$$(1 - \epsilon_e)\bar{v}_e \le v_e \le (1 + \epsilon_e)\bar{v}_e, \forall e \in E_M \tag{3.22}$$

$$\bar{q}^{od} \le q^{od} \le \bar{q}^{od}/(\varrho \Upsilon) \,, \forall od \in OD \tag{3.23}$$

$$h_k^{od} \ge 0, \forall k \in K^{od}, od \in OD \tag{3.24}$$

where

$$\sum_{k \in K^{od}} h_k^{od} = q^{od}, \forall od \in OD \tag{3.25}$$

$$v_s = \sum_{od \in OD} \sum_{k \in K^{od}} a_{sk} h_k^{od} \,, \forall s \in S \tag{3.26}$$

$$v_e = \sum_{od \in OD} \sum_{k \in K^{od}} \sum_{s \in S} \bar{\gamma}_{es} \, a_{sk} h_k^{od} \,, \forall e \in E \tag{3.27}$$

Similar to that in Bell and Iida (1997), the objective function (3.20) also has two terms: the maximizes the path flow entropy to spread the travel demand, and the second assigns the travel demand on the least-cost paths. A detailed description of the side constraints (3.21)-(3.24) and definitional constraints (3.25)-(3.27) can be found in Section 3.3.2.

Proposition 3.1. The distribution of the estimated path flow for [T-PFE] in the optimal solution obeys the multinomial logit model.

Proof. The Lagrange formulation of [T-PFE] with respect to the constraints can be formulated as:

$$\mathcal{L} = \frac{1}{\theta}\sum_{od\in OD}\sum_{k\in K^{od}} h_k^{od}\left(\ln h_k^{od} - 1\right) + \sum_{od\in OD}\sum_{k\in K^{od}} c_k^{od} h_k^{od} + \sum_{e\in E_U} d_e(C_e - v_e)$$
$$+ \sum_{e\in E_M} l_e\left((1-\epsilon_e)\bar{v}_e - v_e\right) + \sum_{e\in E_M} u_e\left((1+\epsilon_e)\bar{v}_e - v_e\right) \quad (3.28)$$
$$+ \sum_{od\in OD} l^{od}(\bar{q}^{od} - q^{od}) + \sum_{od\in OD} u^{od}(\bar{q}^{od}/(\varrho\Upsilon) - q^{od})$$

where d_e, l_e, u_e, l^{od} and u^{od} are the dual variables of constraints (3.21)-(3.23).

After taking the first-order derivative of $\mathcal{L}$ with respect to primal variables h_k^{od}, we obtain:

$$\frac{\partial \mathcal{L}}{\partial h_k^{od}} = c_k^{od} + \frac{1}{\theta}\ln h_k^{od} - \sum_{e\in E_U} d_e \sum_{s\in S}\bar{\gamma}_{es}a_{sk} - \sum_{e\in E_M} l_e \sum_{s\in S}\bar{\gamma}_{es}a_{sk}$$
$$- \sum_{e\in E_M} u_e \sum_{s\in S}\bar{\gamma}_{es}a_{sk} - l^{od} - u^{od} = 0 \quad (3.29)$$

We then have the following path flow expression:

$$h_k^{od} = \exp\left(\theta\left(-c_k^{od} + \sum_{e\in E_U} d_e \sum_{s\in S}\bar{\gamma}_{es}a_{sk} + \sum_{e\in E_M} l_e \sum_{s\in S}\bar{\gamma}_{es}a_{sk}\right.\right.$$
$$\left.\left. + \sum_{e\in E_M} u_e \sum_{s\in S}\bar{\gamma}_{es}a_{sk} + l^{od} + u^{od}\right)\right) \quad (3.30)$$

Thus, the probability of choosing path k for OD pair od is:

$$p_k^{od} = \frac{h_k^{od}}{q^{od}} = \frac{h_k^{od}}{\sum_{k\in K^{od}} h_k^{od}} = \frac{\exp\left(\theta\left(-c_k^{od} + J_k^{od}\right)\right)}{\sum_{k'\in K^{od}}\exp\left(\theta\left(-c_{k'}^{od} + J_{k'}^{od}\right)\right)}, \forall k \in K^{od}, od \in OD \quad (3.31)$$

where $J_k^{od} = \sum_{e\in E_U} d_e \sum_{s\in S}\bar{\gamma}_{es}a_{sk} + \sum_{e\in E_M} l_e \sum_{s\in S}\bar{\gamma}_{es}a_{sk} + \sum_{e\in E_M} u_e \sum_{s\in S}\bar{\gamma}_{es}a_{sk}$, $\forall k \in K^{od}, od \in OD$. Eq. (3.31) implies that the path flow distribution obeys the multinomial logit model.

This completes the proof. □

3.3.4 Congested transit PFE formulation

When considering congestion, the route-section cost function in general has an asymmetric Jacobian. Therefore, the congested transit PFE does not have an equivalent mathematical programming formulation. Here we propose a variational inequality formulation for the congested transit PFE instead.

[VI-T-PFE]

To find an optimal solution h_k^{od*} such that

$$\sum_{od\in OD}\sum_{k\in K^{od}}\left(c_k^{od}(\mathbf{h}^*)+\frac{1}{\theta}\ln h_k^{od*}\right)\left(h_k^{od}-h_k^{od*}\right)\geq 0,\forall h_k^{od}\in\Omega \tag{3.32}$$

where Ω represents the feasible region, i.e. the region where Eqs. (3.21)-(3.27) hold.

Letting $\mathbf{F}(\mathbf{h})=c_k^{od}(\mathbf{h})+\frac{1}{\theta}\ln h_k^{od}$ and $\mathbf{h}=\left[h_k^{od}\right]$, the VI model can be simplified to a standard form:

$$\mathbf{F}(\mathbf{h}^*)^T(\mathbf{h}-\mathbf{h}^*)\geq 0,\forall\mathbf{h}\in\Omega \tag{3.33}$$

Proposition 3.2. The distribution of the estimated path flow for [VI-T-PFE] in the optimal solution follows the multinomial logit model.

Proof. The first-order conditions (Karush-Kuhn-Tucker (KKT) conditions) for the [VI-T-PFE] model are:

$$\begin{aligned}\left(c_k^{od}+\frac{1}{\theta}\ln h_k^{od}-\sum_{e\in E_U}d_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}-\sum_{e\in E_M}l_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}-\sum_{e\in E_M}u_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}\right.\\ \left.-l^{od}-u^{od}\right)h_k^{od}=0,\forall k\in K^{od},od\in OD\end{aligned} \tag{3.34}$$

$$\begin{aligned}c_k^{od}+\frac{1}{\theta}\ln h_k^{od}-\sum_{e\in E_U}d_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}-\sum_{e\in E_M}l_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}-\sum_{e\in E_M}u_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}\\ -l^{od}-u^{od}\geq 0,\forall k\in K^{od},od\in OD\end{aligned} \tag{3.35}$$

Because $h_k^{od}>0$, the equation below is satisfied:

$$\begin{aligned}c_k^{od}+\frac{1}{\theta}\ln h_k^{od}-\sum_{e\in E_U}d_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}-\sum_{e\in E_M}l_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}-\sum_{e\in E_M}u_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}\\ -l^{od}-u^{od}=0\end{aligned} \tag{3.36}$$

Then, the analytical expression of the path flow for each OD pair is

$$\begin{aligned}h_k^{od}=\exp\left(\theta\left(-c_k^{od}+\sum_{e\in E_U}d_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}+\sum_{e\in E_M}l_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}\right.\right.\\ \left.\left.+\sum_{e\in E_M}u_e\sum_{s\in S}\bar{\gamma}_{es}a_{sk}+l^{od}+u^{od}\right)\right)\end{aligned} \tag{3.37}$$

Thus, the probability of choosing path k for OD pair od is given as:

$$p_k^{od} = \frac{h_k^{od}}{q^{od}} = \frac{h_k^{od}}{\sum_{k \in K^{od}} h_k^{od}} = \frac{\exp\left(\theta\left(-c_k^{od} + J_k^{od}\right)\right)}{\sum_{k' \in K^{od}} \exp\left(\theta\left(-c_{k'}^{od} + J_{k'}^{od}\right)\right)}, \forall k \in K^{od}, od \in OD \quad (3.38)$$

which indicates that the estimated path flow in the optimal solution follows the multinomial logit model.

This completes the proof. □

3.4 Solution Algorithm

This paper proposes both uncongested and congested transit PFE models, where the uncongested one is a special case of congested one (see Section 3.4.1 and 3.4.2). Hence, this section mainly presents how to solve the congested transit PFE model. Due to the asymmetric features of the route-section and path cost functions, we adopt the well-known diagonalization method to solve the proposed congested transit PFE formulation [VI-T-PFE] (Florian, 1977). The core procedure of this algorithm is to diagonalize the cost function to get a mathematical programming (MP) formulation in each diagonalized iteration. The diagonalized PFE formulation can then be solved by the partial linearization algorithm embedded with an iterative balancing scheme (Chen et al, 2009).

3.4.1 Diagonalization of cost functions

In this section, we show how the diagonalized [VI-T-PFE] model can be reformulated as a convex mathematical programming model for which effective path-based solution algorithms are available. To do this, we introduce the following proposition.

Proposition 3.3. When the cost function of the [VI-T-PFE] model is diagonalized, [VI-T-PFE] can be reformulated as a convex optimization problem such that

[MP-T-PFE]

$$\min \sum_{s \in S} \int_0^{v_s} \hat{c}_s(\omega)\, d\omega + \frac{1}{\theta} \sum_{od \in OD} \sum_{k \in K^{od}} h_k^{od}\left(\ln h_k^{od} - 1\right) \quad (3.39)$$

subject to Eqs. (3.21)-(3.27).

Proof. [VI-T-PFE] in Eq. (3.32) can be further expressed as

$$\sum_{od \in OD} \sum_{k \in K^{od}} \left(c_k^{od}(\mathbf{h}^*) + \frac{1}{\theta} \ln h_k^{od*}\right)\left(h_k^{od} - h_k^{od*}\right) \geq 0, \forall h_k^{od} \in \Omega \quad (3.40)$$

We separate the term $\left(c_k^{od}(\mathbf{h}^*) + \frac{1}{\theta}\ln h_k^{od*}\right)$ in Eq. (3.40) into two parts, and multiply each part by $\left(h_k^{od} - h_k^{od*}\right)$ to get

$$\sum_{od\in OD}\sum_{k\in K^{od}} c_k^{od}(\mathbf{h}^*)\left(h_k^{od} - h_k^{od*}\right) + \sum_{od\in OD}\sum_{k\in K^{od}} \frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od} - h_k^{od*}\right) \geq 0, \forall h_k^{od} \in \Omega \tag{3.41}$$

For Eq. (3.41), we substitute a route section cost function for the path cost function $c_k^{od}(\mathbf{h}^*)$. Then, Eq. (3.31) can be expressed as

$$\sum_{od\in OD}\sum_{k\in K^{od}}\left(\sum_{s\in S} a_{sk}c_s(\mathbf{h}^*)\right)\left(h_k^{od} - h_k^{od*}\right) + \sum_{od\in OD}\sum_{k\in K^{od}} \frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od} - h_k^{od*}\right) \geq 0, \forall h_k^{od} \in \Omega \tag{3.42}$$

Further, we can calculate the route section flow based on the summation symbols for all paths for one OD pair and for all OD pairs.

$$\sum_{s\in S}\left(\sum_{od\in OD}\sum_{k\in K^{od}} a_{sk}\left(h_k^{od} - h_k^{od*}\right)\right)c_s(\mathbf{h}^*) + \sum_{od\in OD}\sum_{k\in K^{od}} \frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od} - h_k^{od*}\right) \geq 0, \forall h_k^{od} \in \Omega \tag{3.43}$$

Then from Eq. (3.43), Eq. (3.32) can be simplified using the route section and path space as follows:

$$\sum_{s\in S}(v_s - v_s^*)c_s(\mathbf{h}^*) + \sum_{od\in OD}\sum_{k\in K^{od}} \frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od} - h_k^{od*}\right) \geq 0, \forall h_k^{od} \in \Omega \tag{3.44}$$

The route section cost in Eq. (3.44) can be expressed as a function of the route section flow, which can be calculated from the path flow. Thus, Eq. (3.32) is finally expressed as below:

$$\sum_{s\in S}(v_s - v_s^*)c_s(\mathbf{v}^*) + \sum_{od\in OD}\sum_{k\in K^{od}} \frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od} - h_k^{od*}\right) \geq 0, \forall h_k^{od} \in \Omega \tag{3.45}$$

At each iteration the $c_s(\mathbf{v})$ is diagonalized at the current solution, yielding a symmetric assignment problem.

$$\sum_{s\in S}(v_s - v_s^*)\hat{c}_s(\mathbf{v}^*) + \sum_{od\in OD}\sum_{k\in K^{od}} \frac{1}{\theta}\ln h_k^{od*}\left(h_k^{od} - h_k^{od*}\right) \geq 0, \forall h_k^{od} \in \Omega \tag{3.46}$$

where $\hat{c}_s(\mathbf{v})$ is the diagonalized expression of $c_s(\mathbf{v})$.

Thus, the above symmetric assignment problem has an equivalent convex optimization formulation (T-PFE] in Eq. (3.20) is equivalent to the direction finding of the [MP-T-PFE] in Eq. (3.39))

$$\min \sum_{s\in S}\int_0^{v_s}\hat{c}_s(\omega)\,d\omega+\frac{1}{\theta}\sum_{od\in OD}\sum_{k\in K^{od}}h_k^{od}\left(\ln h_k^{od}-1\right) \tag{3.47}$$

subject to Eqs. (3.21)-(3.27).

This completes the proof. □

3.4.2 Overall solution procedure

The whole framework of the solution algorithm for VI formulation is presented in Figure 3.5. The solution procedure can be summarized into the following steps:

Step 0. Initialization. Find an initial feasible solution $(\bar{\mathbf{v}},\bar{\mathbf{h}})$.

Step 1. Diagonalize $c(\mathbf{v})$ at $(\bar{\mathbf{v}},\bar{\mathbf{h}})$.

Step 2. Solve problem [MP-T-PFE] to get $(\hat{\mathbf{v}},\hat{\mathbf{h}})$.

Step 3. Stop test. If $(\bar{\mathbf{v}},\bar{\mathbf{h}})$ and $(\hat{\mathbf{v}},\hat{\mathbf{h}})$ are sufficiently close, stop; otherwise:

Step 4. Make: $(\bar{\mathbf{v}},\bar{\mathbf{h}})\leftarrow(\hat{\mathbf{v}},\hat{\mathbf{h}})$, and return to Step 1.

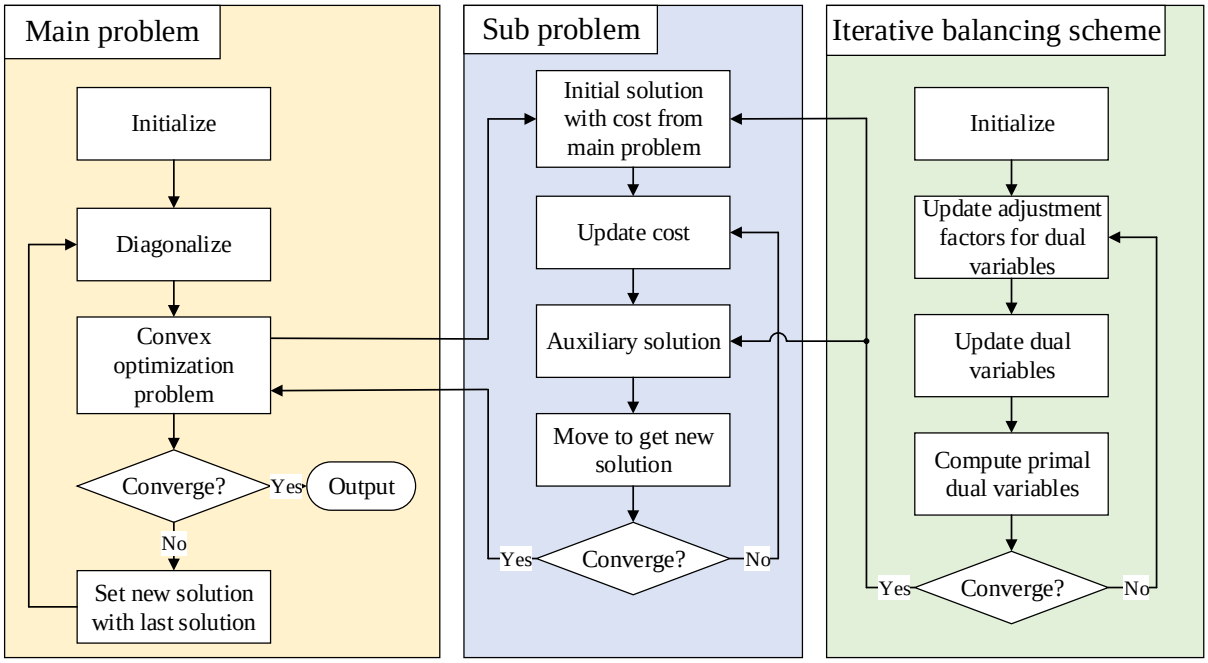


Figure 3.5 Framework of the solution algorithm for [VI-T-PFE]

To solve the subproblem in Step 2 ("Convex optimization problem" in Figure 3.5), we develop the path-based partial linearization algorithm combined with an SRA step-size scheme (Liu et al., 2009), embedded with an iterative balancing scheme and column generation procedure. The line search step determines how far the current solution should move in the

search direction. The new solution is found as a convex combination of the solution of the above subproblem and the current solution. Note that in Step 2, the SRA method determines the step size based on the distance between auxiliary point $\tilde{\mathbf{h}}^i$ and current solution $\mathbf{h}^i$, due to the fact that $\tilde{\mathbf{h}}^i \to \mathbf{h}^*$. The rules for calculating the step size are as follows:

$$\sigma^i = 1/\beta^i \tag{3.48}$$

$$\beta^i = \begin{cases} \beta^{i-1} + \lambda_1, & if \left| \mathbf{h}^i - \tilde{\mathbf{h}}^i \right| \geq \left| \mathbf{h}^{i-1} - \tilde{\mathbf{h}}^{i-1} \right| \\ \beta^{i-1} + \lambda_2, & otherwise \end{cases} \tag{3.49}$$

where $\lambda_1 > 1$ and $0 < \lambda_2 < 1$.

This direction finding problem in Step 2 (see the algorithm in Figure 3.6) must be solved by the iterative balancing scheme used in the original PFE model (Bell and Iida, 1997; Bell et al., 1997) due to the large number of inequality side constraints. The iterative balancing scheme is used with a given path set and fixed cost, and its core procedure is to adjust dual variables related to the side constraints and update the corresponding primal variables at each iteration.

Algorithm. Iterative balancing scheme

1: Initialization.

(a) Set $j = 0$; $(d_e)^j = 0$, $(l_e)^j = 0$, $(u_e)^j = 0$ for all transit line segments; $\left(l^{od}\right)^j = 0$, $\left(u^{od}\right)^j = 0$ for all OD pairs.

(b) compute primal variables:

$\left(h_k^{od}\right)^j = \exp\left(\theta\left(-c_k^{od}\right)\right), \forall k \in K^{od}, od \in OD$

$(v_e)^j = \sum_{s \in S} \bar{\gamma}_{es} \sum_{od \in OD} \sum_{k \in K^{od}} a_{sk} \left(h_k^{od}\right)^j, \forall e \in E$

2: **while** $j \leq Iter_{max}$ and $\varepsilon \geq \underline{\eta}$ and $\varepsilon < \overline{\eta}$ **do**

3: **for** each transit line segment e **do**

update dual variables

$(d_e)^{j+1} = \min\left\{0, (d_e)^j + \frac{1}{\theta} \ln \frac{C_e}{(v_e)^j}\right\}$

$(l_e)^{j+1} = \max\left\{0, (l_e)^j + \frac{1}{\theta} \ln \frac{(1-\epsilon_e)\bar{v}_e}{(v_e)^j}\right\}$

$(u_e)^{j+1} = \min\left\{0, (u_e)^j + \frac{1}{\theta} \ln \frac{(1+\epsilon_e)\bar{v}_e}{(v_e)^j}\right\}$

update primal variables

$\left(h_k^{od}\right)^{j+1} = \exp\left(\theta\left(-c_k^{od} + \sum_{e \in E_U} (d_e)^{j+1} \sum_{s \in S} \bar{\gamma}_{es} a_{sk} + \sum_{e \in E_M} (l_e)^{j+1} \sum_{s \in S} \bar{\gamma}_{es} a_{sk} + \sum_{e \in E_M} (u_e)^{j+1} \sum_{s \in S} \bar{\gamma}_{es} a_{sk} + \left(l^{od}\right)^j + \left(u^{od}\right)^j\right)\right), \forall k \in K^{od}, od \in OD$

$(v_e)^{j+1} = \sum_{s \in S} \bar{\gamma}_{es} \sum_{od \in OD} \sum_{k \in K^{od}} a_{sk} \left(h_k^{od}\right)^{j+1}, \forall e \in E$

4: **for** each OD pair od **do**

$(q^{od})^j = \sum_{k\in K^{od}}(h_k^{od})^{j+1}$

update dual variables

$(l^{od})^{j+1} = \max\left\{0, (l^{od})^j + \frac{1}{\theta}\ln\frac{\bar{q}^{od}}{(q^{od})^j}\right\}$

$(u^{od})^{j+1} = \min\left\{0, (u^{od})^j + \frac{1}{\theta}\ln\frac{\bar{q}^{od}/(\varrho\Upsilon)}{(q^{od})^j}\right\}$

update primal variables

$(h_k^{od})^{j+1} = \exp\left(\theta\left(-c_k^{od} + \sum_{e\in E_U}(d_e)^{j+1}\sum_{s\in S}\bar{\gamma}_{es}a_{sk} + \sum_{e\in E_M}(l_e)^{j+1}\sum_{s\in S}\bar{\gamma}_{es}a_{sk} + \sum_{e\in E_M}(u_e)^{j+1}\sum_{s\in S}\bar{\gamma}_{es}a_{sk} + (l^{od})^{j+1} + (u^{od})^{j+1}\right)\right), \forall k \in K^{od}$

$(q^{od})^{j+1} = \sum_{k\in K^{od}}(h_k^{od})^{j+1}$

$(v_e)^{j+1} = \sum_{s\in S}\bar{\gamma}_{es}\sum_{od\in OD}\sum_{k\in K^{od}}a_{sk}(h_k^{od})^{j+1}, \forall e \in E$

$$\varepsilon = \max\left\{\begin{matrix}\max\limits_{e\in E}\{|(d_e)^{j+1} - (d_e)^j|, |(l_e)^{j+1} - (l_e)^j|, |(u_e)^{j+1} - (u_e)^j|\}, \\ \max\limits_{od\in OD}\left\{\left|(l^{od})^{j+1} - (l^{od})^j\right|, \left|(u^{od})^{j+1} - (u^{od})^j\right|\right\}\end{matrix}\right\}$$

$j \leftarrow j + 1$

5: **end while**

6: where $Iter_{max}$ is the maximum iteration number and $\underline{\eta}, \overline{\eta} \in R_+$ are predetermined tolerance.

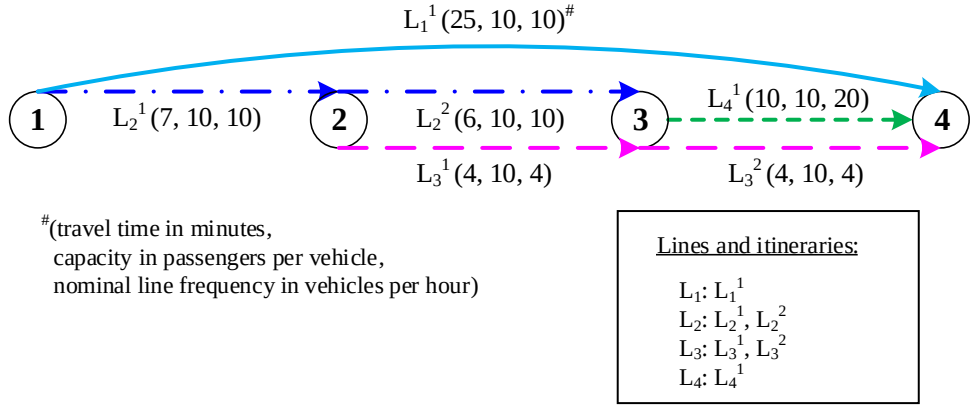
Figure 3.6 Algorithm of the iterative balancing scheme

3.5 Numerical Experiments

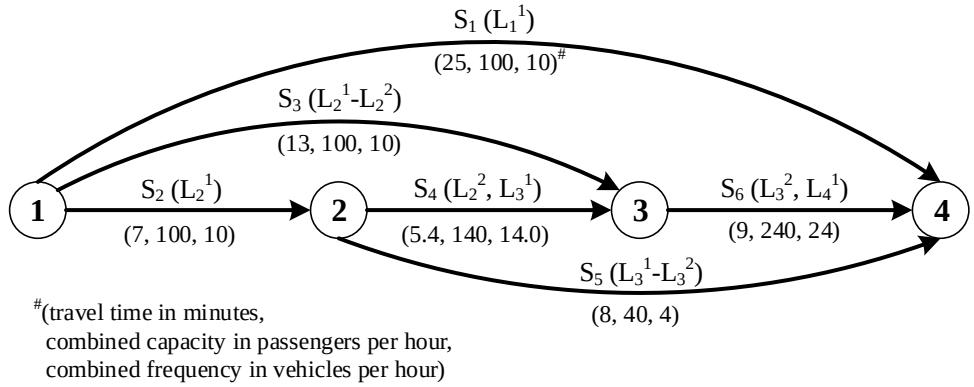
Three networks are used to demonstrate the features of the proposed model and the effectiveness of the solution algorithm. The first is a small example transit network, which is used to illustrate the correctness of the developed solution algorithm. The second one is a hypothetical transit network based on the Sioux Falls network, which is used to evaluate the performance of the proposed model with different configurations of onboard passenger counts and observed partial OD trip matrices from transit APC and AFC data. And the third is the transit network in Winnipeg, Canada, the purpose of which is to show the applicability of the proposed model and solution algorithm. In these three examples, the onboard passenger counts are assumed to be the values generated by the logit-based transit assignment model (See Chapter 2), and the observed partial OD trip matrices are assumed to capture a specific percentage of the true demand. The specific configuration can be found in each case. The measurement error ϵ_e allowed for the passenger count on transit line segment e is set as 5% for all measured transit line segments in these three examples. Finally, the parameter values are set as follows: $\alpha = 1$, $\theta = 0.1$, $\varphi = 10$, and $\vartheta = \varsigma = \varpi = 1$.

3.5.1 Small network

3.5.1.1 Network settings



(a) Network representation using line and itinerary description



(b) Network representation using route sections

Figure 3.7 Example of a small transit network

Table 3.2 Path information for each OD pair

OD	Path	
1-4	1	S_1
	2	S_2, S_5
	3	S_3, S_6
2-4	1	S_5
3-4	1	S_6

The network created by De Cea and Fernández (1993) is adopted to illustrate the performance of the proposed model for OD demand estimation in a congested transit network (Figure 3.7a). It consists of four transit lines (L_1, L_2, L_3 and L_4) and three OD pairs (1-4, 2-4 and 3-4). Figure 3.7b is an alternative representation of the same small example network in

terms of route sections, in which case the OD pairs are 1-4, 2-4 and 3-4. The basic data of the transit lines in the small example network and the basic characteristics of the route-section-based network are also given in Figure 7. The path information of each OD pair is shown in Table 3.2. This example is designed to demonstrate the performance of the proposed method. The tests are based on the following inputs. The true passenger OD demands are 200 (passengers/hour) for 1-4, 20 for 2-4 and 100 for 3-4. The "observed" transit line segment flows are generated through assigning the true passenger OD demands on the example transit network using the logit-based transit assignment. All of the transit line segments are assumed to have measured passenger counts in this example.

3.5.1.2 Results of small network experiment

To assess the combined effect of the onboard passenger counts and the observed partial OD flows on the estimation results, we design four scenarios:

- Scenario I: no observed partial OD flows and all onboard passenger counts
- Scenario II: 92% of true demand as observed partial OD flows and all onboard passenger counts
- Scenario III: 94% of true demand as observed partial OD flows and all onboard passenger counts
- Scenario IV: 96% of true demand as observed partial OD flows and all onboard passenger counts

Table 3.3 Estimated transit OD flows

		OD demands (pass/hr)			**RMSE**
		q_{1-4}	q_{2-4}	q_{3-4}	
	True	200	20	100	
Scenario I	Partial OD	-	-	-	
	Estimated	190	20.377	93.623	6.85
Scenario II	Partial OD	192	19.2	96	
	Estimated	192	19.2	96	5.18
Scenario III	Partial OD	188	18.8	94	
	Estimated	190	20	94	6.73
Scenario IV	Partial OD	184	18.4	92	
	Estimated	190	20.377	93.623	6.85

Table 3.4 Estimated transit line segment flows

		Transit line segment flows (pass/hr)					
		$v_{L_1^1}$	$v_{L_2^1}$	$v_{L_2^2}$	$v_{L_3^1}$	$v_{L_3^2}$	$v_{L_4^1}$
	Measured	111.513	88.487	61.911	46.576	73.561	134.926
Scenario I	Estimated	105.937	84.063	60.192	44.247	69.883	128.18
	ARE	5.00%	5.00%	2.78%	5.00%	5.00%	5.00%

Scenario II	Estimated	106.443	85.557	59.739	45.017	70.974	129.783
	ARE	4.55%	3.31%	3.51%	3.35%	3.52%	3.81%
Scenario III	Estimated	105.937	84.063	59.815	44.247	69.883	128.18
	ARE	5.00%	5.00%	3.39%	5.00%	5.00%	5.00%
Scenario IV	Estimated	105.937	84.063	60.192	44.247	69.883	128.18
	ARE	5.00%	5.00%	2.78%	5.00%	5.00%	5.00%

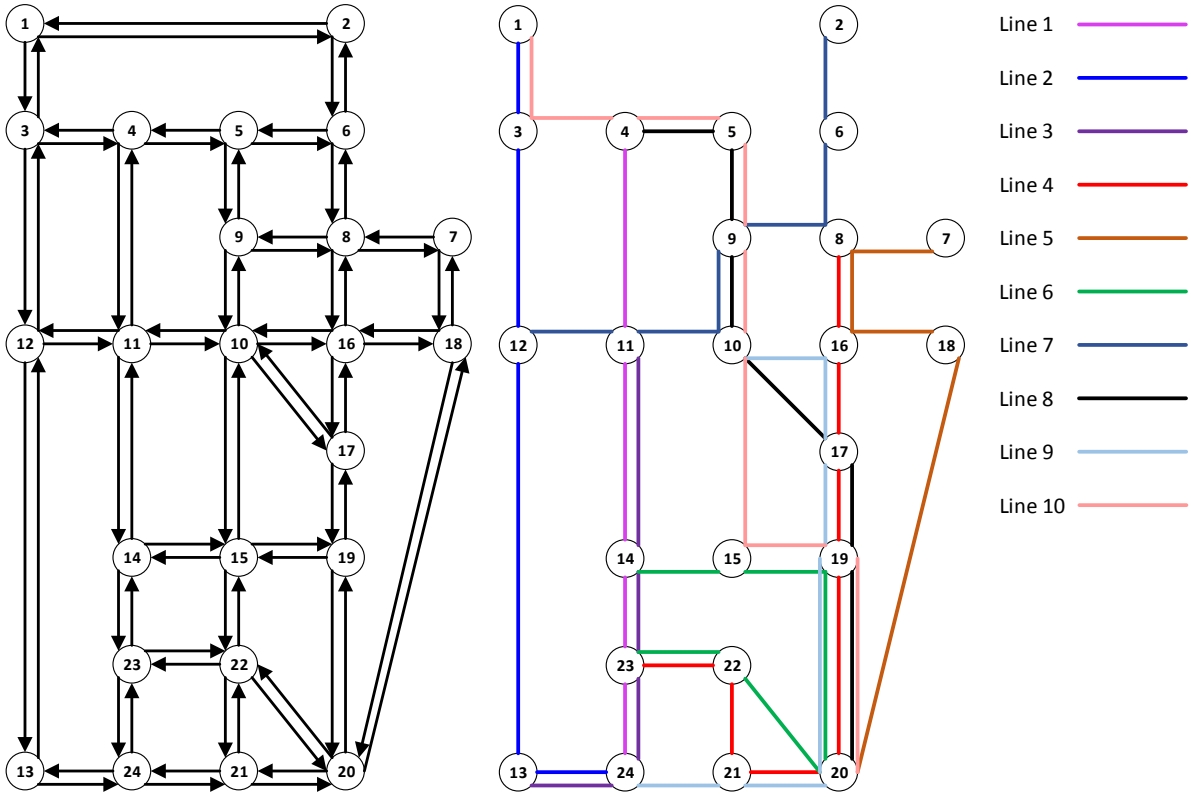
Table 3.3 presents the estimated OD matrices for the four scenarios, together with the observed partial OD flows. The root mean square error (RMSE) between the estimated OD flows and true OD demands is used to evaluate the estimated results. The scenarios can be ranked from lowest to highest RMSE as follows: Scenario II, Scenario III and Scenarios I and IV. Scenarios I and IV obtain the same estimated OD flows because the observed partial OD trip matrices in Scenario IV are too poor-quality to improve the estimated results, while only the onboard passenger counts have an effect on the estimated results same as that in Scenario I. Scenario II obtains the best result because the estimated OD trip matrices are equal to the observed partial OD flows due to the high quality of the observed partial OD flows and no effect of onboard passenger counts. The performance of Scenario III lies between those of Scenario II and Scenarios I/IV because both the onboard passenger counts and observed partial OD flows affect the estimated results. To demonstrate the accuracy of the estimated results, Table 4 presents the estimated results of transit line segment flows for all scenarios. The values of absolute relative error (ARE) are all in the interval of [0, 5%], which is consistent with the predetermined measure error ϵ_e.

3.5.2 Medium-size network

3.5.2.1 Network settings

The highway network of Sioux Falls (http://www.bgu.ac.il/~bargera/tntp/), shown in Figure 3.8, has 76 directed links and 24 nodes on which the itineraries of 10 lines (i.e., 20 itineraries) are defined (Sun and Szeto, 2018). It is assumed that all of the in-vehicle movements on the same highway link have identical travel times. The information of the frequencies, capacities and stop sequences of the lines is given in Table 3.5. The travel time of transit vehicles on the road network is assumed to equal that of private cars. Table 3.6 gives the details of 32 OD pairs with known and positive demands. Again, the counted flows are generated via assigning the true demands on the corresponding transit network using the logit-based transit assignment model. The observed partial OD matrix is obtained by multiplying the true demand by a scaling factor (e.g. 0.85). The product of the success rate and penetration rate

of using smartcards is set as 0.7. The transit path set is generated by adopting the generation algorithm in Chapter 2.3. The maximum number of transfers is 2, and the maximum number of paths between each OD pair is 30.



(a) Road network (b) Transit network

Figure 3.8 Sioux-Falls road and transit networks

Table 3.5 Transit line data for Sioux Falls network

Line	Line ID	Frequency (vehicle/hr)	Capacity (pass/vehicle)	Stop sequence
1	1	10	50	4 11 14 23 24
	2			24 23 14 11 4
2	3	10	50	1 3 12 13 24
	4			24 13 12 3 1
3	5	10	50	11 14 23 24 13
	6			13 24 23 14 11
4	7	12	50	8 16 17 19 20 21 22 23
	8			23 22 21 20 19 17 16 8
5	9	10	50	7 8 16 18 20
	10			20 18 16 8 7
6	11	10	50	14 15 19 20 22 23
	12			23 22 20 19 15 14
7	13	20	50	2 6 8 9 10 11 12
	14			12 11 10 9 8 6 2
8	15	20	50	4 5 9 10 17 19 20

	16			20 19 17 10 9 5 4
9	17	20	50	10 16 17 19 20 21 24
	18			24 21 20 19 17 16 10
10	19	20	50	1 3 4 5 9 10 15 19 20
	20			20 19 15 10 9 5 4 3 1

Table 3.6 True demand matrix for Sioux Falls

OD pair	Demand (pass/hr)	OD pair	Demand (pass/hr)	OD pair	Demand (pass/hr)	OD pair	Demand (pass/hr)
(1 - 13)	200	(3- 13)	200	(13 - 1)	200	(21 - 1)	200
(1 - 20)	200	(3 - 20)	200	(13 - 2)	200	(21 - 2)	200
(1 - 21)	200	(3 - 21)	200	(13 - 3)	200	(21 - 3)	200
(1 - 24)	200	(3 - 24)	200	(13 - 4)	200	(21 - 4)	200
(2 - 13)	200	(4 - 13)	200	(20 - 1)	200	(24 - 1)	200
(2 - 20)	200	(4 - 20)	200	(20 - 2)	200	(24 - 2)	200
(2 - 21)	200	(4 - 21)	200	(20 - 3)	200	(24 - 3)	200
(2 - 24)	200	(4 - 24)	200	(20 - 4)	200	(24 - 4)	200

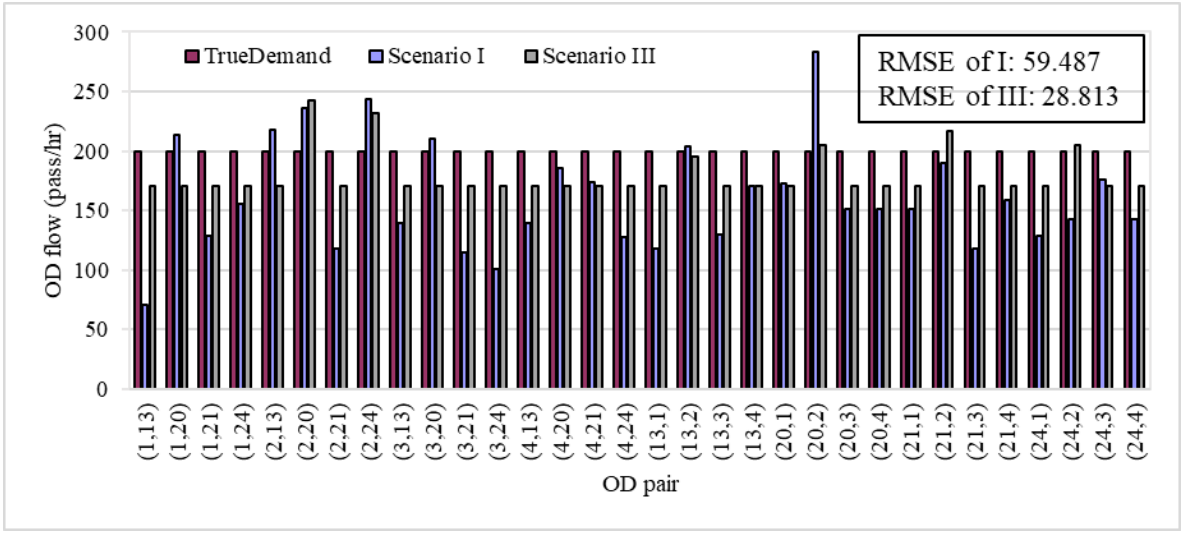
3.5.2.2 Effect of data availability on the estimated results

Four scenarios are designed to illustrate the performance of the proposed model and its dependence on the availability of count information for the transit line segments and observed partial OD trip matrices:

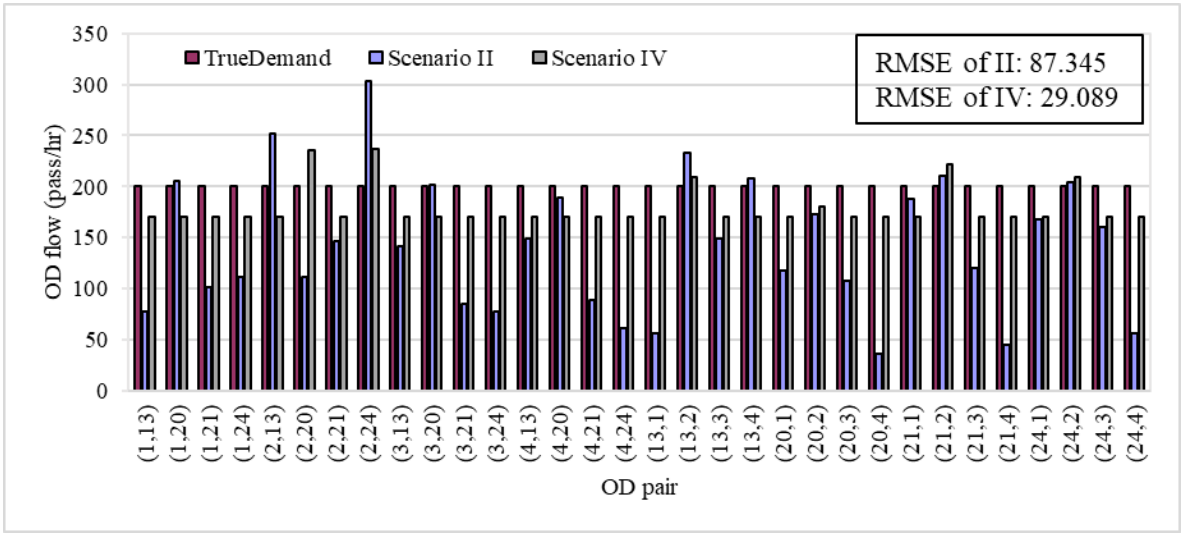
- **Scenario I**: using the count information of all of the transit line segments as the available counts
- **Scenario II**: using the count information of the 50% of the transit line segments with the most counted flows
- **Scenario III**: Scenario I with observed partial OD trip matrices
- **Scenario IV**: Scenario II with observed partial OD trip matrices

Scenarios I and II are designed to compare the effect of the degree of count information availability, and Scenarios III and IV illustrate the effect of including observed partial OD trip matrices. Figure 3.9 shows the estimation results of these four scenarios. In general, the scenarios with observed partial OD trip matrices (Scenarios III and IV) yield better results than those without (Scenarios I and II). Specifically, the RMSEs between each scenario I and the true demand are 59.487, 87.345, 28.813 and 29.089 for Scenarios I, II, III and IV, respectively. We also observe that the scenarios with the count information of all transit line segments (Scenarios I and III) yield better results than those with only half of this information (Scenarios II and IV). Figure 3.10 and 3.11 display the estimated transit line segment flows of line 4 for these four scenarios. For the scenarios with the count information of all of the transit line

segments (Scenarios I and III), the estimated transit line segment flows remains within the [-5%, +5%] area of the observed counts. However, for Scenarios II and IV, some of the estimated segment flows fall outside the [-5%, +5%] area of the observed counts. These transit line segments are among those for which the count information is not included in the side constraints, so the observed data are not available to minimize the error.

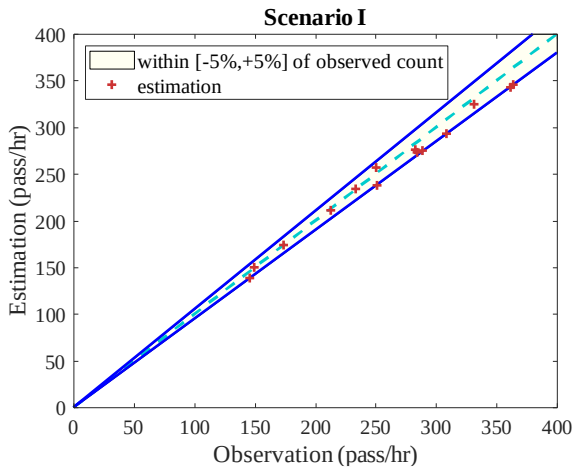


(a) Comparison of estimation accuracy between Scenarios I and III

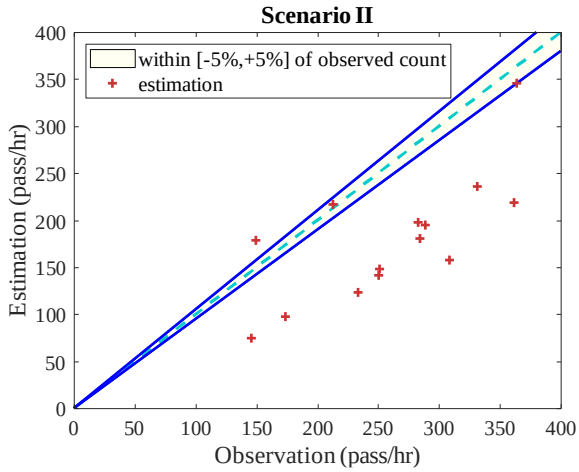


(b) Comparison of estimation accuracy between Scenarios II and IV

Figure 3.9 Estimation of four scenarios on the Sioux Falls network

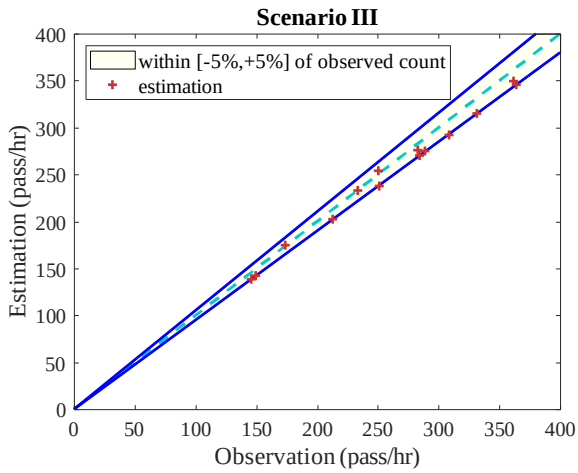


(a) Scenario I

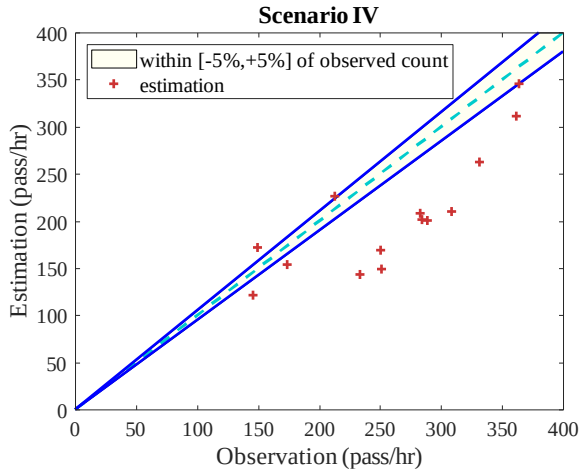


(b) Scenario II

Figure 3.10 Estimated vs. observed line segment flow of transit line 4 (both directions, Scenario I and II)



(a) Scenario III



(b) Scenario IV

Figure 3.11 Estimated vs. observed line segment flow of transit line 4 (both directions, Scenario III and IV)

3.5.2.3 Effect of different levels of the observed partial OD trip matrix

This section mainly examines the effect of different values of the observed OD trip matrix on the estimation. We vary the value of the partial OD trip matrix obtained from AFC data from 130 to 180 in intervals of 10 (pass/hr). Figure 3.12 displays the RMSE values of the estimated OD flows with different values of the observed partial OD trip matrix. The estimation results improve with increasing values of the partial OD trip matrix. This is expected because, in our proposed model, the observed partial OD trip matrix is used as a side constraint to obtain

the optimal solutions. When the matrix is poor-quality, the error bound within which the estimated results fall is large, which increases the probability of inaccurate results.

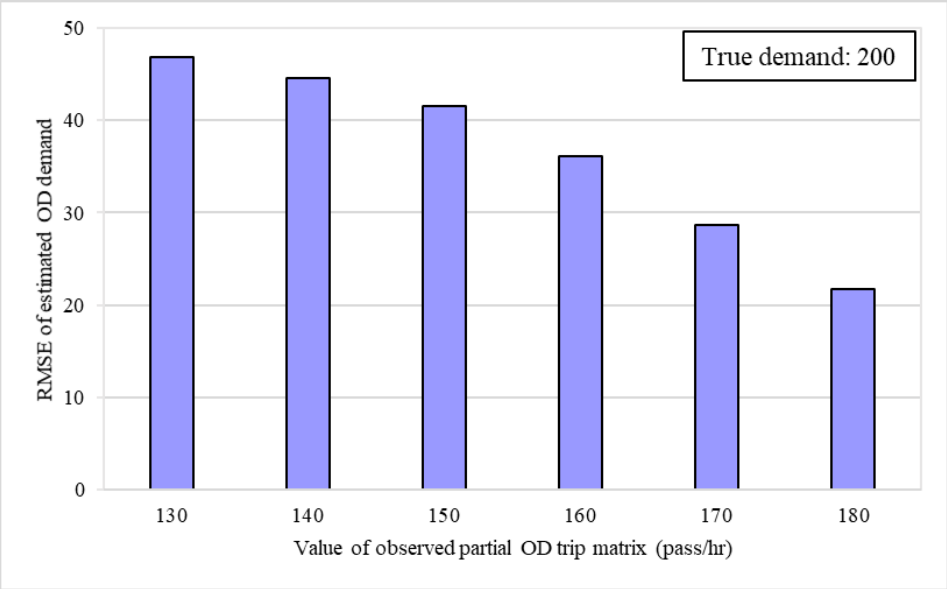


Figure 3.12 Effect of different values of the observed partial OD trip matrix

3.5.2.4 Effect of number of transfers

As the number of transfers is accounted for in the transit path set generation, we examine three path set strategies with different numbers of transfers. The maximum number of paths for all three strategies is 20.

- Strategy I: maximum of one transfer
- Strategy II: maximum of two transfers
- Strategy III: no maximum number of transfers

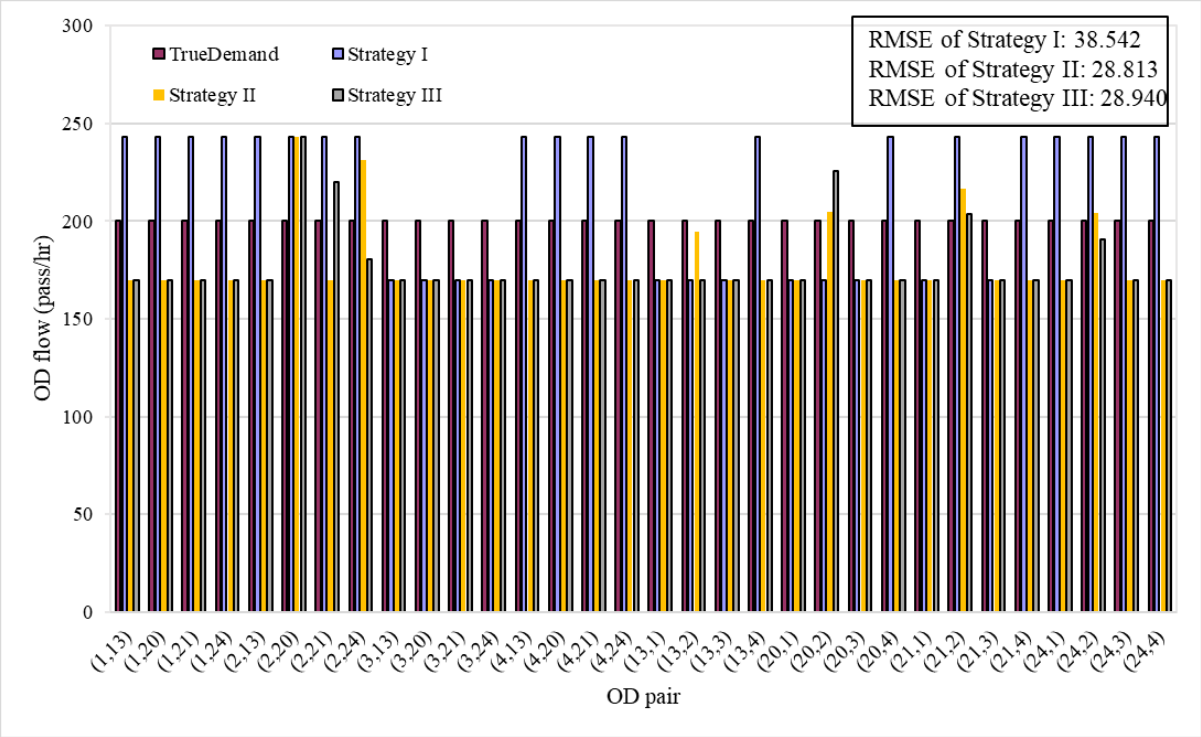


Figure 3.13 Effect of the number of transfers on the estimation results

Figure 3.13 compares the estimated OD flows obtained from three different path sets with the above three strategies. As expected, Strategy II performs the best because the observations of partial OD trip matrices and onboard passenger counts are generated by logit-based transit assignment model using strategy II. Interestingly, Strategy III performs only slightly worse than Strategy II, while Strategy I performs much worse than Strategies II and III. Because Strategy I allows a maximum of one transfer for the transit path, the transit path set narrowed considerably, which has a major effect on the resulting flow patterns.

3.5.3 Large network

This subchapter applies the proposed model to a real-world transit network in the city of Winnipeg, Canada. Shown in Figure 3.14, the transit network is extracted from Emme V4.3.2. It consists of 130 transit lines, 4187 transit line segments and 924 transit stops. The number of origins is 106, and that of OD pairs is 5303. To connect the origins and the transit stops, the network also contains 803 walking segments. After conducting the route-section network construction, the network consists of 44408 route sections, in which each walking segment is seen as a special case of a transit route section with the cost set as 0. The solution algorithm is implemented in Microsoft Visual Studio 2015 and run on a 2.7 GHz processor with 20.00 GB of RAM.

The transit line segments with volume/capacity (V/C) ratio larger than 0.1 are included as the observation constraints. The observed demand inferred from smart card data is assumed to be 85% of the true demand. The product of the transaction record use rate ϱ and smartcard penetration rate Υ is again assumed to be 0.7. Thus, the upper bound of the estimated demand is 1.214 times the true demand.

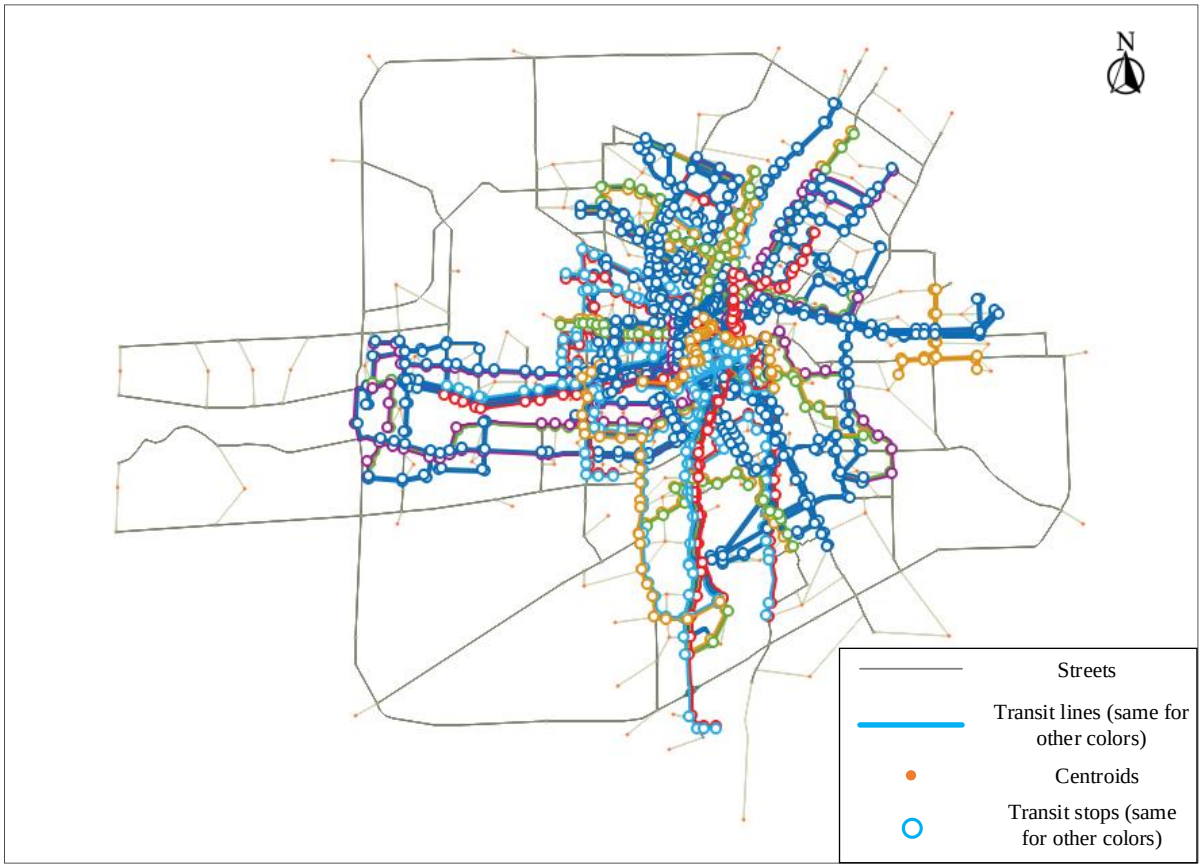


Figure 3.14 Winnipeg transit network: different colors denote different lines

To assess the agreement between the estimated and observed values, we provide a scatter plot in Figure 3.15 to compare the transit line segment flows. As can be seen, most of the estimated transit line segment flows fall within the area of [-5%, +5%] of the observed counts (i.e. the light blue dashed line at 45° represents that estimations equal observations), and the estimated OD flow is also within the error bound of the true demand (determined by the observed partial OD flow and the success rate-dependent and penetration rate-dependent upper bound).

Figure 3.16 displays the values of ARE for the estimated OD trip matrices. Compared with the true demand, 31.59% of the OD pairs reach the lower bound of the estimated demand constraints (ARE = 0.15), and 55.48% reach the upper bound (ARE = 0.2143). This illustrates

again that the estimated OD flows are mostly confined within the area of lower and upper bounds. However, the estimated results for 12.94% of the OD pairs are below the upper bound or above the lower bound, which implies that the estimated results would improve with the inclusion of the observed onboard passenger counts of the transit line segments. Of course, the performance would then depend on the quality of the observed OD flows and onboard passenger counts of line segments, but this does not affect the applicability of the proposed model. Moreover, the RMSE of the estimated results is 1.1602, which is relatively acceptable.

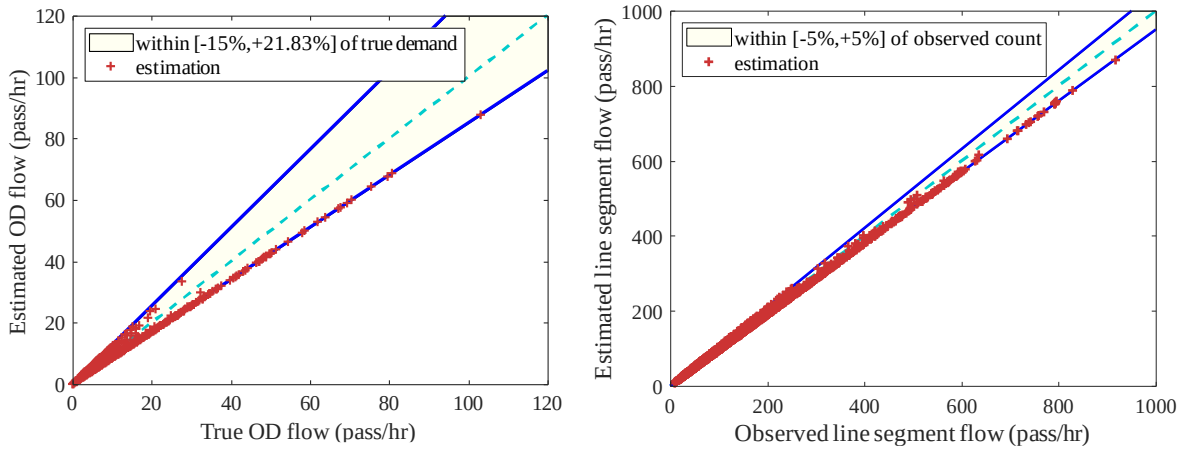


Figure 3.15 Comparison of the observed and estimated line segment flows for the Winnipeg network

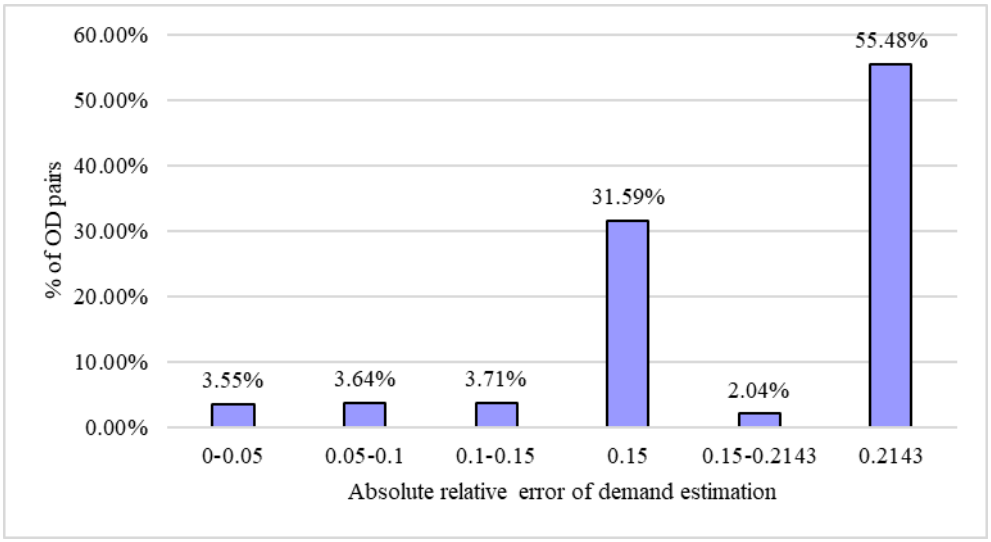


Figure 3.16 Distribution of the absolute relative error of demand estimation

3.6 Chapter Summary

In this chapter, we propose a frequency-based PFE framework for OD trip matrix estimation in transit networks. A frequency-based transit PFE formulation in variational inequality form is proposed, incorporating observed partial OD demand matrices and onboard passenger flow observations as side constraints. The observed partial OD demand matrix is inferred from the smartcard (AFC) data together with AVL data based on inference of the alighting stops, while the observed onboard passenger flow is calculated from APC data and AVL data. To solve the proposed model, the diagonalization method is adopted, and the diagonalized subproblem is solved by a path-based partial linearization solution algorithm embedded with an iterative balancing scheme to handle the various side constraints.

Numerical examples are provided to illustrate the performance of the proposed model and its applicability in a real-world transit network. The results show that the configurations of onboard passenger counts and observed partial OD flows affect the estimated results. Overall, the results of a large-network indicate that the estimation of OD flows can be improved by using onboard passenger counts and observed partial OD flows together.

To the best knowledge of the authors, this is the first attempt to build a framework integrating congested transit choice behavior, APC data and AFC data to estimate transit OD flow. The model presented in this chapter is mainly designed for transit network-level OD flow estimation. Of course, route-level estimation could be seen as a special case of our proposed model. Although the numerical examples in this chapter do not use real APC and AFC data, this does not affect the applicability of the proposed model, because real APC and AFC data can be transformed to the onboard passenger counts and observed partial OD trip matrices in real-world applications.

CHAPTER 4

MULTI-MODAL PATH FLOW ESTIMATOR FOR ESTIMATING OD DEMAND IN URBAN TRANSPORTATION NETWORKS

This paper proposes an approach to estimate multi-modal origin-destination (OD) trip matrices in urban transportation networks. The model, called multi-modal path flow estimator (MM-PFE), is formulated as a variational inequality (VI) problem based on a single-level structure; moreover, it incorporates limited available observations (i.e., road link traffic counts, onboard passenger counts of bus and metro line segments, the mode-specific target OD demand, and zonal production and attraction) as side constraints. The interactions of private cars and bus vehicles, car and transit mode choice behaviours, and mode similarity are modeled in the congested network. A nested logit model is adopted for the mode choice, and a multinomial logit model is used for the route choice. To solve the MM-PFE problem, a diagonalization approach is adopted; in each diagonalized iteration, the VI-based MM-PFE problem is reformulated as a convex optimization problem, which is solved using a developed path-based partial linearization algorithm. To handle various inequality/equality side constraints, a three-layer iterative balancing scheme is developed to obtain the adjustment factors for updating dual variables based on the duality theory. Finally, a computational test on the proposed model and the developed solution algorithm is conducted using data from the hypothetical multi-modal transportation network of Sioux Falls.

4.1 Introduction

Multi-modal transportation provides people with multiple substitutable modes of travel (e.g., car, metro, bus) for between their origins and their destinations, and it has become increasingly popular in most cities around the world. As a sustainable alternative model to private transport, public transport plays an important role in travel demand sharing in many large cities. For example, in Hong Kong, approximately 90% of daily trips are made using multiple public transport modes. The prevalence of public transport demonstrates the importance of conducting multi-modal transportation network analyses, such as transport policies evaluation, for urban transportation planning and management. However, these analyses are primarily focused on multi-modal travel demand forecasting (Oppenheim, 1995; Kitthamkesorn et al., 2016; Wang et al., 2018). For this forecasting, the origin-destination (OD) trip table is a critical input that can rarely be directly obtained from the real world. The quality of the OD demand has an important impact on travel demand model accuracy.

In the literature, almost all of the existing OD demand estimation models focus on estimating the OD matrix of a single mode for urban transportation systems. Several studies have been conducted on traffic and transit OD demand estimations, as seen below:

- For traffic OD demand estimation in road networks, bi-level models (Fisk, 1988; Yang, 1995; Yang et al., 2001; Lundgren and Peterson, 2008) and path flow estimators (PFEs) (Bell and Iida, 1997; Chen et al., 2005, 2009, 2010) are the two most used types of methods. Regarding bi-level models, in the upper-level, the travel demand matrix is estimated using a least-squares formulation, generalized least-squares function, or maximum likelihood/entropy function, while in the lower-level, the traffic assignment problem is modeled with the user equilibrium principle. The critical drawback of these models is that the heuristic solution algorithms cannot necessarily converge to the global optimal solution. Regarding the path flow estimator (PFE), the multinomial logit-based stochastic user equilibrium principle is adopted, and its solution algorithm can converge to a unique global optimal solution, fundamentally owing to the single-level structure.
- For transit OD demand estimation, the methods can be classified into data-based models (Barry et al., 2002; Zhao et al., 2007; Trépanier et al., 2007; Munizaga and Palma, 2012) and network-based models (Lam et al., 2003; Wu and Lam, 2006; Babazadeh et al., 2010). The data-based models use automatic passenger count data, automatic fare collection data, and automatic vehicle location data to estimate a demand matrix via trip chain

reconstruction. However, these models can only obtain an incomplete demand matrix owing to the penetration rate of smartcards and the success rate of alighting stops identifications. Regarding the network-based model, bi-level formulations with the same structure as traffic OD demand estimation models are widely used.

Even though the mode-specific OD demand can be independently estimated, the mode choice behavior can rarely be guaranteed, e.g. multinomial logit (MNL) model (Oppenheim, 1995; Wu and Lam, 2003; Wang et al., 2018), nested logit (NL) model (Ben-Akiva and Lerman, 1985; Kitthamkesorn et al., 2016), cross-nested logit model (Vovsha, 1997) and nested weibit model (Kitthamkesorn and Chen, 2017). Mode choice is one of the most critical components of the travel demand modeling process, whereby the OD demand is split into trips using car, transit, or other emerging travel modes (e.g., ridesharing platforms). Forecasting the mode-specific travel demand independently will ignore the interaction of vehicles from different modes (e.g., private cars and bus vehicles) on the road network. This kind of interaction will affect the travel cost (disutility) and consequently influence people's degree of satisfaction with the mode choice. This implies that we cannot simply conduct a mode-specific network equilibrium analysis independently in a multi-modal transportation system. Thus, systematic demand modeling methods are needed to estimate the urban multi-modal travel demand.

However, a few studies have focused on the estimation of multi-modal OD matrices. The current practices in estimating multi-modal OD matrices use a four-step model based on trip rates; a sequential framework consisting of trip generation, trip distribution, modal split and traffic assignment is adopted. This practical sequential procedure has several drawbacks: (1) it requires iterative feedback mechanisms to obtain consistent solutions of various flow patterns (e.g. OD demand, mode-specific OD demand, mode-specific path flow, link flow) at different spatial levels; (2) it cannot utilize the information contained in the observations; and (3) its operation usually requires a lengthy calibration process and specialized technical staffs. Furthermore, with the concept of combined network equilibrium, García-Ródenas and Marín (2009) proposed a calibration and demand adjustment model based on bi-level programming for the simultaneous estimation of an OD matrix and its parameters. Owing to the poor mathematical properties of the bi-level model, the authors developed a heuristic column generation algorithm by reformulating the bi-level model into a single-level one. Unfortunately, the heuristic nature of the solution algorithm still cannot be ignored, and the algorithm does not obtain a global optimal solution.

Therefore, this paper aims to explore the multi-modal OD matrix estimation problem via a single-level model. Specifically, we propose an NL-based multi-modal path flow estimator

(MM-PFE) to estimate the OD matrix in an urban congested transportation network. The interaction of bus vehicles and private cars on the road network is considered for travel cost modeling, and strategy-based transit behavior is incorporated for the transit component. Moreover, various kinds of data sources are incorporated to estimate the OD trip matrix, namely zonal production and attraction flows, a target OD mode-specific trip table, and mode-specific link flow observations. These data sources will serve as side constraints, which will help to reproduce the flow patterns according to the observations.

An NL model is used for the mode choice to handle the mode similarity issue, and an MNL model is used for the route choice. Owing to the asymmetric cost function, the MM-PFE is formulated as a variational inequality (VI) problem. Then, a diagonalized approach is adopted to solve the VI formulation of the MM-PFE, and the diagonalized convex optimization problem is solved using a path-based partial linearization algorithm embedded with a self-regulated averaging scheme and iterative balancing scheme.

The contributions of this paper can be summarized as follows: (a) an NL-based PFE for urban multi-modal OD demand estimation problem is proposed; (b) a diagonalization algorithm with a three-level iterative balancing scheme is developed to solve the MM-PFE; (c) the proposed model and developed algorithm are tested in the hypothetical multi-modal transportation network of Sioux Falls.

4.2 Multi-Modal Transportation Network Modelling

4.2.1 Notations

This subchapter provides a list of notation used in this chapter unless otherwise specified.

Sets	
O	set of origins, $O \subseteq N$
D	set of destinations, $D \subseteq N$
OD	set of origin-destination (OD) pairs, $od \subseteq OD$
U^{od}	set of nests connecting OD pair od
$\breve{U}_u^{od}$	set of modes among nest u connecting OD pair od, $\breve{U} = \{c, b, m\}$
$K_{u\breve{u}}^{od}$	set of routes of mode $\breve{u}$ among nest u connecting OD pair od
E_b	set of line segments of bus mode b
E_m	set of line segments of metro mode m
S	set of arcs in the route-section-based multi-modal transportation network, i.e., $S = S_c \cup S_b \cup S_m$
S_c	set of road links of car mode c
S_b	set of route sections of bus mode b
S_m	set of route sections of metro mode m
$\bar{S}_c, \bar{\bar{S}}_c$	set of measured and unmeasured road links of car mode c, i.e., $S_c = \bar{S}_c \cup \bar{\bar{S}}_c$
$\bar{E}_b, \bar{\bar{E}}_b$	set of measured and unmeasured line segments of bus mode b, i.e., $E_b = \bar{E}_b \cup \bar{\bar{E}}_b$

$\bar{E}_m, \bar{\bar{E}}_m$	set of measured and unmeasured line segments of metro mode m, i.e., $E_m = \bar{E}_m \cup \bar{\bar{E}}_m$
Parameters and inputs	
$\theta_{u\breve{u}}^{od}$	dispersion parameter for route choice of mode $\breve{u}$ among nest u connecting OD pair od
ρ_u^{od}	degree of independence in unobserved utility among the alternatives in nest u connecting OD pair od
$C_{c,s}$	capacity on road link $s \in S_c$
$C_{b,e}$	capacity on bus line segment $e \in E_b$
$C_{m,e}$	capacity on metro line segment $e \in E_m$
f_l	frequency of transit line l
$\kappa_{b,l}$ ($\kappa_{m,l}$)	capacity of line l for transit mode b (m)
$\bar{v}_{c,s}$	traffic count on measured road link $s \in \bar{S}_c$
$\bar{v}_{b,e}$	onboard passenger count on measured bus line segment $e \in \bar{E}_b$
$\bar{v}_{m,e}$	onboard passenger count on measured metro line segment $e \in \bar{E}_m$
$z_{\breve{u}u}^{od}$	observed OD flow of mode $\breve{u}$ among nest u connecting target OD pair od
z^o	observed trip production of origin o
z^d	observed trip attraction of destination d
Intermediate variables	
c_s	travel time on arc $s \in S$
$c_{u\breve{u},k}^{od}$	travel cost of route k of mode $\breve{u}$ among nest u connecting OD pair od
$v_{c,s}$	flow on road link $s \in S_c$
$v_{b,s}$	flow on route section $s \in S_b$
$v_{m,s}$	flow on route section $s \in S_m$
$v_{b,e}$	flow on bus line segment $e \in E_b$
$v_{m,e}$	flow on bus line segment $e \in E_m$
q^{od}	OD flow between origin o and destination d
P^o	trip production of origin o
A^d	trip attraction of destination d
Decision variables	
$h_{u\breve{u},k}^{od}$	flow on route k of mode $\breve{u}$ among nest u connecting OD pair od
$q_{u\breve{u}}^{od}$	OD flow of mode $\breve{u}$ among nest u connecting OD pair od
q_u^{od}	OD flow of nest u connecting OD pair od

4.2.2 Illustration of interaction between road and bus networks

In the multi-modal transportation network, vehicles of road-based transit systems (i.e., bus network in this study) interact with private cars on the road network (Figure 4.1). This means that their travel time on the road links is affected by not only their own flow but also the flow of the other modes (e.g., bus vehicles). Moreover, the road link connecting two adjacent bus stops on a specific bus line usually contains more than one element. This means that the in-vehicle travel time of the bus line segment is the summation of the bus travel times on these road links. For example, in Figure 4.1, line segment (n1-n2) of bus line 1 contains two road links, 2 and 4. The relationships among route section, section line segment, and line segment are also demonstrated, which is consistent with the definition in Chapter 2.2.2.

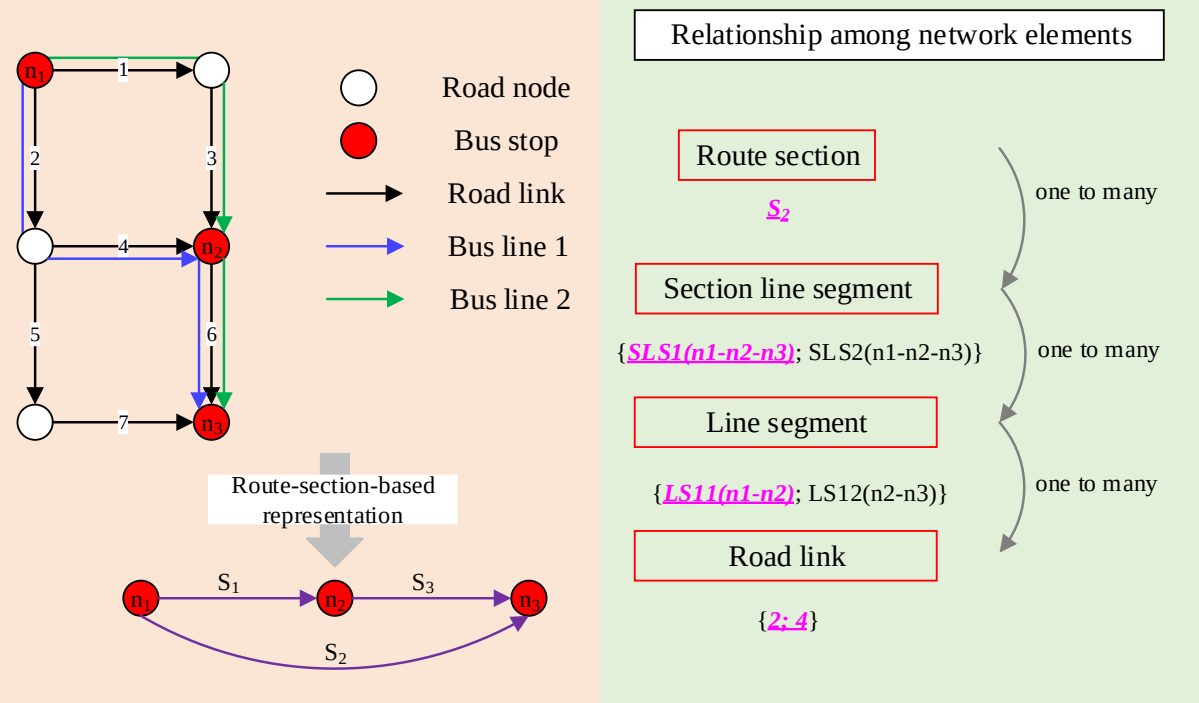


Figure 4.1 Relationship between road and bus networks

4.2.3 Path travel time for cars

The travel time of cars on the road network is affected by bus vehicles running on the same road link; therefore, we need to first obtain the flow of private cars and bus vehicles on that road link. The flow of private cars on the road link can be obtained by simply summing up the flows of the paths of all OD pairs going through the road link:

$$v_{c,s} = \sum_{od \in OD} \sum_{k \in K_c^{od}} h_{c,k}^{od} \eta_{sk}^{od}, \forall s \in S_c \tag{4.1}$$

Assuming that the frequency of each transit line is fixed, the passenger car equivalent flow of bus vehicles is the summation of the bus vehicles of bus lines passing through the road link multiplied by the value of the passenger car equivalent:

$$F_{c,s} = \sum_{l' \in \Psi_s} f_{l'} \cdot PCE \tag{4.2}$$

where Ψ_s is the set of bus lines going through road link s, and PCE is the passenger car equivalent for a bus vehicle.

Thus, the link travel time of the car in the road network is:

$$c_{c,s} = t_{c,s}^0 \left[1 + \gamma\left(\left(F_{c,s} + v_{c,s}\right)/C_{c,s}\right)^{\beta}\right], \forall s \in S_c \tag{4.3}$$

where $t_{c,s}^0$ is the free-flow travel time of car on road link s, $C_{c,s}$ is the capacity of road link s, and γ and β are parameters for the travel time function.

With the travel time of the car on road link, the travel time of path k between OD pair od for mode c can be expressed as:

$$c_{c,k}^{od}(\mathbf{v}) = \sum_{s \in S_c} a_{sk} c_{c,s}(\mathbf{v}), \forall k \in K_c^{od}, od \in OD \tag{4.4}$$

4.2.4 Path travel time for bus and metro

Using the route-section-based transit network representation in Chapter 2.2.2, we formulate the path travel time for bus and metro. The route section cost consists of in-vehicle travel time, waiting time, and perceived congestion time. Here we use a unified symbol $\bar{u} \in \bar{U} = \{b, m\}$ to represent the mode. For route section s of transit mode $\bar{u}$, the expected total travel time is given by:

$$c_{\bar{u},s}(\mathbf{v}) = t_{\bar{u},s}(\mathbf{v}) + w_{\bar{u},s} + \phi_{\bar{u},s}(\mathbf{v}), \forall \bar{u} \in \bar{U} \tag{4.5}$$

The three components of the route section cost in Eq. (4.5) are specified below:

- In-vehicle travel time

The allocation ratio between section line segments and route section is directly proportional to their frequencies:

$$x_{\bar{u},s}^{l} = \frac{f_l}{\sum_{j \in A_{\bar{u}}^s} f_j}, \forall l \in A_{\bar{u}}^s, s \in S_{\bar{u}}, \bar{u} \in \bar{U} \tag{4.6}$$

(1) Bus: Based on the discussion in Chapter 4.2.2, the in-vehicle travel time of a bus is associated with the car and bus vehicle flows on the road network:

$$\bar{t}_{b,s} = t_{b,s}^0 \left[1 + \gamma\left(\left(F_{c,s} + v_{c,s}\right)/C_{c,s}\right)^{\beta}\right], \forall s \in S_c \tag{4.7}$$

where $t_{b,s}^0$ is the free-flow travel time of the bus vehicle on road link s.

Note that each bus line segment (between two adjacent stops) will usually pass across several road links; accordingly, the in-vehicle travel time of the bus line segment is the summation of the travel times on the different road links:

$$\tilde{t}_{b,l}^{s} = \sum_{s' \in \mho(s,l)} \bar{t}_{b,s'}, \forall l \in A_b^s, s \in S_b \tag{4.8}$$

where $\mho(s, l)$ denotes a set of directed road links for bus line l, which is included in the set of attractive bus lines associated with route section s; that is, $l \in A_b^s$.

As each route section consists of one or several section line segments, the in-vehicle travel time of route section s is the weighted summation of those of the section line segments:

$$t_{b,s} = \sum_{l \in A_b^s} \tilde{t}_{b,l}^s x_{b,s}^l, \forall s \in S_b \tag{4.9}$$

(2) Metro: The in-vehicle travel time of route section s in a metro network is defined as:

$$t_{m,s} = \sum_{l \in A_m^s} t_{m,l}^0 x_{m,s}^l, \forall s \in S_m \tag{4.10}$$

where $t_{m,l}^0$ is the in-vehicle travel time on section line segment l of the metro network, and A_m^s is the set of attractive metro lines.

- Waiting time

The waiting time of route section s is defined as the product of α and the inverse of its combined frequency:

$$w_{\bar{u},s} = \frac{\alpha}{\sum_{l \in A^s} f_l}, \forall s \in S_{\bar{u}}, \bar{u} \in \bar{U} \tag{4.11}$$

where $\alpha = 1$ means that the transit vehicle headway follows an exponential distribution, and $\alpha = 0.5$ means that it follows a uniform distribution.

- Perceived congestion time

The perceived congestion time of the route section involves the additional waiting time due to vehicle congestion, which is a function of its own flow and that of its competing route sections. First the flow on route section s of transit mode $\bar{u}$ is given as:

$$v_{\bar{u},s} = \sum_{od \in OD} \sum_{k \in K_{\bar{u}}^{od}} a_{\bar{u},sk} h_{\bar{u},k}^{od}, \forall s \in S_{\bar{u}}, \bar{u} \in \bar{U} \tag{4.12}$$

where the path-section incidence $a_{\bar{u},sk}$ equals 1 if section s lies on path k; otherwise, it equals 0.

There are three main groups of passengers competing with route section s (de Cea and Fernández, 1993) and the competing section flow is given as:

$$\tilde{v}_{\bar{u},s} = \sum_{\check{s} \neq s \in S_{\bar{u}}} \delta_s^{\check{s}} \sum_{l \in A_{\bar{u}}^s \cap A_{\bar{u}}^{\check{s}}} v_{\bar{u},\check{s}l}, \forall s \in S_{\bar{u}}, \bar{u} \in \bar{U} \tag{4.13}$$

where $\delta_s^{\check{s}}$ is the competing section indicator, such that $\delta_s^{\check{s}} = 1$ means that section $\check{s}$ is a competing section of section s, and otherwise $\delta_s^{\check{s}} = 0$.

The section line segment flow $v_{\bar{u},sl}$ in Eq. (4.13) is determined by:

$$v_{\bar{u},sl} = v_{\bar{u},s} x_{\bar{u},s}^l, \forall l \in A_{\bar{u}}^s, s \in S_{\bar{u}}, \bar{u} \in \bar{U} \tag{4.14}$$

Then the perceived congestion time function for route section s of transit mode $\bar{u}$ is expressed as:

$$\phi_{\bar{u},s}(\mathbf{v}) = \varphi_{\bar{u},s}\left(\frac{\lambda v_{\bar{u},s} + \varsigma \tilde{v}_{\bar{u},s}}{\sum_{l \in A_{\bar{u}}^{s}} f_l \, \kappa_{\bar{u},l}}\right)^{\zeta}, \forall s \in S_{\bar{u}}, \bar{u} \in \bar{U} \tag{4.15}$$

where parameters ϑ, ς, $\varphi_{\bar{u},s}$, and ϖ are used to represent different effects of various flows on the perceived congestion time, and $\kappa_{\bar{u},l}$ is the capacity of the transit vehicle.

With the expected route section time function, the expected travel time associated with path k of transit mode $\bar{u}$ between OD pair od can be expressed as:

$$c_{\bar{u},k}^{od}(\mathbf{v}) = \sum_{s \in S_{\bar{u}}} a_{\bar{u},sk} c_{\bar{u},s}(\mathbf{v}), \forall k \in K_{\bar{u}}^{od}, \bar{u} \in \bar{U}^{od}, od \in OD \tag{4.16}$$

4.3 Multi-Modal Path Flow Estimator Formulation

4.3.1 Mode and route choice modelling

To consider the similarities between the bus and metro modes, we use an NL choice model to model the mode choice (Train, 2003). Three important definitions are presented first:

- *Nest*. We define the nest as a set consisting of one mode or several similar (correlated) modes.
- *Mode*. The specific travel mode in reality, e.g., car, metro, and bus.
- *Route*. For the car mode, route is a sequence of road links for each OD pair, whereas for the metro and bus modes, route is a sequence of route sections.

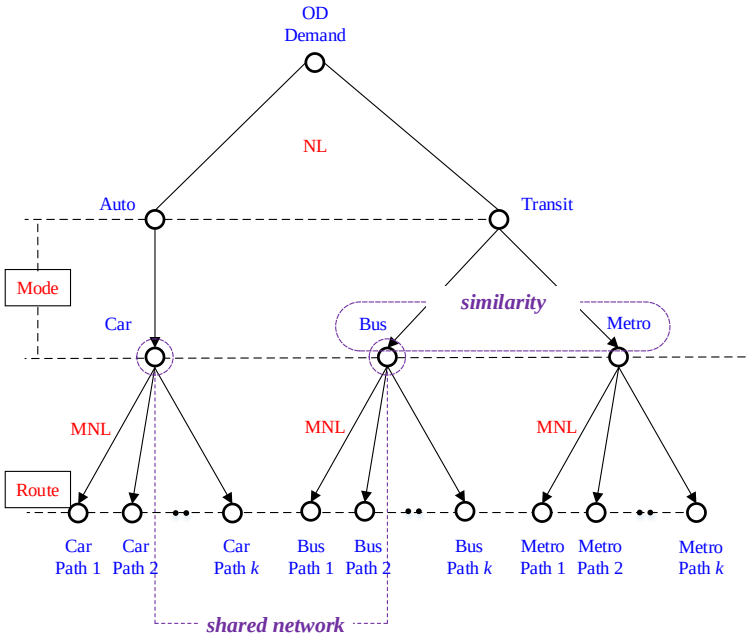


Figure 4.2 Illustration of mode and route choice

In the NL model, the upper nest consists of auto and transit. Here, auto contains only the private car mode, while transit contains the bus and metro modes, which are similar. This means that the combination of the bus and metro modes will compete with the car mode. In other words, car, bus and metro are not seen as three independent alternatives for travelers. Regarding the path (route) choice, we adopt the MNL model to address the issue of stochastic perception variance of travelers regarding path travel time. A detailed framework of the mode and route choices in a multi-modal transportation network is displayed in Figure 4.2.

To describe this three-layer structure of mode and route choices for each OD pair od, we introduce the following four probabilities:

- The marginal probability of choosing a nest (p_u^{od})
- The conditional probability of choosing a mode with a given nest ($p_{\breve{u}|u}^{od}$)
- The probability of choosing a mode ($p_{u\breve{u}}^{od} = p_{\breve{u}|u}^{od} \cdot p_u^{od}$)
- The probability of choosing a path for each mode ($p_{u\breve{u},k}^{od}$)

The first three probabilities are for mode choice, and they are related as follows: $p_{u\breve{u}}^{od} = p_{\breve{u}|u}^{od} \cdot p_u^{od}$. The fourth probability is for route choice. Moreover, we model the interactions of private cars and buses sharing the same roadway space on a road network. Thus, their travel times on road link will be affected by each other and not only by their own flows.

4.3.2 Flow conservation in a multi-modal transportation network at different spatial levels

Based on Figure 4.2, the relationship between the OD demand q^{od} and the upper-nest-specific OD demand q_u^{od} ($u \in U^{od}, U^{od} = \{auto, transit\}$) is expressed as follows:

$$\sum_{u \in U^{od}} q_u^{od} = q^{od}, od \in OD \tag{4.17}$$

Moreover, the relationship between the upper-nest-specific OD demand q_u^{od} and the lower-nest-specific (i.e., mode-specific) OD demand $q_{u\breve{u}}^{od}$ ($\breve{u} \in \breve{U}_u^{od}$ is the mode alternative in nest u) is as follows:

$$\sum_{\breve{u} \in \breve{U}_u^{od}} q_{u\breve{u}}^{od} = q_u^{od}, \forall u \in U^{od}, od \in OD \tag{4.18}$$

For each mode in the multi-modal transportation network, we continue to illustrate the relationship between the mode-specific OD demand and the model-specific path flow:

$$\sum_{k \in K_{u\breve{u}}^{od}} h_{u\breve{u},k}^{od} = q_{u\breve{u}}^{od}, \forall \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in OD \tag{4.19}$$

Regarding the line segment flow in the bus and metro network, let $\Gamma = (\gamma_{es})$ denote the line segment-route section incidence matrix, which is equal to 1 when line segment e of line l lies on route section s; otherwise, it is equal to 0. Then, the line segment flow expression is given as:

$$v_{b,e} = \sum_{s \in S_b} \gamma_{b,es} y_{b,l}^{e} x_{b,s}^{l} v_{b,s}, \forall e \in E_b \tag{4.20}$$

$$v_{m,e} = \sum_{s \in S_m} \gamma_{m,es} y_{m,l}^{e} x_{m,s}^{l} v_{m,s}, \forall e \in E_m \tag{4.21}$$

where $y_{b,l}^{e} = 1$ ($y_{m,l}^{e} = 1$) indicates that the transit line segment e is on transit line l.

Furthermore, we set the planning survey data as side constraints. These constraints include the zonal production and attraction flows, which are expressed as follows:

$$P^o = z^o, \forall o \in O \tag{4.22}$$

$$A^d = z^d, \forall d \in D \tag{4.23}$$

where the zonal model-specific production and attraction flows are defined as:

$$P^o = \sum_{d \in D} \sum_{u \in U^{od}} q_u^{od}, \forall o \in O \tag{4.24}$$

$$A^d = \sum_{o \in O} \sum_{u \in U^{od}} q_u^{od}, \forall d \in D \tag{4.25}$$

The results obtained from the planning survey data usually do not fit the field data, e.g., the target OD demand, traffic counts of road links (Ryu et al., 2014b) and onboard passenger counts of transit line segment. It does not require the estimated flows to be the same as the traffic observations exactly. Therefore, for the target OD mode-specific trip table, we refine the estimated mode-specific OD demand with lower and upper bounds about the observation:

$$\left(1 - \varepsilon_{u\breve{u}}^{od}\right) z_{u\breve{u}}^{od} \le q_{u\breve{u}}^{od} \le \left(1 + \varepsilon_{u\breve{u}}^{od}\right) z_{u\breve{u}}^{od}, \forall \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in \overline{OD} \tag{4.26}$$

We use the same method to refine the estimated results for the road links and transit line segments with observations:

$$\left(1 - \varepsilon_{c,s}\right) \cdot \bar{v}_{c,s} \le v_{c,s} + F_{c,s} \le \left(1 + \varepsilon_{c,s}\right) \cdot \bar{v}_{c,s}, \forall s \in \bar{S}_c \tag{4.27}$$

$$\left(1 - \varepsilon_{b,e}\right) \cdot \bar{v}_{b,e} \le v_{b,e} \le \left(1 + \varepsilon_{b,e}\right) \cdot \bar{v}_{b,e}, \forall e \in \bar{E}_b \tag{4.28}$$

$$\left(1 - \varepsilon_{m,e}\right) \cdot \bar{v}_{m,e} \le v_{m,e} \le \left(1 + \varepsilon_{m,e}\right) \cdot \bar{v}_{m,e}, \forall e \in \bar{E}_m \tag{4.29}$$

In addition, for the road links and transit line segments without observations, we set a constraint that their estimated flow cannot exceed their own capacity:

$$v_{c,s} + F_{c,s} \le C_{c,s}, \forall s \in \bar{\bar{S}}_c \tag{4.30}$$

$$v_{b,e} \le C_{b,e}, \forall e \in \bar{\bar{E}}_b \tag{4.31}$$

$$v_{m,e} \le C_{m,e}, \forall e \in \bar{\bar{E}}_m \tag{4.32}$$

where the line segment capacities of the bus and metro modes are defined as $C_{b,e} = f_l y_{b,l}^e \kappa_{b,l}$ and $C_{m,e} = f_l y_{m,l}^e \kappa_{m,l}$, respectively.

4.3.3 Formulated as a variational inequality problem

Based on the mode and route choice models above and the asymmetric travel time function of path, we adopt a VI formulation for the MM-PFE.

[MM-PFE]

Find $\left(\mathbf{q}_u^*, \mathbf{q}_{u\breve{u}}^*, \mathbf{h}_{u\breve{u},k}^*\right) \in \Omega$ such that

$$\begin{aligned}
&\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\sum_{k\in K_{u\breve{u}}^{od}}\left(c_{u\breve{u},k}^{od*} + \frac{1}{\theta_{u\breve{u}}^{od}}\ln h_{u\breve{u},k}^{od*}\right)\left(h_{u\breve{u},k}^{od} - h_{u\breve{u},k}^{od*}\right)\\
&+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\left(\rho_u^{od} - \frac{1}{\theta_{u\breve{u}}^{od}}\right)\ln q_{u\breve{u}}^{od*}\left(q_{u\breve{u}}^{od} - q_{u\breve{u}}^{od*}\right)\\
&-\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\mathrm{MSC}_{u\breve{u}}^{od}\left(q_{u\breve{u}}^{od} - q_{u\breve{u}}^{od*}\right)\\
&+\sum_{od\in OD}\sum_{u\in U^{od}}(1-\rho_u^{od})\ln q_u^{od*}\left(q_u^{od} - q_u^{od*}\right) \ge 0\\
&\forall\left(\mathbf{q}_u, \mathbf{q}_{u\breve{u}}, \mathbf{h}_{u\breve{u},k}\right) \in \Omega
\end{aligned} \tag{4.33}$$

where $\theta_{u\breve{u}}^{od}$ is the dispersion parameter for mode $\breve{u}$ in nest u of OD pair od; ρ_u^{od} is the parameter for nest u of each OD pair od; $\mathrm{MSC}_{u\breve{u}}^{od}$ is the mode-specific constant for mode $\breve{u}$ in nest u of OD pair od and it represents the exogenous modal attractiveness, which is an important component in the modal-split problem; and Ω is the set of feasible solutions that satisfy Eqs. (4.18)-(4.19), (4.22)-(4.23), and (4.26)-(4.32) and non-negative constraints

$$h_{u\breve{u},k}^{od} \ge 0, \forall k \in K_{u\breve{u}}^{od}, \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in OD \tag{4.34}$$

$$q_{u\breve{u}}^{od} \ge 0, \forall \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in OD \tag{4.35}$$

$$q_u^{od} \ge 0, \forall u \in U^{od}, od \in OD \tag{4.36}$$

[MM-PFE] can be simplified to the vector form:

$$F(\mathbf{z}^*)^T \cdot (\mathbf{z} - \mathbf{z}^*) \ge 0, \forall \mathbf{z} \in \Omega \tag{4.37}$$

where

$$\mathbf{z} = \left(\mathbf{q}_{\mathrm{u}}, \mathbf{q}_{u\breve{u}}, \mathbf{h}_{u\breve{u},k}\right)^T \tag{4.38}$$

$$F(\mathbf{z}) = \left((1-\boldsymbol{\rho}_u)\ln \mathbf{q}_{\mathrm{u}}, \left(\boldsymbol{\rho}_{u\breve{u}} - \frac{1}{\theta_{u\breve{u}}^{od}}\right)\ln \mathbf{q}_{u\breve{u}} - \mathrm{MSC}_{u\breve{u}}^{od}, \mathbf{c}_{u\breve{u},k}(\mathbf{h}) + \frac{1}{\theta_{u\breve{u}}^{od}}\ln \mathbf{h}_{u\breve{u},k} \right)^T \tag{4.39}$$

For the [MM-PFE] model, the mode-specific demand is distributed according to an NL model, and the path flow of each model follows the MNL model. To show this, we provide the following proposition:

Proposition 4.1. The optimal solution of the [MM-PFE] fulfills the NL-based mode choice and the MNL-based route choice.

Proof. The first-order conditions (Karush-Kuhn-Tucker conditions) for the [MM-PFE] model are as follows:

$$\left(c_{u\breve{u},k}^{od} + \frac{1}{\theta_{u\breve{u}}^{od}}\ln h_{u\breve{u},k}^{od} - \chi_{u\breve{u}}^{od} + J_{u\breve{u},k}^{od}\right) h_{u\breve{u},k}^{od} = 0, \forall k \in K_{u\breve{u}}^{od}, \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in OD \tag{4.40}$$

$$c_{u\breve{u},k}^{od} + \frac{1}{\theta_{u\breve{u}}^{od}}\ln h_{u\breve{u},k}^{od} - \chi_{u\breve{u}}^{od} + J_{u\breve{u},k}^{od} \geq 0, \forall k \in K_{u\breve{u}}^{od}, \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in OD \tag{4.41}$$

$$\left(\left(\rho_u^{od} - \frac{1}{\theta_{u\breve{u}}^{od}}\right)\ln q_{u\breve{u}}^{od} - \mathrm{MSC}_{u\breve{u}}^{od} + \chi_{u\breve{u}}^{od} - \varpi_u^{od} - \epsilon_{u\breve{u}}^{od-} + \epsilon_{u\breve{u}}^{od+}\right) q_{u\breve{u}}^{od} = 0, \forall \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in OD \tag{4.42}$$

$$\left(\rho_u^{od} - \frac{1}{\theta_{u\breve{u}}^{od}}\right)\ln q_{u\breve{u}}^{od} - \mathrm{MSC}_{u\breve{u}}^{od} + \chi_{u\breve{u}}^{od} - \varpi_u^{od} - \epsilon_{u\breve{u}}^{od-} + \epsilon_{u\breve{u}}^{od+} \geq 0, \forall \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in OD \tag{4.43}$$

$$\left((1-\rho_u^{od})\ln q_u^{od} + \varpi_u^{od} + \tau^o + \varrho^d\right) q_u^{od} = 0, \forall u \in U^{od}, od \in OD \tag{4.44}$$

$$(1-\rho_u^{od})\ln q_u^{od} + \varpi_u^{od} + \tau^o + \varrho^d \geq 0, \forall u \in U^{od}, od \in OD \tag{4.45}$$

where $J_{u\breve{u},k}^{od}$ $\left(\forall k \in K_{u\breve{u}}^{od}, \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in OD\right)$ has the following expression

$$J_{c,k}^{od} = \sum_{s \in \bar{\bar{S}}_c} \omega_{c,s}\eta_{c,sk}^{od} - \sum_{s \in \bar{S}_c} \mu_{c,s}^{-}\eta_{c,sk}^{od} + \sum_{s \in \bar{S}_c} \mu_{c,s}^{+}\eta_{c,sk}^{od} \tag{4.46}$$

$$J_{b,k}^{od} = \sum_{e \in \bar{\bar{E}}_b} \omega_{b,e} \sum_{s \in S_b} \bar{\gamma}_{b,es} a_{b,sk} - \sum_{e \in \bar{E}_b} \psi_{b,e}^{-} \sum_{s \in S_b} \bar{\gamma}_{b,es} a_{b,sk} + \sum_{e \in \bar{E}_b} \psi_{b,e}^{+} \sum_{s \in S_b} \bar{\gamma}_{b,es} a_{b,sk} \tag{4.47}$$

$$J_{m,k}^{od} = \sum_{e\in\bar{\bar{E}}_m} \omega_{m,e} \sum_{s\in S_m} \bar{\gamma}_{m,es} a_{m,sk} - \sum_{e\in\bar{E}_m} \psi_{m,e}^{-} \sum_{s\in S_b} \bar{\gamma}_{m,es} a_{m,sk} + \sum_{e\in\bar{E}_m} \psi_{m,e}^{+} \sum_{s\in S_m} \bar{\gamma}_{m,es} a_{m,sk} \tag{4.48}$$

The distribution of the mode-specific path flow for each OD pair is derived as follows.

As $h_{u\breve{u},k}^{od} > 0$, Eq. (4.49) is satisfied,

$$c_{u\breve{u},k}^{od} + \frac{1}{\theta_{u\breve{u}}^{od}} \ln h_{u\breve{u},k}^{od} - \chi_{u\breve{u}}^{od} + J_{u\breve{u},k}^{od} = 0 \tag{4.49}$$

Then, the analytical expression for the mode-specific path flow of each OD pair is:

$$h_{u\breve{u},k}^{od} = \exp\left(\theta_{u\breve{u}}^{od}\left(-c_{u\breve{u},k}^{od} + \chi_{u\breve{u}}^{od} - J_{u\breve{u},k}^{od}\right)\right) \tag{4.50}$$

Thus, the probability of choosing path k for mode $\breve{u}$ in nest u between OD pair od is given as:

$$p_{u\breve{u},k}^{od} = \frac{h_{u\breve{u},k}^{od}}{\sum_{p\in K_{u\breve{u}}^{od}} h_{u\breve{u},p}^{od}} = \frac{\exp\left(\theta_{u\breve{u}}^{od}\left(-c_{u\breve{u},k}^{od} - J_{u\breve{u},k}^{od}\right)\right)}{\sum_{p\in K_{u\breve{u}}^{od}} \exp\left(\theta_{u\breve{u}}^{od}\left(-c_{u\breve{u},p}^{od} - J_{u\breve{u},p}^{od}\right)\right)} \tag{4.51}$$

The distribution of the mode-specific demand for each OD pair is derived as follows.

Combining Eqs. (4.19) and (4.50), we obtain the expression of dual variable $\chi_{u\breve{u}}^{od}$:

$$\begin{aligned} &\sum_{k\in K_{u\breve{u}}^{od}} h_{u\breve{u},k}^{od} = \sum_{k\in K_{u\breve{u}}^{od}} \exp\left(\theta_{u\breve{u}}^{od}\left(-c_{u\breve{u},k}^{od} + \chi_{u\breve{u}}^{od} - J_{u\breve{u},k}^{od}\right)\right) \\ &\Longrightarrow \chi_{u\breve{u}}^{od} = \frac{1}{\theta_{u\breve{u}}^{od}} \ln q_{u\breve{u}}^{od} - \frac{1}{\theta_{u\breve{u}}^{od}} \ln \sum_{k\in K_{u\breve{u}}^{od}} \exp\left(-\theta_{u\breve{u}}^{od}\left(c_{u\breve{u},k}^{od} + J_{u\breve{u},k}^{od}\right)\right) \end{aligned} \tag{4.52}$$

Because $q_{u\breve{u}}^{od} > 0$, from Eqs. (4.42)-(4.43), we have

$$\left(\rho_u^{od} - \frac{1}{\theta_{u\breve{u}}^{od}}\right) \ln q_{u\breve{u}}^{od} - \mathrm{MSC}_{u\breve{u}}^{od} + \chi_{u\breve{u}}^{od} - \varpi_u^{od} - \epsilon_{u\breve{u}}^{od-} + \epsilon_{u\breve{u}}^{od+} = 0 \tag{4.53}$$

By substituting Eq. (4.52) into Eq. (4.53), we obtain:

$$\begin{aligned} &\left(\rho_u^{od} - \frac{1}{\theta_{u\breve{u}}^{od}}\right) \ln q_{u\breve{u}}^{od} - \mathrm{MSC}_{u\breve{u}}^{od} + \frac{1}{\theta_{u\breve{u}}^{od}} \ln q_{u\breve{u}}^{od} \\ &\quad - \frac{1}{\theta_{u\breve{u}}^{od}} \ln \sum_{k\in K_{u\breve{u}}^{od}} \exp\left(-\theta_{u\breve{u}}^{od}\left(c_{u\breve{u},k}^{od} + J_{u\breve{u},k}^{od}\right)\right) - \varpi_u^{od} - \epsilon_{u\breve{u}}^{od-} + \epsilon_{u\breve{u}}^{od+} \\ &\quad = 0 \end{aligned} \tag{4.54}$$

The analytical expression for mode-specific OD flow is presented as:

$$q_{u\breve{u}}^{od} = \exp\left(\frac{1}{\rho_u^{od}}\left(\frac{1}{\theta_{u\breve{u}}^{od}}\ln\sum_{k\in K_{u\breve{u}}^{od}}\exp\left(-\theta_{u\breve{u}}^{od}\left(c_{u\breve{u},k}^{od}+J_{u\breve{u},k}^{od}\right)\right)+\mathrm{MSC}_{u\breve{u}}^{od}+\varpi_u^{od}\right.\right.$$
$$\left.\left.+\epsilon_{u\breve{u}}^{od-}-\epsilon_{u\breve{u}}^{od+}\right)\right) \tag{4.55}$$

We obtain the expression of q_u^{od} from Eqs. (4.44) and (4.45) as follows:

$$\begin{aligned}&(1-\rho_u^{od})\ln q_u^{od}+\varpi_u^{od}+\tau^o+\varrho^d=0\\&\Rightarrow q_u^{od}=\exp\left(-\frac{\varpi_u^{od}+\tau^o+\varrho^d}{1-\rho_u^{od}}\right)\end{aligned} \tag{4.56}$$

Here, we set

$$\Lambda_{u\breve{u}}^{od}=\frac{1}{\theta_{u\breve{u}}^{od}}\ln\sum_{k\in K_{u\breve{u}}^{od}}\exp\left(-\theta\left(c_{u\breve{u},k}^{od}+J_{u\breve{u},k}^{od}\right)\right)+\mathrm{MSC}_{u\breve{u}}^{od} \tag{4.57}$$

As $\sum_{\breve{u}\in\breve{U}_u^{od}} q_{u\breve{u}}^{od}=q_u^{od}$, ϖ_u^{od} is expressed as:

$$\begin{aligned}\varpi_u^{od}=&-(1-\rho_u^{od})\rho_u^{od}\ln\sum_{\breve{u}\in\breve{U}_u^{od}}\exp\left(\frac{1}{\rho_u^{od}}\left(\Lambda_{u\breve{u}}^{od}+\epsilon_{u\breve{u}}^{od-}-\epsilon_{u\breve{u}}^{od+}\right)\right)\\&-\rho_u^{od}(\tau^o+\varrho^d)\end{aligned} \tag{4.58}$$

Let

$$I_u^{od}=\ln\sum_{\breve{u}\in\breve{U}_u^{od}}\exp\left(\frac{1}{\rho_u^{od}}\left(\Lambda_{u\breve{u}}^{od}+\epsilon_{u\breve{u}}^{od-}-\epsilon_{u\breve{u}}^{od+}\right)\right) \tag{4.59}$$

Equation (4.58) is simplified as

$$\varpi_u^{od}=-(1-\rho_u^{od})\rho_u^{od}I_u^{od}-\rho_u^{od}(\tau^o+\varrho^d) \tag{4.60}$$

The analytical expression of the nest-specific OD flow is presented as:

$$q_u^{od}=\exp\left(-\frac{\varpi_u^{od}+\tau^o+\varrho^d}{1-\rho_u^{od}}\right)=\exp(\rho_u^{od}I_u^{od})\cdot\exp(-\tau^o)\cdot\exp(-\varrho^d) \tag{4.61}$$

Thus, the marginal probability p_u^{od} and conditional probability $p_{\breve{u}|u}^{od}$ for mode choice are as follows:

$$p_u^{od}=\frac{q_u^{od}}{\sum_{v\in U^{od}}q_v^{od}}=\frac{\exp(\rho_u^{od}I_u^{od})}{\sum_{v\in U^{od}}\exp\left(\rho_v^{od}I_v^{od}\right)} \tag{4.62}$$

$$p_{\breve{u}|u}^{od}=\frac{q_{u\breve{u}}^{od}}{\sum_{l\in\breve{U}_u^{od}}q_{ul}^{od}}=\frac{\exp\left(\Lambda_{u\breve{u}}^{od}/\rho_u^{od}\right)}{\sum_{l\in\breve{U}_u^{od}}\exp\left(\Lambda_{ul}^{od}/\rho_u^{od}\right)} \tag{4.63}$$

Furthermore, we can rewrite $q_{u\breve{u}}^{od}$ as:

$$q_{u\breve{u}}^{od} = \exp\left(\left(\Lambda_{u\breve{u}}^{od} + \epsilon_{u\breve{u}}^{od-} - \epsilon_{u\breve{u}}^{od+}\right)/\rho_u^{od}\right) \cdot \exp(\tau^o + \varrho^d) \cdot \left(\sum_{\breve{u}\in\breve{U}_u^{od}} \exp\left(\frac{\Lambda_{u\breve{u}}^{od} + \epsilon_{u\breve{u}}^{od-} - \epsilon_{u\breve{u}}^{od+}}{\rho_u^{od}}\right)\right)^{\rho_u^{od}-1} \tag{4.64}$$

Thus, the NL probability expression for mode choice can be obtained as follows:

$$\begin{aligned} p_{u\breve{u}}^{od} &= \frac{q_{u\breve{u}}^{od}}{\sum_{v\in U^{od}} \sum_{l\in\breve{U}_v^{od}} q_{vl}^{od}} \\ &= \frac{\exp\left(\frac{\Lambda_{u\breve{u}}^{od} + \epsilon_{u\breve{u}}^{od-} - \epsilon_{u\breve{u}}^{od+}}{\rho_u^{od}}\right) \cdot \left(\sum_{l\in\breve{U}_u^{od}} \exp\left(\frac{\Lambda_{ul}^{od} + \epsilon_{ul}^{od-} - \epsilon_{ul}^{od+}}{\rho_u^{od}}\right)\right)^{\rho_u^{od}-1}}{\sum_{v\in U^{od}} \left(\sum_{l\in\breve{U}_v^{od}} \exp\left(\frac{\Lambda_{vl}^{od} + \epsilon_{vl}^{od-} - \epsilon_{vl}^{od+}}{\rho_v^{od}}\right)\right)^{\rho_v^{od}}} \\ &= p_{\breve{u}|u}^{od} \cdot p_u^{od} \end{aligned} \tag{4.65}$$

This completes the proof. □

Remarks

(1) There are three kinds of dual variables in this study: free variables ($\chi_{u\breve{u}}^{od}, \varpi_u^{od}, \tau^o$ and ϱ^d) associated with equality constraints representing the positive or negative intrinsic attractiveness; positive variables ($\epsilon_{u\breve{u}}^{od-}, \epsilon_{u\breve{u}}^{od+}, \mu_{c,s}^-, \mu_{c,s}^+, \psi_{b,e}^-, \psi_{b,e}^+, \xi_{m,e}^-, \xi_{m,e}^+$) associated with the lower and upper bound constraints keeping the estimated flow being higher than the lower bound and lower than the upper bound; and positive variables ($\omega_{c,s}, \omega_{b,e}, \omega_{m,e}$) associated with the capacity constraint representing queuing when the total link flow reaches its link capacity.

(2) With the expression of q_u^{od}, we obtain the total demand for each OD pair as

$$q^{od} = \sum_{u\in U^{od}} q_u^{od} = \sum_{u\in U^{od}} \exp(\rho_u^{od} I_u^{od}) \cdot \exp(-\tau^o) \cdot \exp(-\varrho^d) \tag{4.66}$$

Let $B^o P^o = \exp(-\tau^o)$, $B^d A^d = \exp(-\varrho^d)$, and $\pi^{od} = -\ln\sum_{u\in U^{od}} \exp(\rho_u^{od} I_u^{od})$, where B^o and B^d are the balancing factors; P^o and A^d are the production and attraction flows, respectively; and π^{od} is the multi-modal OD cost between origin o and destination d. Thus, the total demand for each OD pair can be rewritten as follows:

$$q^{od} = B^o P^o B^d A^d \exp(-\pi^{od}) \tag{4.67}$$

This is similar to the convex programming formulation provided by Ryu et al. (2014) for the small community PFE that finds the most probable demand pattern based on Smith's efficiency principle. The main difference between the formulations is that the multi-modal PFE

framework presented in this paper considers the interactions of multiple modes sharing the roadway; MM-PFE cannot be formulated as a convex program and requires a solution algorithm that can handle asymmetric and non-separable link travel time functions.

4.4 Solution Algorithm

Owing to the asymmetric travel time functions of road link, route section and path, a diagonalization method is adopted to solve the proposed model in this section (Florian, 1977; Florian and Spiess, 1982; de Cea and Fernández, 1993; de Cea et al., 2005). In each outer iteration, the travel time function of the route section is diagonalized, and the VI formulation is reformulated as a convex optimization model. Then, the reformulated convex optimization problem can be solved using a path-based partial linearization solution algorithm (Chen et al., 2009) embedded with an iterative balancing scheme for direction finding. An iterative balancing scheme is used to handle various inequality side constraints in the optimization programming model.

4.4.1 Diagonalization of cost function

In this subchapter, we demonstrate that the diagonalized [MM-PFE] problem can be reformulated as a convex programming problem. In literature, PFE in mathematical programming formulations have been well solved by path-based solution algorithms (Chen et al., 2009, 2012; Ryu et al., 2014b). To do this, we have the following proposition.

Proposition 4.2. When the cost function of the [MM-PFE] model is diagonalized, [MM-PFE] can be reformulated as a convex optimization problem such that

[MM-PFE-D]

$$\begin{aligned}\min Z = &\sum_{s\in S}\int_0^{v_s}\hat{c}_s(\omega)\,d\omega + \sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\sum_{k\in K_{u\breve{u}}^{od}}\frac{1}{\theta_{u\breve{u}}^{od}}h_{u\breve{u},k}^{od}\left(\ln h_{u\breve{u},k}^{od}-1\right)\\ &+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\left(\rho_u^{od}-\frac{1}{\theta_{u\breve{u}}^{od}}\right)q_{u\breve{u}}^{od}\left(\ln q_{u\breve{u}}^{od}-1\right)\\ &-\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\mathrm{MSC}_{u\breve{u}}^{od}q_{u\breve{u}}^{od}\\ &+\sum_{od\in OD}\sum_{u\in U^{od}}(1-\rho_u^{od})q_u^{od}(\ln q_u^{od}-1)\end{aligned} \tag{4.68}$$

s.t. Eqs. (4.18)-(4.19), (4.22)-(4.23), (4.26)-(4.32), and (4.34)-(4.36).

Proof. Problem [MM-PFE] can be further expressed by separating $\left(c_{u\breve{u},k}^{od*} + \frac{1}{\theta_{u\breve{u}}^{od}} \ln h_{u\breve{u},k}^{od*}\right)$ into two terms:

$$\begin{aligned}
&\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in \breve{U}_u^{od}}\sum_{k\in K_{u\breve{u}}^{od}} c_{u\breve{u},k}^{od*}\left(h_{u\breve{u},k}^{od} - h_{u\breve{u},k}^{od*}\right)\\
&+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in \breve{U}_u^{od}}\sum_{k\in K_{u\breve{u}}^{od}} \frac{1}{\theta_{u\breve{u}}^{od}} \ln h_{u\breve{u},k}^{od*}\left(h_{u\breve{u},k}^{od} - h_{u\breve{u},k}^{od*}\right)\\
&+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in \breve{U}_u^{od}}\left(\rho_u^{od} - \frac{1}{\theta_{u\breve{u}}^{od}}\right)\ln q_{u\breve{u}}^{od*}\left(q_{u\breve{u}}^{od} - q_{u\breve{u}}^{od*}\right)\\
&-\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in \breve{U}_u^{od}} \mathrm{MSC}_{u\breve{u}}^{od}\left(q_{u\breve{u}}^{od} - q_{u\breve{u}}^{od*}\right)\\
&+\sum_{od\in OD}\sum_{u\in U^{od}}(1-\rho_u^{od})\ln q_u^{od*}\left(q_u^{od} - q_u^{od*}\right) \geq 0
\end{aligned} \tag{4.69}$$

Then, we substitute the arc cost function for the path cost function as follows:

$$\begin{aligned}
&\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in \breve{U}_u^{od}}\sum_{k\in K_{u\breve{u}}^{od}}\left(\sum_{s\in S} a_{sk}c_s\right)\left(h_{u\breve{u},k}^{od} - h_{u\breve{u},k}^{od*}\right)\\
&+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in \breve{U}_u^{od}}\sum_{k\in K_{u\breve{u}}^{od}} \frac{1}{\theta_{u\breve{u}}^{od}} \ln h_{u\breve{u},k}^{od*}\left(h_{u\breve{u},k}^{od} - h_{u\breve{u},k}^{od*}\right)\\
&+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in \breve{U}_u^{od}}\left(\rho_u^{od} - \frac{1}{\theta_{u\breve{u}}^{od}}\right)\ln q_{u\breve{u}}^{od*}\left(q_{u\breve{u}}^{od} - q_{u\breve{u}}^{od*}\right)\\
&-\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in \breve{U}_u^{od}} \mathrm{MSC}_{u\breve{u}}^{od}\left(q_{u\breve{u}}^{od} - q_{u\breve{u}}^{od*}\right)\\
&+\sum_{od\in OD}\sum_{u\in U^{od}}(1-\rho_u^{od})\ln q_u^{od*}\left(q_u^{od} - q_u^{od*}\right) \geq 0
\end{aligned} \tag{4.70}$$

Furthermore, we can calculate arc flows based on the mode-specific path flows of all OD pairs and simplify [MM-PFE] using the arc and path space.

$$\begin{aligned}
&\sum_{s\in S}\left(\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in \breve{U}_u^{od}}\sum_{k\in K_{u\breve{u}}^{od}} a_{sk}\left(h_{u\breve{u},k}^{od} - h_{u\breve{u},k}^{od*}\right)\right)c_s\\
&+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in \breve{U}_u^{od}}\sum_{k\in K_{u\breve{u}}^{od}} \frac{1}{\theta_{u\breve{u}}^{od}} \ln h_{u\breve{u},k}^{od*}\left(h_{u\breve{u},k}^{od} - h_{u\breve{u},k}^{od*}\right)
\end{aligned} \tag{4.71}$$

$$+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\left(\rho_u^{od}-\frac{1}{\theta_{u\breve{u}}^{od}}\right)\ln q_{u\breve{u}}^{od*}\left(q_{u\breve{u}}^{od}-q_{u\breve{u}}^{od*}\right)$$

$$-\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\mathrm{MSC}_{u\breve{u}}^{od}\left(q_{u\breve{u}}^{od}-q_{u\breve{u}}^{od*}\right)$$

$$+\sum_{od\in OD}\sum_{u\in U^{od}}(1-\rho_u^{od})\ln q_u^{od*}\left(q_u^{od}-q_u^{od*}\right)\geq 0$$

At each iteration of the diagonalization loop, the travel time function c_s results in a diagonalized cost function $\hat{c}_s$ (i.e., separable function) that can be used to formulate a convex program for the combined modal split and assignment problem with an NL mode choice model and an MNL route choice model.

$$\begin{aligned}&\sum_{s\in S}(v_s-v_s^*)\hat{c}_s+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\sum_{k\in K_{u\breve{u}}^{od}}\frac{1}{\theta_{u\breve{u}}^{od}}\ln h_{u\breve{u},k}^{od*}\left(h_{u\breve{u},k}^{od}-h_{u\breve{u},k}^{od*}\right)\\&+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\left(\rho_u^{od}-\frac{1}{\theta_{u\breve{u}}^{od}}\right)\ln q_{u\breve{u}}^{od*}\left(q_{u\breve{u}}^{od}-q_{u\breve{u}}^{od*}\right)\\&-\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\mathrm{MSC}_{u\breve{u}}^{od}\left(q_{u\breve{u}}^{od}-q_{u\breve{u}}^{od*}\right)\\&+\sum_{od\in OD}\sum_{u\in U^{od}}(1-\rho_u^{od})\ln q_u^{od*}\left(q_u^{od}-q_u^{od*}\right)\geq 0,\forall\left(\mathbf{q}_u,\mathbf{q}_{u\breve{u}},\mathbf{h}_{u\breve{u},k}\right)\in\Omega\end{aligned}\tag{4.72}$$

Thus, it has an equivalent mathematical programming formulation as follows

[MM-PFE-D]

$$\begin{aligned}\min Z=&\sum_{s\in S}\int_0^{v_s}\hat{c}_s(\omega)\,d\omega+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\sum_{k\in K_{u\breve{u}}^{od}}\frac{1}{\theta_{u\breve{u}}^{od}}h_{u\breve{u},k}^{od}\left(\ln h_{u\breve{u},k}^{od}-1\right)\\&+\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\left(\rho_u^{od}-\frac{1}{\theta_{u\breve{u}}^{od}}\right)q_{u\breve{u}}^{od}\left(\ln q_{u\breve{u}}^{od}-1\right)\\&-\sum_{od\in OD}\sum_{u\in U^{od}}\sum_{\breve{u}\in\breve{U}_u^{od}}\mathrm{MSC}_{u\breve{u}}^{od}q_{u\breve{u}}^{od}\\&+\sum_{od\in OD}\sum_{u\in U^{od}}(1-\rho_u^{od})q_u^{od}(\ln q_u^{od}-1)\end{aligned}\tag{4.73}$$

s.t. Eqs. (18)-(19), (22)-(23), (26)-(32), and (34)-(36).

This completes the proof. □

4.4.2 Overall solution algorithm framework

The whole framework of the solution algorithm for [MM-PFE] is summarized into the following steps, as shown in Figure 4.3:

Step 0. (initialization). Find an initial feasible solution $(\bar{\mathbf{q}}, \bar{\mathbf{h}}, \bar{\mathbf{v}})$.

Step 1. Diagonalize $c(\mathbf{v})$ at $(\bar{\mathbf{q}}, \bar{\mathbf{h}}, \bar{\mathbf{v}})$.

Step 2. Solve problem [MM-PFE-*D*] to get $(\hat{\mathbf{q}}, \hat{\mathbf{h}}, \hat{\mathbf{v}})$.

Step 3. Stop test. If $(\bar{\mathbf{q}}, \bar{\mathbf{h}}, \bar{\mathbf{v}})$ and $(\hat{\mathbf{q}}, \hat{\mathbf{h}}, \hat{\mathbf{v}})$ are sufficiently close, stop; otherwise:

Step 4. Make: $(\bar{\mathbf{q}}, \bar{\mathbf{h}}, \bar{\mathbf{v}}) \leftarrow (\hat{\mathbf{q}}, \hat{\mathbf{h}}, \hat{\mathbf{v}})$, and return to Step 1.

Algorithm Diagonalization algorithm for the MM-PFE
1: Initialization. Find an initial feasible solution $(\bar{\mathbf{q}}, \bar{\mathbf{h}}, \bar{\mathbf{v}})$
2: **while** $err > \varepsilon$
3: Diagonalize $c(\mathbf{v})$ at $(\bar{\mathbf{q}}, \bar{\mathbf{h}}, \bar{\mathbf{v}})$
4: Solve problem [MM-PFE-D] to get $(\hat{\mathbf{q}}, \hat{\mathbf{h}}, \hat{\mathbf{v}})$
5: $err = \max\{|\hat{\mathbf{q}} - \bar{\mathbf{q}}|, |\hat{\mathbf{h}} - \bar{\mathbf{h}}|, |\hat{\mathbf{v}} - \bar{\mathbf{v}}|\}$
6: $(\bar{\mathbf{q}}, \bar{\mathbf{h}}, \bar{\mathbf{v}}) \leftarrow (\hat{\mathbf{q}}, \hat{\mathbf{h}}, \hat{\mathbf{v}})$
7: **end while**
8: **return** $(\mathbf{q}^*, \mathbf{h}^*, \mathbf{v}^*) \approx (\bar{\mathbf{q}}, \bar{\mathbf{h}}, \bar{\mathbf{v}})$
where $\varepsilon \in R_+$ is the tolerance

Figure 4.3 Solution algorithm for MM-PFE

To solve the [MM-PFE] problem in Step 2, we adopt a path-based partial linearization algorithm (Chen et al., 2009) combined with an iterative balancing search direction scheme (Bell and Iida, 1997) and a self-regulated averaging step-size scheme (Liu et al., 2009). The partial linearization subproblem is a convex program with linear equity and inequality constraints, and it can be efficiently solved using the three-layer iterative balancing scheme, as presented in subchapter 4.4.3.

4.4.3 Direction finding: iterative balancing scheme

As the [MM-PFE-D] model has a three-layer structure (two layers for the mode choice and one layer for the route choice), the traditional iterative balancing scheme (Bell, 1995) cannot be used directly used here. This section mainly illustrates how to develop the Lagrange dual formulation to obtain the adjustment factors and update the dual variables. This approach is based on the work by Li (2016), and differs from Bell's approach (Bell, 1995; Bell et al., 1997; Chen et al., 2005, 2009, 2010; Ryu et al., 2014b).

4.4.3.1 Dual formulation of [MM-PFE-D]

$Y = \left(\dots h_{u\breve{u},k}^{od} \dots q_{u\breve{u}}^{od} \dots q_{u}^{od} \dots \right)^T$ is a vector of primal decision variables in the [MM-PFE-D] model, and $W = \begin{pmatrix} \dots \chi_{u\breve{u}}^{od} \dots \varpi_u^{od} \dots \tau^o \dots \varrho^d \dots \epsilon_{u\breve{u}}^{od-} \dots \epsilon_{u\breve{u}}^{od+} \dots \omega_{c,s} \dots \omega_{b,e} \dots \\ \omega_{m,e} \dots \mu_{c,s}^- \dots \mu_{c,s}^+ \dots \psi_{b,e}^- \dots \psi_{b,e}^+ \dots \xi_{m,e}^- \dots \xi_{m,e}^+ \dots \end{pmatrix}^T$ is a vector of the dual variables (Lagrange multipliers). For given W, we define $H(W,Y)$ as a function of Y (i.e., $h_{u\breve{u},k}^{od}$s, $q_{u\breve{u}}^{od}$s and q_u^{od}s), which is essentially the Lagrange formulation of [MM-PFE-D]:

$$H(W,Y) = Z(Y) + \sum_{od\in OD} \sum_{u\in U^{od}} \sum_{\breve{u}\in \breve{U}_u^{od}} \chi_{u\breve{u}}^{od} \left(q_{u\breve{u}}^{od} - \sum_{k\in K_{u\breve{u}}^{od}} h_{u\breve{u},k}^{od} \right)$$

$$+ \sum_{od\in OD} \sum_{u\in U^{od}} \varpi_u^{od} \left(q_u^{od} - \sum_{\breve{u}\in \breve{U}_u^{od}} q_{u\breve{u}}^{od} \right) + \sum_{o\in O} \tau^o (P^o - z^o) + \sum_{d\in D} \varrho^d (A^d - z^d)$$

$$+ \sum_{od\in OD} \sum_{u\in U^{od}} \sum_{\breve{u}\in \breve{U}_u^{od}} \epsilon_{u\breve{u}}^{od-} \left(\left(1 - \varepsilon_{\breve{u}u}^{od}\right) \cdot z_{\breve{u}u}^{od} - q_{u\breve{u}}^{od} \right)$$

$$+ \sum_{od\in OD} \sum_{u\in U^{od}} \sum_{\breve{u}\in \breve{U}_u^{od}} \epsilon_{u\breve{u}}^{od+} \left(q_{u\breve{u}}^{od} - \left(1 + \varepsilon_{\breve{u}u}^{od}\right) \cdot z_{\breve{u}u}^{od} \right)$$

$$+ \sum_{s\in \bar{\bar{S}}_c} \omega_{c,s} \left(v_{c,s} + F_{c,s} - C_{c,s} \right) + \sum_{e\in \bar{\bar{E}}_b} \omega_{b,e} \left(v_{b,e} - C_{b,e} \right) + \sum_{e\in \bar{\bar{E}}_m} \omega_{m,e} \left(v_{m,e} - C_{m,e} \right)$$

$$+ \sum_{s\in \bar{S}_c} \mu_{c,s}^- \left(\left(1 - \varepsilon_{c,s}\right) \cdot \bar{v}_{c,s} - v_{c,s} - F_{c,s} \right) + \sum_{s\in \bar{S}_c} \mu_{c,s}^+ \left(v_{c,s} + F_{c,s} - \left(1 + \varepsilon_{c,s}\right) \cdot \bar{v}_{c,s} \right)$$

$$+ \sum_{e\in \bar{E}_b} \psi_{b,e}^- \left(\left(1 - \varepsilon_{b,e}\right) \cdot \bar{v}_{b,e} - v_{b,e} \right) + \sum_{e\in \bar{E}_b} \psi_{b,e}^+ \left(v_{b,e} - \left(1 + \varepsilon_{b,e}\right) \cdot \bar{v}_{b,e} \right)$$

$$+ \sum_{e\in \bar{E}_m} \xi_{m,e}^- \left(\left(1 - \varepsilon_{m,e}\right) \cdot \bar{v}_{m,e} - v_{m,e} \right) + \sum_{e\in \bar{E}_m} \xi_{m,e}^+ \left(v_{m,e} - \left(1 + \varepsilon_{m,e}\right) \cdot \bar{v}_{m,e} \right)$$

The dual problem of [MM-PFE-D] is defined as

(DP) max $\Theta(W)$

s.t. $W \in D_W$

where

$D_W = \left\{ W: \epsilon_{u\breve{u}}^{od-}, \epsilon_{u\breve{u}}^{od+}, \omega_{c,s}, \omega_{b,e}, \omega_{m,e}, \mu_{c,s}^-, \mu_{c,s}^+, \psi_{b,e}^-, \psi_{b,e}^+, \xi_{m,e}^-, \xi_{m,e}^+ \geq 0, \text{free } \chi_{u\breve{u}}^{od}, \varpi_u^{od}, \tau^o \text{ and } \varrho^d \right\}$ is the feasible region, and $\Theta(W) = \inf\left\{ H(W,Y), Y \in \left(R_+^h \cup R_+^q \right) \right\}$.

4.4.3.2 Dual variables adjustment factor

Proposition 4.3. The concave function $\Theta(W)$ is differentiable with respect to W. For the dual problem (DP), to gain the optimal solution, $\nabla\Theta(W)$ should satisfy the following conditions:

$$\frac{\partial\Theta(W)}{\partial\chi_{u\breve{u}}^{od}} = q_{u\breve{u}}^{od} - \sum_{k\in K_{u\breve{u}}^{od}} h_{u\breve{u},k}^{od} = 0 \tag{4.74}$$

$$\frac{\partial\Theta(W)}{\partial\varpi_{u}^{od}} = q_{u}^{od} - \sum_{\breve{u}\in\breve{U}_{u}^{od}} q_{u\breve{u}}^{od} = 0 \tag{4.75}$$

$$\frac{\partial\Theta(W)}{\partial\tau^{o}} = P^{o} - z^{o} = 0 \tag{4.76}$$

$$\frac{\partial\Theta(W)}{\partial\varrho^{d}} = A^{d} - z^{d} = 0 \tag{4.77}$$

$$\frac{\partial\Theta(W)}{\partial\epsilon_{u\breve{u}}^{od-}} = \left(1-\varepsilon_{\breve{u}u}^{od}\right)\cdot z_{\breve{u}u}^{od} - q_{u\breve{u}}^{od} = \begin{cases}\leq 0, if\ \epsilon_{u\breve{u}}^{od-} = 0\\ = 0, if\ \epsilon_{u\breve{u}}^{od-} > 0\end{cases} \tag{4.78}$$

$$\frac{\partial\Theta(W)}{\partial\epsilon_{u\breve{u}}^{od+}} = q_{u\breve{u}}^{od} - \left(1+\varepsilon_{\breve{u}u}^{od}\right)\cdot z_{\breve{u}u}^{od} = \begin{cases}\leq 0, if\ \epsilon_{u\breve{u}}^{od+} = 0\\ = 0, if\ \epsilon_{u\breve{u}}^{od+} > 0\end{cases} \tag{4.79}$$

$$\frac{\partial\Theta(W)}{\partial\omega_{c,s}} = v_{c,s} + F_{c,s} - C_{c,s} = \begin{cases}\leq 0, if\ \omega_{c,s} = 0\\ = 0, if\ \omega_{c,s} > 0\end{cases} \tag{4.80}$$

$$\frac{\partial\Theta(W)}{\partial\omega_{b,e}} = v_{b,e} - C_{b,e} = \begin{cases}\leq 0, if\ \omega_{b,e} = 0\\ = 0, if\ \omega_{b,e} > 0\end{cases} \tag{4.81}$$

$$\frac{\partial\Theta(W)}{\partial\omega_{m,e}} = v_{m,e} - C_{m,e} = \begin{cases}\leq 0, if\ \omega_{m,e} = 0\\ = 0, if\ \omega_{m,e} > 0\end{cases} \tag{4.82}$$

$$\frac{\partial\Theta(W)}{\partial\mu_{c,s}^{-}} = \left(1-\varepsilon_{c,s}\right)\cdot\bar{v}_{c,s} - v_{c,s} - F_{c,s} = \begin{cases}\leq 0, if\ \mu_{c,s}^{-} = 0\\ = 0, if\ \mu_{c,s}^{-} > 0\end{cases} \tag{4.83}$$

$$\frac{\partial\Theta(W)}{\partial\mu_{c,s}^{+}} = v_{c,s} + F_{c,s} - \left(1+\varepsilon_{c,s}\right)\cdot\bar{v}_{c,s} = \begin{cases}\leq 0, if\ \mu_{c,s}^{+} = 0\\ = 0, if\ \mu_{c,s}^{+} > 0\end{cases} \tag{4.84}$$

$$\frac{\partial\Theta(W)}{\partial\psi_{b,e}^{-}} = \left(1-\varepsilon_{b,e}\right)\cdot\bar{v}_{b,e} - v_{b,e} = \begin{cases}\leq 0, if\ \psi_{b,e}^{-} = 0\\ = 0, if\ \psi_{b,e}^{-} > 0\end{cases} \tag{4.85}$$

$$\frac{\partial\Theta(W)}{\partial\psi_{b,e}^{+}} = v_{b,e} - \left(1+\varepsilon_{b,e}\right)\cdot\bar{v}_{b,e} = \begin{cases}\leq 0, if\ \psi_{b,e}^{+} = 0\\ = 0, if\ \psi_{b,e}^{+} > 0\end{cases} \tag{4.86}$$

$$\frac{\partial\Theta(W)}{\partial\xi_{m,e}^{-}} = \left(1-\varepsilon_{m,e}\right)\cdot\bar{v}_{m,e} - v_{m,e} = \begin{cases}\leq 0, if\ \xi_{m,e}^{-} = 0\\ = 0, if\ \xi_{m,e}^{-} > 0\end{cases} \tag{4.87}$$

$$\frac{\partial\Theta(W)}{\partial\xi_{m,e}^{+}} = v_{m,e} - \left(1+\varepsilon_{m,e}\right)\cdot\bar{v}_{m,e} = \begin{cases}\leq 0, if\ \xi_{m,e}^{+} = 0\\ = 0, if\ \xi_{m,e}^{+} > 0\end{cases} \tag{4.88}$$

Based on Proposition 4.3, we can classify the dual variables into two types: (1) Eqs. (74)-(77), and (2) Eqs. (78)-(88). Hence, we present the following propositions to derive the adjustment factors for the dual variables $\chi_{u\breve{u}}^{od}$ in Eq. (74) and $\epsilon_{u\breve{u}}^{od-}$ in Eq. (78). The expressions of the adjustment factors for the remaining dual variables are given in Table 4.1.

Proposition 4.4. If the optimality condition given by Eq. (74) is not 0, assume that $\chi_{u\breve{u}}^{od}$ is the *l*th component of W; then, an adjustment $\tilde{\chi}_{u\breve{u}}^{od}$ can be made to $\chi_{u\breve{u}}^{od}$ such that

$\tilde{\chi}_{u\breve{u}}^{od} = \frac{1}{\theta_{u\breve{u}}^{od}} \ln\left(\frac{q_{u\breve{u}}^{od}}{\sum_{k\in K_{u\breve{u}}^{od}} h_{u\breve{u},k}^{od}}\right),\ \tilde{\chi}_{u\breve{u}}^{od} \times \frac{\partial\Theta(W)}{\partial\chi_{u\breve{u}}^{od}} > 0\ ,\ \tilde{\chi}_{u\breve{u}}^{od} \times \frac{\partial\Theta\left(W-\tilde{\chi}_{u\breve{u}}^{od}\cdot e_l\right)}{\partial\chi_{u\breve{u}}^{od}} = 0\ ,\ \Theta\left(W-\tilde{\chi}_{u\breve{u}}^{od}\cdot e_l\right) \geq \Theta\left(W-\tilde{\chi}_{u\breve{u}}^{od\prime}\cdot e_l\right), \tilde{\chi}_{u\breve{u}}^{od\prime} \in R,$

where e_l is the unit vector with the *l*th component equal to 1.

Proof. After adding an adjustment factor $\tilde{\chi}_{u\breve{u}}^{od}$ to $\chi_{u\breve{u}}^{od}$, we obtain:

$$\frac{\partial\Theta\left(W-\tilde{\chi}_{u\breve{u}}^{od}\cdot e_l\right)}{\partial\chi_{u\breve{u}}^{od}} = 0 \tag{4.89}$$

From Eqs. (4.74) and (4.89), we have:

$$q_{u\breve{u}}^{od} = \sum_{k\in K_{u\breve{u}}^{od}} \exp\left(-\theta_{u\breve{u}}^{od}\left(c_{u\breve{u},k}^{od} - \chi_{u\breve{u}}^{od} - \tilde{\chi}_{u\breve{u}}^{od} + J_{u\breve{u},k}^{od}\right)\right) = \exp\left(\theta_{u\breve{u}}^{od}\tilde{\chi}_{u\breve{u}}^{od}\right) \sum_{k\in K_{u\breve{u}}^{od}} h_{u\breve{u},k}^{od} \tag{4.90}$$

Let $q = \sum_{k\in K_{u\breve{u}}^{od}} h_{u\breve{u},k}^{od}$ in Eq. (4.90), then we have:

$$\tilde{\chi}_{u\breve{u}}^{od} \begin{cases} > 0, & if\ q_{u\breve{u}}^{od} > q, \text{that is}, \partial\Theta/\partial\chi_{u\breve{u}}^{od} > 0 \\ < 0, & if\ q_{u\breve{u}}^{od} < q, , \text{that is}, \partial\Theta/\partial\chi_{u\breve{u}}^{od} < 0 \end{cases}$$

This implies that $\tilde{\chi}_{u\breve{u}}^{od} \times \partial\Theta(W)/\partial\chi_{u\breve{u}}^{od} > 0$. Note that $\tilde{\chi}_{u\breve{u}}^{od}$ is the solution of Eq. (4.90); therefore, we have $\tilde{\chi}_{u\breve{u}}^{od} \times \partial\Theta\left(W-\tilde{\chi}_{u\breve{u}}^{od}\cdot e_l\right)/\partial\chi_{u\breve{u}}^{od} = 0$.

As $\Theta(W)$ is concave, we have $\Theta\left(W-\tilde{\chi}_{u\breve{u}}^{od}\cdot e_l\right) \geq \Theta\left(W-\tilde{\chi}_{u\breve{u}}^{od\prime}\cdot e_l\right), \tilde{\chi}_{u\breve{u}}^{od\prime} \in R$. Therefore, from Eq. (4.90), we have:

$$\tilde{\chi}_{u\breve{u}}^{od} = \frac{1}{\theta_{u\breve{u}}^{od}} \ln\left(\frac{q_{u\breve{u}}^{od}}{\sum_{k\in K_{u\breve{u}}^{od}} h_{u\breve{u},k}^{od}}\right)$$

This completes the proof. □

Proposition 4.5. If the optimality condition given by Eq. (78) is not satisfied, assume that $\epsilon_{u\breve{u}}^{od-}$ is the *l*th component of W; then, an adjustment $\tilde{\epsilon}_{u\breve{u}}^{od-}$ can be made to $\epsilon_{u\breve{u}}^{od-}$ such that

$\tilde{\epsilon}_{u\breve{u}}^{od-} = \max\left\{-\epsilon_{u\breve{u}}^{od-}, \rho_u^{od} \ln\left(\frac{\left(1-\varepsilon_{u\breve{u}}^{od}\right)\cdot z_{u\breve{u}}^{od}}{q_{u\breve{u}}^{od}}\right)\right\}\ ,\ \tilde{\epsilon}_{u\breve{u}}^{od-} \times \frac{\partial\Theta(W)}{\partial\epsilon_{u\breve{u}}^{od-}} > 0\ ,\ \tilde{\epsilon}_{u\breve{u}}^{od-} \times \frac{\partial\Theta\left(W+\tilde{\epsilon}_{u\breve{u}}^{od-}\cdot e_l\right)}{\partial\epsilon_{u\breve{u}}^{od-}} \geq 0\ ,$

$\left(W+\tilde{\epsilon}_{u\breve{u}}^{od-}\cdot e_l\right) \geq \left(W+\tilde{\epsilon}_{u\breve{u}}^{od-\prime}\cdot e_l\right), \tilde{\epsilon}_{u\breve{u}}^{od-\prime} \in D_l$, where $D_l = \left\{\epsilon_{u\breve{u}}^{od-} + \tilde{\epsilon}_{u\breve{u}}^{od-\prime} \geq 0\right\}$.

Proof. Eq. (4.78) can be rewritten as:

$$\begin{aligned}\frac{\partial \Theta(W)}{\partial \epsilon_{u\breve{u}}^{od-}} &= \left(1-\varepsilon_{\breve{u}\breve{u}}^{od}\right)\cdot z_{\breve{u}\breve{u}}^{od} - q_{u\breve{u}}^{od} \\ &= \left(1-\varepsilon_{\breve{u}\breve{u}}^{od}\right)\cdot z_{\breve{u}\breve{u}}^{od} \\ &\quad - \exp\left(\frac{1}{\rho_u^{od}}\left(\frac{1}{\theta_{u\breve{u}}^{od}}\ln\sum_{k'\in K_{u\breve{u}}^{od}}\exp\left(-\theta_{u\breve{u}}^{od}\left(c_{u\breve{u},k}^{od}+J_{u\breve{u},k}^{od}\right)\right)+\mathrm{MSC}_{u\breve{u}}^{od}\right.\right. \\ &\quad \left.\left. + \varpi_u^{od} + \epsilon_{u\breve{u}}^{od-} - \epsilon_{u\breve{u}}^{od+}\right)\right)\end{aligned} \tag{4.91}$$

We need to determine the value of $\tilde{\epsilon}_{u\breve{u}}^{od-}$ based on $\epsilon_{u\breve{u}}^{od-}$:

(1) When $\epsilon_{u\breve{u}}^{od-}=0$, we have $\partial\Theta(W)/\partial\epsilon_{u\breve{u}}^{od-}>0$ because Eq. (4.91) is violated. Following the proof in Proposition 4.4, we have

$$\tilde{\epsilon}_{u\breve{u}}^{od-} = \rho_u^{od}\ln\left(\frac{\left(1-\varepsilon_{\breve{u}\breve{u}}^{od}\right)\cdot z_{\breve{u}\breve{u}}^{od}}{q_{u\breve{u}}^{od}}\right)$$

Note that, in this case, $\tilde{\epsilon}_{u\breve{u}}^{od-}$ is always positive since $\left(1-\varepsilon_{\breve{u}\breve{u}}^{od}\right)\cdot z_{\breve{u}\breve{u}}^{od}/q_{u\breve{u}}^{od}>1$ (i.e., $\partial\Theta(W)/\partial\epsilon_{u\breve{u}}^{od-}>0$).

(2) When $\epsilon_{u\breve{u}}^{od-}>0$, Eq. (4.91) has to be 0. Similar to (1), we can have $\tilde{\epsilon}_{u\breve{u}}^{od-} = \rho_u^{od}\ln\left(\left(1-\varepsilon_{\breve{u}\breve{u}}^{od}\right)\cdot z_{\breve{u}\breve{u}}^{od}/q_{u\breve{u}}^{od}\right)$. $\tilde{\epsilon}_{u\breve{u}}^{od-}$ will be negative if Eq. (4.91) is negative (i.e., $\partial\Theta(W)/\partial\epsilon_{u\breve{u}}^{od-}<0$) which implies that using equation $\epsilon_{u\breve{u}}^{od-}=\epsilon_{u\breve{u}}^{od-}+\tilde{\epsilon}_{u\breve{u}}^{od-}$ to update $\epsilon_{u\breve{u}}^{od-}$ may violate the constraint $\epsilon_{u\breve{u}}^{od-}\geq 0$. To avoid this violation, we set

$$\tilde{\epsilon}_{u\breve{u}}^{od-} = \begin{cases} \rho_u^{od}\ln\left(\left(1-\varepsilon_{\breve{u}\breve{u}}^{od}\right)\cdot z_{\breve{u}\breve{u}}^{od}/q_{u\breve{u}}^{od}\right), & \text{if } \left(1-\varepsilon_{\breve{u}\breve{u}}^{od}\right)\cdot z_{\breve{u}\breve{u}}^{od} > q_{u\breve{u}}^{od}, \text{that is } \partial\Theta(W)/\partial\epsilon_{u\breve{u}}^{od-}>0 \\ \max\left\{-\epsilon_{u\breve{u}}^{od-}, \rho_u^{od}\ln\left(\left(1-\varepsilon_{\breve{u}\breve{u}}^{od}\right)\cdot z_{\breve{u}\breve{u}}^{od}/q_{u\breve{u}}^{od}\right)\right\}, & \text{if } \left(1-\varepsilon_{\breve{u}\breve{u}}^{od}\right)\cdot z_{\breve{u}\breve{u}}^{od} < q_{u\breve{u}}^{od}, \text{that is } \partial\Theta(W)/\partial\epsilon_{u\breve{u}}^{od-}<0 \end{cases}$$

Therefore, based on (1) and (2), $\tilde{\epsilon}_{u\breve{u}}^{od-} = \max\left\{-\epsilon_{u\breve{u}}^{od-}, \rho_u^{od}\ln\left(\left(1-\varepsilon_{\breve{u}\breve{u}}^{od}\right)\cdot z_{\breve{u}\breve{u}}^{od}/q_{u\breve{u}}^{od}\right)\right\}$, and $\left(W+\tilde{\epsilon}_{u\breve{u}}^{od-}\cdot e_l\right)\geq\left(W+\tilde{\epsilon}_{u\breve{u}}^{od-\prime}\cdot e_l\right), \tilde{\epsilon}_{u\breve{u}}^{od-\prime}\in D_l$, where $D_l=\left\{\epsilon_{u\breve{u}}^{od-}+\tilde{\epsilon}_{u\breve{u}}^{od-\prime}\geq 0\right\}$.

This completes the proof. □

Table 4.1 Adjustment factors of dual variables

Dual variable	Equation number	Adjustment factor
ϖ_u^{od}	(75)	$\tilde{\varpi}_u^{od} = \rho_u^{od}\ln\left(\frac{q_u^{od}}{\sum_{\breve{u}\in\breve{U}_u^{od}} q_{u\breve{u}}^{od}}\right)$
τ^o	(76)	$\tilde{\tau}^o = -\ln\left(\frac{z^o}{P^o}\right)$

ϱ^d	(77)	$\tilde{\varrho}^d = -\ln\left(\frac{z^d}{A^d}\right)$
$\epsilon_{u\breve{u}}^{od+}$	(79)	$\tilde{\epsilon}_{u\breve{u}}^{od+} = \max\left\{-\epsilon_{u\breve{u}}^{od+}, -\rho_u^{od}\ln\left(\frac{\left(1+\varepsilon_{u\breve{u}}^{od}\right)\cdot z_{u\breve{u}}^{od}}{q_{u\breve{u}}^{od}}\right)\right\}$
$\omega_{c,s}$	(80)	$\tilde{\omega}_{c,s} = \max\left\{-\omega_{c,s}, -\frac{1}{\theta_c^{od}}\ln\left(\frac{C_{c,s}-F_{c,s}}{v_{c,s}}\right)\right\}$
$\omega_{b,e}$	(81)	$\tilde{\omega}_{b,e} = \max\left\{-\omega_{b,e}, -\frac{1}{\theta_b^{od}}\ln\left(\frac{C_{b,e}}{v_{b,e}}\right)\right\}$
$\omega_{m,e}$	(82)	$\tilde{\omega}_{m,e} = \max\left\{-\omega_{m,e}, -\frac{1}{\theta_m^{od}}\ln\left(\frac{C_{m,e}}{v_{m,e}}\right)\right\}$
$\mu_{c,s}^-$	(83)	$\tilde{\mu}_{c,s}^- = \max\left\{-\mu_{c,s}^-, \frac{1}{\theta_c^{od}}\ln\left(\frac{\left(1-\varepsilon_{c,s}\right)\cdot\bar{v}_{c,s}-F_{c,s}}{v_{c,s}}\right)\right\}$
$\mu_{c,s}^+$	(84)	$\tilde{\mu}_{c,s}^+ = \max\left\{-\mu_{c,s}^+, -\frac{1}{\theta_c^{od}}\ln\left(\frac{\left(1+\varepsilon_{c,s}\right)\cdot\bar{v}_{c,s}-F_{c,s}}{v_{c,s}}\right)\right\}$
$\psi_{b,e}^-$	(85)	$\tilde{\psi}_{b,e}^- = \max\left\{-\psi_{b,e}^-, \frac{1}{\theta_b^{od}}\ln\left(\frac{\left(1-\varepsilon_{b,e}\right)\cdot\bar{v}_{b,e}}{v_{b,e}}\right)\right\}$
$\psi_{b,e}^+$	(86)	$\tilde{\psi}_{b,e}^+ = \max\left\{-\psi_{b,e}^+, -\frac{1}{\theta_b^{od}}\ln\left(\frac{\left(1+\varepsilon_{b,e}\right)\cdot\bar{v}_{b,e}}{v_{b,e}}\right)\right\}$
$\xi_{m,e}^-$	(87)	$\tilde{\xi}_{m,e}^- = \max\left\{-\xi_{m,e}^-, \frac{1}{\theta_m^{od}}\ln\left(\frac{\left(1-\varepsilon_{m,e}\right)\cdot\bar{v}_{m,e}}{v_{m,e}}\right)\right\}$
$\xi_{m,e}^+$	(88)	$\tilde{\xi}_{m,e}^+ = \max\left\{-\xi_{m,e}^+, -\frac{1}{\theta_m^{od}}\ln\left(\frac{\left(1+\varepsilon_{m,e}\right)\cdot\bar{v}_{m,e}}{v_{m,e}}\right)\right\}$

4.4.3.3 Detailed steps

Iterative balancing is centered on the use of dual variables to analytically determine primal variables. By initializing the dual variables, we can obtain the initialized primal variables according to their analytical expressions. After the adjustment factors for the dual variables are obtained via the dual formulation, the updated dual variables can be used to analytically determine the primal variables. When the convergent (or stopping) criterion is reached, the solution (primal and dual variables) will be obtained. Although solution frameworks of iterative balancing have been well presented (Chen et al., 2005, 2009, 2010), they focus on a one-layer structure (i.e., only for the route choice). Thus, it is necessary to present the detailed steps of three-layer iterative balancing (Figure 4.4). The convergence of the iterative balancing scheme with NL model has been proved by Li (2016) (see Proposition 3.11). The algorithm has been coded using C# in Microsoft Visual Studio 2015 and run on a computer with a 2.7 GHz processor and 20.00 GB RAM.

To supplement the steps in the flow chart (Figure 4.4), the analytical expressions of $h_{u\breve{u},k}^{od}$, $q_{u\breve{u}}^{od}$ and q_u^{od} are used to represent the combined mode and route choices in the MM-PFE:

$$h_{u\breve{u},k}^{od} = \exp\left(-\theta_{u\breve{u}}^{od}\left(c_{u\breve{u},k}^{od} - \chi_{u\breve{u}}^{od} + J_{u\breve{u},k}^{od}\right)\right) = \Xi_h\left(\chi_{u\breve{u}}^{od}, J_{u\breve{u},k}^{od}\right) \tag{4.92}$$

$$q_{u\breve{u}}^{od} = \exp\left(\frac{1}{\rho_u^{od}}\left(\frac{1}{\theta_{u\breve{u}}^{od}}\ln\sum_{k'\in K_{u\breve{u}}^{od}}\exp\left(-\theta\left(c_{u\breve{u},k}^{od} + J_{u\breve{u},k}^{od}\right)\right) + \mathrm{MSC}_{u\breve{u}}^{od} + \varpi_u^{od} + \epsilon_{u\breve{u}}^{od-} - \epsilon_{u\breve{u}}^{od+}\right)\right) = \Xi_{q_{u\breve{u}}}\left(\varpi_u^{od}, \epsilon_{u\breve{u}}^{od-}, \epsilon_{u\breve{u}}^{od+}, J_{u\breve{u},k}^{od}\right) \tag{4.93}$$

$$q_u^{od} = \exp(\rho_u^{od} I_u^{od}) \cdot \exp(-\tau^o) \cdot \exp(-\varrho^d) = \Xi_{q_u}\left(\tau^o, \varrho^d, \epsilon_{u\breve{u}}^{od-}, \epsilon_{u\breve{u}}^{od+}, J_{u\breve{u},k}^{od}\right) \tag{4.94}$$

Algorithm. Three-layer iterative balancing scheme

1: Initialization.

(a) Set the value of all dual variables as 0 and $n = 1$.

(b) Compute primal variables:

$h_{u\breve{u},k}^{od} = \Xi_h(\chi_{u\breve{u}}^{od}, J_{u\breve{u},k}^{od}), \forall k \in K_{u\breve{u}}^{od}, \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in OD$

$q_{u\breve{u}}^{od} = \Xi_{q_{u\breve{u}}}(\varpi_u^{od}, \epsilon_{u\breve{u}}^{od-}, \epsilon_{u\breve{u}}^{od+}, J_{u\breve{u},k}^{od}), \forall \breve{u} \in \breve{U}_u^{od}, u \in U^{od}, od \in OD$

$q_u^{od} = \Xi_{q_u}(\tau^o, \varrho^d, \epsilon_{u\breve{u}}^{od-}, \epsilon_{u\breve{u}}^{od+}, J_{u\breve{u},k}^{od}), \forall u \in U^{od}, od \in OD$

2: while $n < Iter_{max}$ and $\Delta_1 \geq \eta$ and $\Delta_2 < \overline{\eta}$ do

3: # update dual variables related to nest and calculate nest flows

```
for o in O
    update τ^o
    for u in U^od
        q_u^od = Ξ_{q_u}(τ^o, ϱ^d, ε_{uŭ}^{od-}, ε_{uŭ}^{od+}, J_{uŭ,k}^{od})
    end for
end for
```

```
for d in D
    update ϱ^d
    for u in U^od
        q_u^od = Ξ_{q_u}(τ^o, ϱ^d, ε_{uŭ}^{od-}, ε_{uŭ}^{od+}, J_{uŭ,k}^{od})
    end for
end for
```

Layer 1

4: # update dual variables related to mode and calculate mode flows

```
for od in OD
  for u in U^od
    update ϖ_u^od
    for ŭ in ŭ ∈ Ŭ_u^od
      q_{uŭ}^od = Ξ_{q_{uŭ}}(ϖ_u^od, ε_{uŭ}^{od-}, ε_{uŭ}^{od+}, J_{uŭ,k}^od)
    end for
  end for
end for
```

```
for od in OD
  for u in U^od
    for ŭ in ŭ ∈ Ŭ_u^od
      update ε_{uŭ}^{od-}, ε_{uŭ}^{od+}
      q_{uŭ}^od = Ξ_{q_{uŭ}}(ϖ_u^od, ε_{uŭ}^{od-}, ε_{uŭ}^{od+}, J_{uŭ,k}^od)
    end for
  end for
end for
```

Layer 2

5: # update dual variables related to mode and calculate path flows

```
for od in OD
  for u in U^od
    for ŭ in ŭ ∈ Ŭ_u^od
      update χ_{uŭ}^od
      for k in K_{uŭ}^od
        h_{uŭ,k}^od = Ξ_h(χ_{uŭ}^od, J_{uŭ,k}^od)
      end for
    end for
  end for
end for
```

Layer 3

6: # update dual variables related to arcs and calculate path flows

```
for s in S_c
  update ω_{c,s}, μ_{c,s}^-, μ_{c,s}^+
  for od in OD
    for k in K_c^od
      h_{c,k}^od = Ξ_h(χ_c^od, J_{c,k}^od)
    end for
  end for
end for
```

```
for e in E_b
  update ω_{b,e}, ψ_{b,e}^-, ψ_{b,e}^+
  for od in OD
    for k in K_c^od
      h_{b,k}^od = Ξ_h(χ_b^od, J_{b,k}^od)
    end for
  end for
end for
```

```
for e in E_m
  update ω_{m,e}, ξ_{m,e}^-, ξ_{m,e}^+
  for od in OD
    for k in K_c^od
      h_{m,k}^od = Ξ_h(χ_m^od, J_{m,k}^od)
    end for
  end for
end for
```

calculate mode and nest flow with the updated dual variables related to arcs

```
for o in O
  for u in U^od
    q_u^od = Ξ_{q_u}(τ^o ϱ^d, ε_{uŭ}^{od-}, ε_{uŭ}^{od+}, J_{uŭ,k}^od)
    for ŭ in ŭ ∈ Ŭ_u^od
      q_{uŭ}^od = Ξ_{q_{uŭ}}(ϖ_u^od, ε_{uŭ}^{od-}, ε_{uŭ}^{od+}, J_{uŭ,k}^od)
    end for
  end for
end for
```

7: Determine the maximum adjustment Δ_1 of all dual variables and maximum dual variable Δ_2

8: n← $n + 1$

9: end while

10: where $Iter_{max}$ is the maximum iteration number and $\eta, \overline{\eta} \in R_+$ are predetermined tolerance.

Figure 4.4 Algorithm of the three-layer iterative balancing scheme

4.5 Numerical Experiment

In this subchapter, the hypothetical multi-modal transportation network of Sioux Falls is used to illustrate the features of the proposed model and the solution algorithm performance.

According to Vovsha (1997), the mode-specific constants can be calibrated using the household survey data, and they are assumed as: $\mathrm{MSC}_c^{od} = 0$, $\mathrm{MSC}_b^{od} = 8$, and $\mathrm{MSC}_m^{od} = 12$ in this paper. The parameters are set as: $\rho_u^{od} = 0.5$, $\theta_{u\tilde{u}}^{od} = 1.5$, and $PCE = 3$.

The travel time functions for the road link and route section for the multi-modal transportation network are as follows:

- Road network: $c_{c,s} = t_{c,s}^0 \left(1 + 0.15 \cdot \left(\frac{F_{c,s}+v_{c,s}}{C_{c,s}}\right)^4\right), \forall s \in S_c$
- Bus network: $\phi_{b,s}(\mathbf{v}) = 10 \cdot \left(\frac{v_{b,s}+\tilde{v}_{b,s}}{\sum_{l\in A_b^s} f_l \kappa_{b,l}}\right), \forall s \in S_b$
- Metro network: $\phi_{m,s}(\mathbf{v}) = 10 \cdot \left(\frac{v_{m,s}+\tilde{v}_{m,s}}{\sum_{l\in A_m^s} f_l \kappa_{m,l}}\right), \forall s \in S_m$

4.5.1 Experiment setting

The multi-modal transportation network of Sioux Falls is displayed in Figure 4.5. It comprises a highway network with 76 directed links and 24 nodes (downloaded from http://www.bgu.ac.il/~bargera/tntp/), a metro network with 5 metro lines (10 itineraries), and a bus network with 9 bus lines (18 itineraries). In Table 2, there are 32 OD pairs given with 8 original zones. The path set of the private car mode was generated using seSue, an open-source software program obtained from Ahipasaoglu et al. (2016). A combination of the link penalty and link elimination methods is used here, and the default setting is adopted. The maximum number of transfers in metro and bus networks is 2, and the maximum number of paths between each OD pair is 20 (metro and bus) and 30 (car).

We have generated the inputs of the model in the multi-modal transportation network with some supposed parameters and zonal production and attraction. Specifically, the input used comprises the link counts (i.e., road link traffic counts, onboard passenger counts of bus and metro line segments), mode-specific target OD demand, and zonal production and attraction observations. Without loss of generality, the link counts and mode-specific target OD demand are randomly multiplied by a number between 0.85 and 1.15. The error bounds for these two kinds of observed data are set as $[-15\%, +15\%]$ and $[-10\%, +10\%]$, respectively.

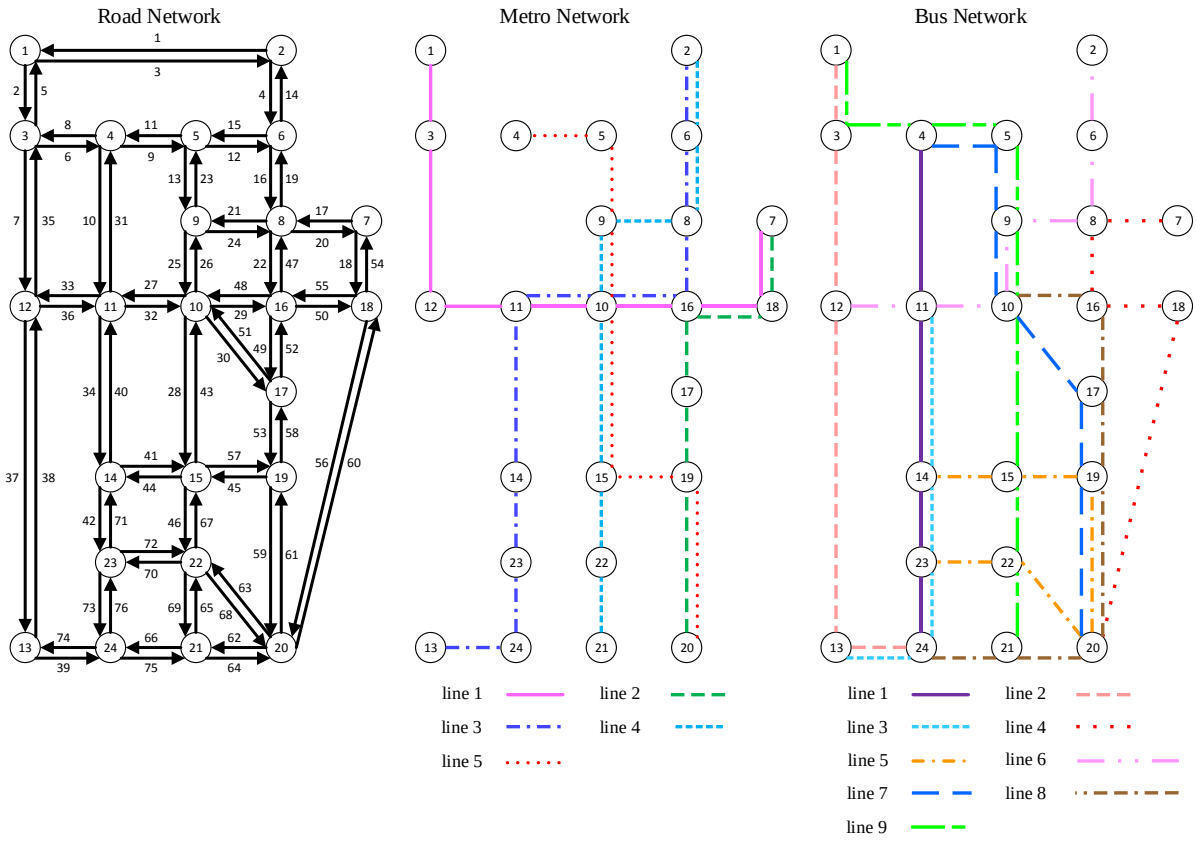


Figure 4.5 Hypothetical multi-modal transportation network of Sioux Falls

Table 4.2 OD pair setting in Sioux Falls

OD pair	OD pair	OD pair	OD pair
(1, 13)	(3, 13)	(13, 1)	(21, 1)
(1, 20)	(3, 20)	(13, 2)	(21, 2)
(1, 21)	(3, 21)	(13, 3)	(21, 3)
(1, 24)	(3, 24)	(13, 4)	(21, 4)
(2, 13)	(4, 13)	(20, 1)	(24, 1)
(2, 20)	(4, 20)	(20, 2)	(24, 2)
(2, 21)	(4, 21)	(20, 3)	(24, 3)
(2, 24)	(4, 24)	(20, 4)	(24, 4)

4.5.2 Result analysis

This subchapter presents the result to demonstrate the applicability of the proposed model and evaluate the capability of the algorithm for finding good solutions.

4.5.2.1 Convergence characteristics

To evaluate the performance of the solution algorithm, we not only plot the convergence curve of the root mean square error (RMSE) of the mode-specific path flow for the proposed model, but also randomly select four constraints to illustrate the convergence characteristics of the corresponding dual variables. The convergence curves are presented in Figure 4.6.

All three curves of RMSE versus the number of iterations decrease below 1E-6 in a few iterations. The values of the selected dual variables at each iteration are also shown in Figure 4.6. All of the dual variables converge. The values of all selected dual variables are larger than or equal to 0:

- The dual variable corresponding to origin 2 (zonal production) is positive, with a value of 21.08; this indicates that origin 2 has an intrinsic attractiveness to travelers.
- The dual variable corresponding to target OD pair (1, 21) (OD demand) of the metro mode equals 0, which means that the estimated OD demand of the metro mode is within the specific bound (i.e., an internal solution).
- The dual variable (upper bound) corresponding to road link 74 (flow) is 7.36, larger than 0, which indicates that the estimated flow of road link 74 reaches the upper bound.
- The dual variable (upper bound) corresponding to line segment 1 of bus line 3 (flow) is 0.44, which indicates that the estimated flow of this bus line segment reaches the upper bound.

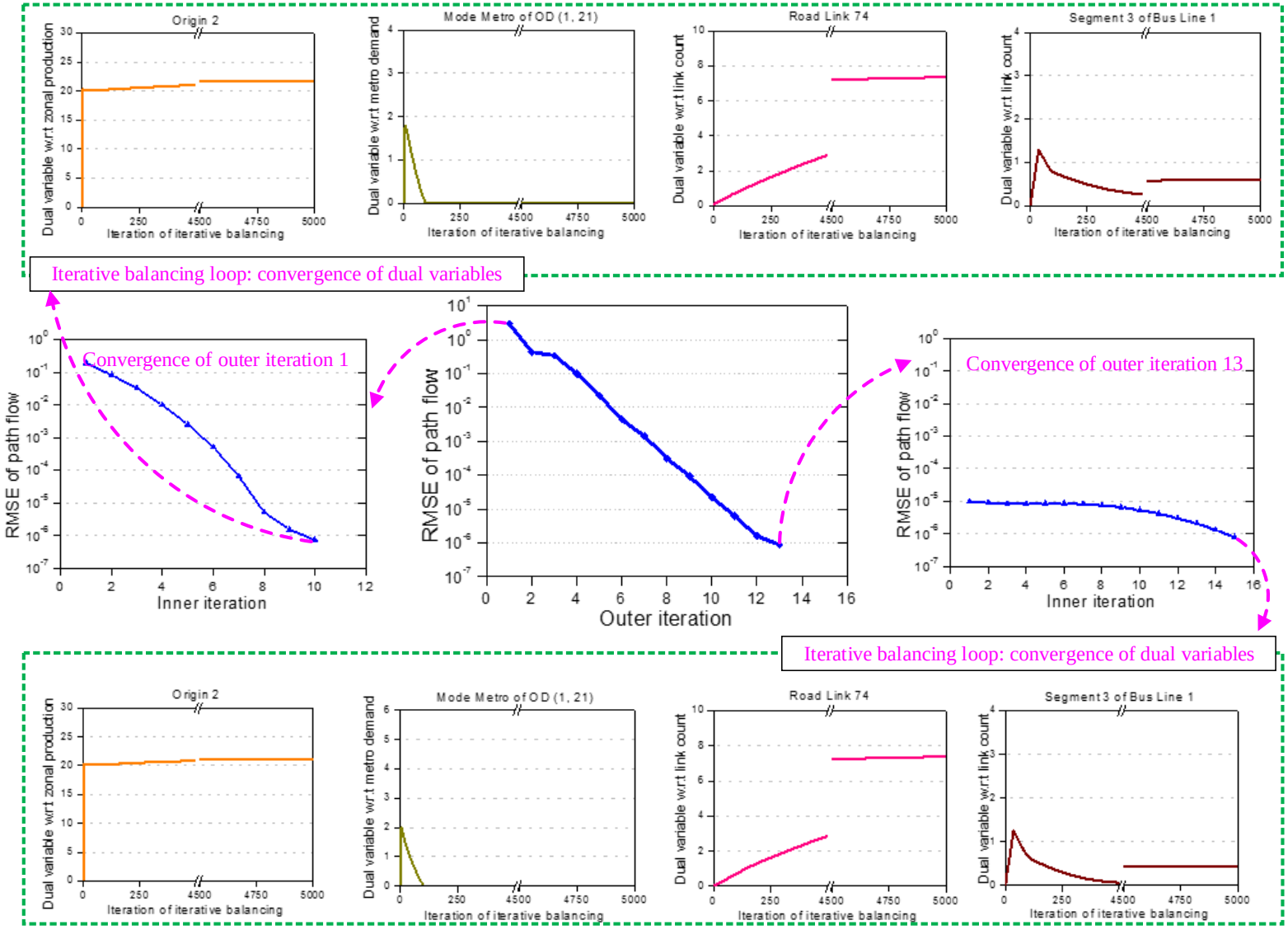
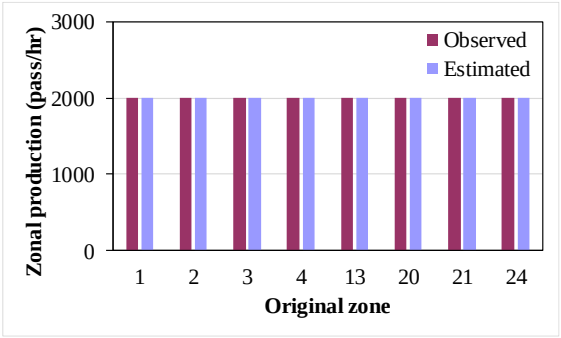


Figure 4.6 Convergence characteristics of the solution algorithm

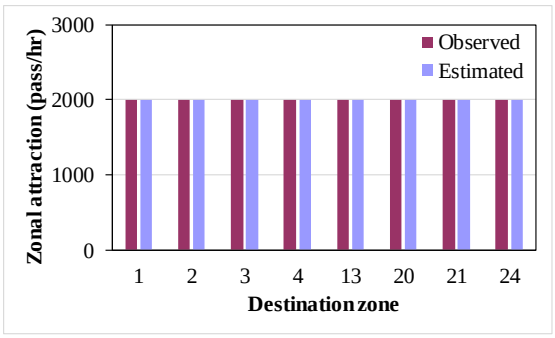
4.5.2.2 Model accuracy

The model accuracy is verified from four perspectives: zonal production and attraction, target mode-specific OD demand, estimated mode-specific OD demand, and observed link (line segment) flow.

- Figure 4.7 compares the observed zonal productions and attractions with those calculated using the estimated mode-specific OD demand. The observed zonal productions and attractions are equal to the estimated values. This indicates that both the estimated zonal production and attraction meet their constraints, i.e., Eqs. (4.22)-(4.23).
- Figure 4.8 compares the observed target mode-specific OD demand with the estimated values. The scatters lie in the region of [-10%, +10%] of the observed count, which meets the requirement of the constraint in Eq. (4.26).
- Figure 4.9 compares the estimated mode-specific OD demand with those calculated using the estimated mode-specific path flow. The scatter points lie on the 45° line (i.e., $Y = X$). This is consistent with the flow conservation constraint in Eq. (4.19).
- Figure 4.10 compares the observed and estimated link (line segment) flows. The scatters lie in the region of [-15%, +15%] of the observed count. This also accords with Eqs. (4.27)-(4.29).

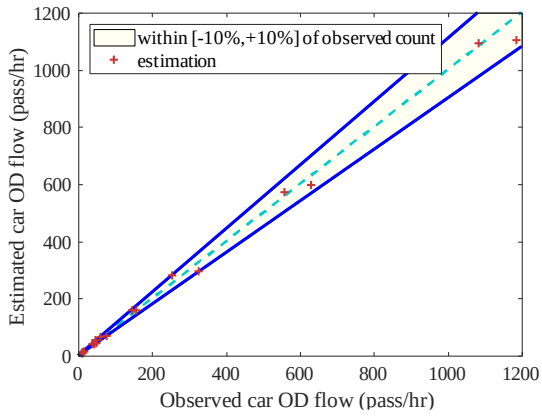


(a) Original zone

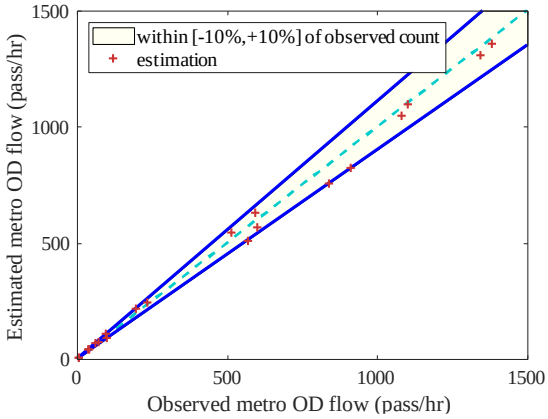


(b) Destination zone

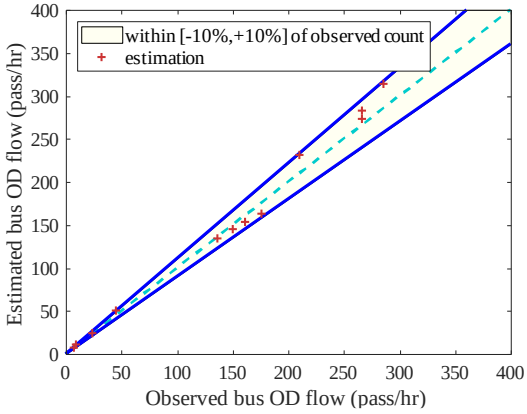
Figure 4.7 Comparison between the observed zonal productions and attractions and those calculated using the estimated mode-specific OD demand



(a) Mode car of OD pair

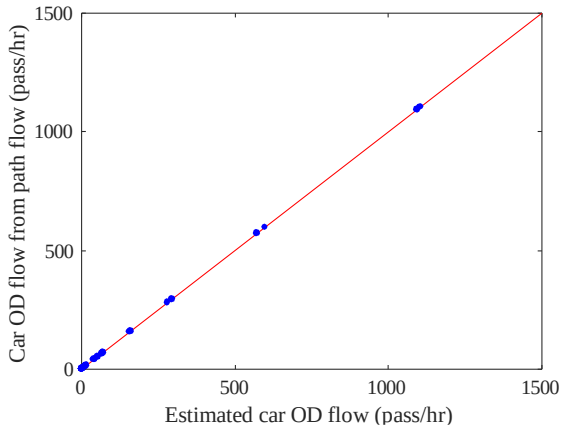


(b) Mode metro of OD pair

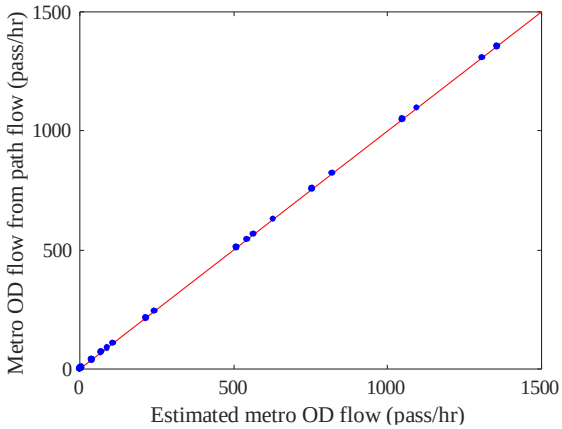


(c) Mode bus of OD pair

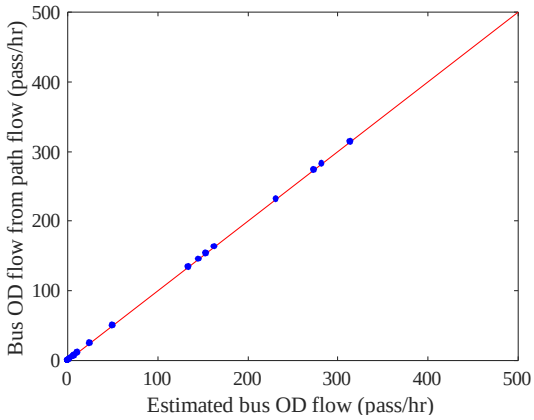
Figure 4.8 Comparison between the observed and estimated target mode-specific OD demands



(a) Mode car of OD pair

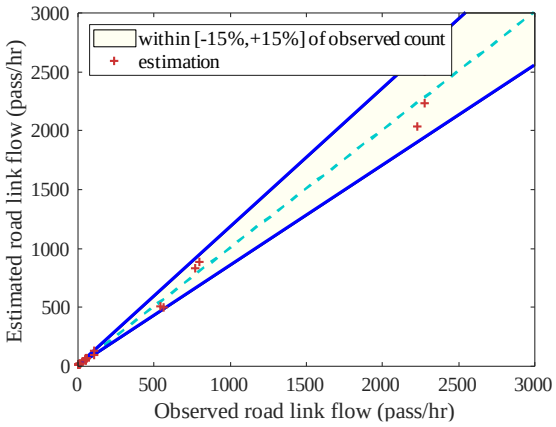


(b) Mode metro of OD pair

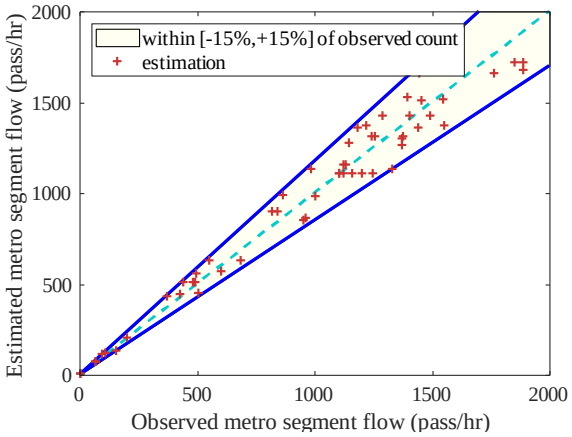


(c) Mode bus of OD pair

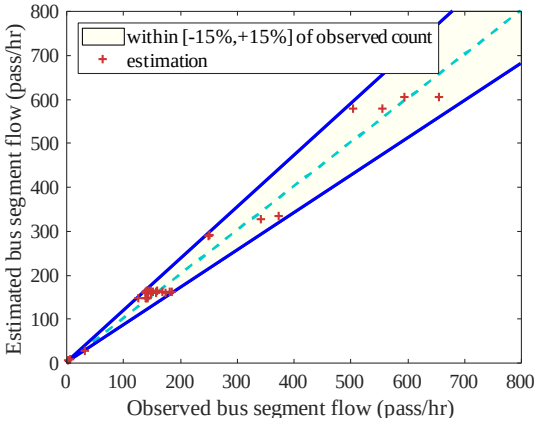
Figure 4.9 Comparison between the estimated mode-specific OD demands and those calculated using the estimated mode-specific path flow



(a) Road link



(b) Metro line segment



(c) Bus line segment

Figure 4.10 Comparison between the observed and estimated link (line segment) flows

4.6 Chapter Summary

This chapter addresses the multi-modal OD demand estimation problem in urban transportation networks. A NL-based PFE is formulated as a VI problem based on a single-level structure. In the proposed model, the interaction of private cars and bus vehicles, private car and transit route choice behaviours, and mode similarity are modeled. Moreover, the estimated demand distribution in the optimal solution is proved: an NL model is used for the mode demand, and an MNL model is used for the mode-specific path flow of each OD pair. The model input for demand estimation includes information about the link counts (i.e., road link traffic counts, metro and bus line segment flow observations), the mode-specific target OD demand, and zonal production and attraction observations.

A diagonalization approach is developed for the proposed MM-PFE model. Using the diagonalized approach, the variation inequality formulation is reformulated as a convex optimization problem in each diagonalized iteration. Then, the reformulated problem is solved using a path-based partial linearization algorithm embedded with a three-layer iterative balancing scheme, which can handle various inequality/equality side constraints.

The MM-PFE is applied to a hypothetical multi-modal transportation network based on the Sioux Falls network. The numerical results demonstrate that the proposed MM-PFE can obtain a suitable multi-modal OD trip matrix with limited available observations and that the developed solution algorithm can solve the proposed model.

CHAPTER 5

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

This chapter summarizes the work and conclusions achieved in this thesis and presents several future research directions.

5.1 Conclusion

The research presented in this thesis focuses on multi-modal travel demand estimation methodology to address the weaknesses of the traditional transport planning in a sequential manner and activity-based models. Compared with the traditional four-step model, the integrated PFE framework is consistent and can incorporate various side constraints related to multiple data sources. Compared with the activity-based model, the integrated PFE framework not only take less time for computation, but also can use the information contained within the observational data and express the path flow (mode-specific OD demand) analytically. Three research problems are comprehensively modeled: (a) transit travel behavior; (b) transit origin-destination (OD) demand estimation; and (c) multi-modal demand estimation. Three research components are undertaken to answer the abovementioned problems as follows.

(1) Development of a strategy-based stochastic transit equilibrium model with capacity and number-of-transfers constraints

Chapter 2 presents a strategy-based transit stochastic user equilibrium model with capacity and number-of-transfers constraints. The logit-based stochastic path choice behavior and in-vehicle congestion cost are taken into account, a strict capacity constraint of transit line segments is added to handle the overload problem, and a number-of-transfers constraint is

considered for transit path finding. This transit equilibrium problem is formulated as a logit-based variational inequality problem. A transit path-set generation procedure based on the *k*-shortest path algorithm is introduced, which also considers the features of a route-section-based transit network and number-of-transfers constraint. The diagonalization method is adopted to solve the proposed model, and the diagonalized subproblem is solved using a path-based partial linearization solution algorithm embedded with an iterative balancing scheme to handle the capacity constraints.

Numerical examples are provided to demonstrate the features of the proposed model and evaluate the performance of the developed solution algorithm. The results indicate that the number of transfers constraint changes the components of the transit path set, which strongly impacts the passenger flow patterns. The results show that the line capacity constraint also affects flow patterns, which revises the evaluation of some transit management strategies. The results from the real-case transit network further verify the applicability of the developed solution algorithm. Overall, the numerical examples highlight the importance of using capacity and number-of-transfers constraints in transit equilibrium problems.

(2) Frequency-based path flow estimator for transit OD demand estimation

Chapter 3 presents a frequency-based path flow estimator (PFE) framework to estimate the OD trip matrix in a transit network. A frequency-based transit PFE formulation in variational inequality form is proposed that incorporate the observed partial OD demand matrices and onboard passenger flow observations as side constraints. The observed partial OD demand matrix is inferred from the automatic fare collection (i.e., smartcard) data together with automatic vehicle location (AVL) data based on the inference of the alighting stops, while the observed onboard passenger flow is calculated from the automatic passenger counting and AVL data. The diagonalization method is adopted to solve the proposed model, and the diagonalized subproblem is solved using a path-based partial linearization solution algorithm embedded with an iterative balancing scheme to handle the various side constraints.

Numerical examples are provided to illustrate the performance of the proposed model and its applicability in a real-world transit network. The results show that the configurations of the onboard passenger counts and observed partial OD flows affect the estimated results. Overall, the large-network results indicate that the OD flow estimations can be improved using a combination of the onboard passenger counts and observed partial OD flows.

(3) Multi-modal PFE for OD demand estimation

In Chapter 4, a multi-modal PFE with a nested logit (NL) choice model is formulated as a variational inequality problem based on a single-level structure. The proposed model

addresses the interaction of private cars and bus vehicles, route choice behavior of private cars and transit modes, and mode similarity. The estimated demand distribution in the optimal solution is verified. An NL model is used to estimate the mode demand, and a multinomial logit model is used to estimate the mode-specific path flow of each OD pair. The model input for the demand estimation includes information regarding the link counts (e.g., road link traffic counts and metro and bus line segment flow observations), mode-specific target OD demand, and zonal production and attraction observations.

A diagonalization approach is developed for the proposed multi-modal PFE model that reformulates the variational inequality formulation as a convex optimization problem in each diagonalized iteration. The reformulated problem is then further solved using a path-based partial linearization algorithm embedded with a three-layer iterative balancing scheme, which can handle various inequality/equality side constraints.

The multi-modal PFE is applied to a hypothetical multi-modal transportation network based on the Sioux Falls (USA) network. The numerical results demonstrate that the proposed MM-PFE can obtain a suitable multi-modal OD trip matrix with limited available observations, and the developed solution algorithm is capable of solving the proposed model.

5.2 Further Research Directions

In this subchapter, we discuss two main potential further: extended integrated PFE and integrated PFE applications.

5.2.1 Extended integrated PFE

(1) Trip-chain-based PFE: Trip chaining behavior has been incorporated into network equilibrium models (Maruyama and Harata, 2005, 2006; Maruyama and Sumalee, 2007; He et al., 2015; Lu et al., 2015; Shimamoto et al., 2016; Wang et al., 2016; Xie et al., 2017; Gao et al., 2019). However, the issue of demand inconsistency might arise in the travel demand estimation problem due to the ignored linkage between adjacent trips. Hence, one further direction is to explore a PFE that considers trip chaining behavior. According to Primerano et al. (2008), trip chains contain more general patterns with more secondary activities for which the scheduling order is flexible. This requires a comprehensive integrated trip-chain-based PFE framework to capture these features.

(2) Advanced discrete choice model: The models proposed in this thesis adopt the logit model for passenger route choice behavior, which assumes that random error terms are

independent and identically distributed. This raises two issues of route overlapping and route-specific perception variance. Thus, one possible direction is to relax the assumption of identical distribution and adopt a weibit-based model (Castillo et al., 2008; Kitthamkesorn and Chen, 2013, 2014; Kitthamkesorn et al., 2015).

(3) Real network validation: The proposed models were tested with hypothetical transportation networks, which demonstrates the necessity to integrate data from a real-case network to calibrate the model parameters. The proposed models adopt the predetermined measurement errors for the known information, which is not quite suitable for the real data case, and norm approximation techniques (Chen et al., 2009) can be developed to handle the various data inconsistencies. A direct comparison between the estimated and observed values is not possible because the true OD demands (benchmarks) are unknown. The confidence interval estimation approach (Chootinan and Chen, 2011) can be adopted in a multi-modal transportation network to assess the reliability of the demand estimation results.

(4) More travel choices: More mode choices exist in a multi-modal urban transportation system (e.g., more transit modes, park and ride, and ride-sharing/ride-hailing). For example, the pick-up/drop-off behavior in ride-sharing mode should be considered, especially for the case of multiple pick-ups/drop-offs for an individual ride-sharing driver, and a more in-depth integrated PFE should be proposed.

5.2.2 Applications of an integrated PFE

This thesis focuses on an integrated PFE framework for multi-modal travel demand estimation, which also serves as a network equilibrium analysis tool with corresponding side constraints. Network equilibrium models are important for other urban disciplines including transportation management and land use.

(1) Network and service operation design: For existing bi-level transportation network (e.g., new roads and bus network) and service operation design (e.g., pricing for connected and autonomous vehicles, transit line frequency, and subsidies for transit passengers) problems, the lower-level model usually adopts a network equilibrium model which does not contain the traffic count information related to the historical demand. A corresponding PFE model can thus fill this gap. Moreover, the above nonlinear bi-level models are consistently solved heuristically. An exact algorithmic framework can be developed for these cases using a nested branch-and-bound tree and the piecewise linear approximation method (Dan et al., 2021).

(2) Combined PFE and land use: Transportation and land use are two closely connected components in urban cities. Mobility or accessibility will affect land use, and land use changes the travel demand distribution, which can affect the mobility or accessibility. Hence, it is important to explore the possibility of an integrated framework of a PFE and land use model to make good use of the advantages of the PFE approach, especially in regions with accelerating urbanization.

REFERENCES

Alameda-Contra Costa Transit District and Metropolitan Transportation Commission, 2018. 2017 AC transit on-board transit survey. http://www.actransit.org/wp-content/uploads/board_memos/17-231a%20Rider%20Survey%20Atch%201.pdf.

An, Q., Fu, X., Huang, D., Cheng, Q., Liu, Z. Y., 2020. Analysis of adding-runs strategy for peak-hour regular bus services. *Transportation Research Part E: Logistics and Transportation Review* 143, 102100. https://doi.org/10.1016/j.tre.2020.102100

Ahipasaoglu, S. D., Arikan, U., Natarajan, K., 2016. On the flexibility of using marginal distribution choice models in traffic equilibrium. *Transportation Research Part B: Methodological* 91, 130-158. https://doi.org/10.1016/j.trb.2016.05.002

Babazadeh, A., Aashtiani, H .Z., 2005. Algorithm for equilibrium transit assignment problem. *Transportation Research Record: Journal of the Transportation Research Board* 1923, 227-235. https://doi.org/10.1177/0361198105192300124

Babazadeh, A., Khodakarami, M. A., Aashtiani, H. Z., 2010. Passenger origin-destination matrix estimation employing the path-based formulation of transit assignment problem. Presented at the *89th Aunual Meeting of Transportation Research Board*, Washington D.C., Jannuary 10-14.

Barry, J. J., Newhouser, R., Rahbee, A., Sayeda, S., 2002. Origin and destination estimation in New York City with automated fare system data. *Transportation Research Record* 1817, 183-187. https://doi.org/10.3141/1817-24

Bell, M. G. H., 1995. Stochastic user equilibrium assignment in network with queues. *Transportation Research Part B: Methodological* 29, 125-137. https://doi.org/10.1016/0191-2615(94)00030-4

Bell, M. G. H, Cassir, C., 1998. The use of the path flow estimator in multi-modal networks. In *Proceedings of the 1998 Conference on Traffic and Transportation Studies*, ICTTS, July 27, 1998-July 29, 1998. Beijing, China: ASCE.

Bell, M. G. M., Cassir, C., Iida, Y., Lam, W. H. K., 1999. A sensitivity based approach to network reliability assessment. In *Proceedings of the 14th International Symposium of Transportation and Traffic Theory*, edited by A. Ceder, Elsevier, 283-300.

Bell, M. G. H., Iida, Y., 1997. Transportation Network Analysis. John Wiley & Sons, New York.

Bell, M. G. M., Lam, W. H. K., Iida, Y. 1996. A time-dependent multi-class path flow estimator. In *Proceedings of the 13th International Symposium of Transportation and Traffic Theory*, edited by J.B. Lesort, Elsevier, 173-193.

Bell, M. G. H., Liu, X., Angeloudis, P., Fonzone, A., Hosseinloo, S. H., 2011. A frequency-based maritime container assignment model. *Transportation Research Part B: Methodological* 45(8), 1152-1161. https://doi.org/10.1016/j.trb.2011.04.002

Bell, M. G. H., Shield, C. M., Busch, F., Kruse, G., 1997. A stochastic user equilibrium path flow estimator. *Transportation Research Part C: Emerging Technologies* 5(34), 197-210. https://doi.org/10.1016/S0968-090X(97)00009-0

Beltrán, P., Cortes, C., Gschwender, A., Ibarra, R., Munizaga, M., Ortega, M., Palma, C., Zuñiga, M., 2011. Obtención de información valiosa a partir de datos de Transantiago, XV Congreso Chileno de Ingeniería de Transporte.

Ben-Akiva M., Lerman, S. R., 1985. Discrete Choice Analysis. MIT Press, Cambridge.

Ben-Akiva, M. E., Macke, P. P., Hsu, P. S., 1985. Alternative methods to estimate route-level trip tables and expand on-board surveys. *Transportation Research Record Journal of the Transportation Research Board* 1037(1037), 1-11.

Bouzaïene-Ayari, B., Gendreau, M., Nguyen, S., 1995. An equilibrium-fixed point model for passenger assignment in congested transit networks. *Technical Report CRT-95-57*, U. de Montréal.

Canca, D., De-Los-Santos, A., Laporte, G., Mesa, J. A., 2019. Integrated railway rapid transit network design and line planning problem with maximum profit. *Transportation Research Part E: Logistics and Transportation Review* 127, 1-30. https://doi.org/10.1016/j.tre.2019.04.007

Carey, M., Hendrickson, C., Siddharthan, K., 1981. A method for direct estimation of origin-destination trip matrices. *Transportation Science* 15(1), 32-49. https://doi.org/10.1287/trsc.15.1.32

Carrese, S., Gori, S., 2002. An urban bus network design procedure. *Applied Optimization* 64, 177-196. https://doi.org/10.1007/0-306-48220-7_11

Cascetta, E., Nguyen, S., 1988. A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B: Methodological* 22(6), 437-455. https://doi.org/10.1016/0191-2615(88)90024-0

Castillo, E., Menéndez, J. M., Jiménez, P., Rivas, A., 2008. Closed form expressions for choice probabilities in the Weibull case. *Transportation Research Part B: Methodological* 42(4), 373-380. https://doi.org/10.1016/j.trb.2007.08.002

Cepeda, M., Cominetti, R., Florian, M., 2006. A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria. *Transportation Research Part B: Methodological* 40(6), 437-459. https://doi.org/10.1016/j.trb.2005.05.006

Chen, A., Chootinan, P., Recker, W., 2005. Examining the quality of synthetic origin-destination trip table estimated by path flow estimator. *Journal of Transportation Engineering* 131(7), 506-513. https://doi.org/10.1061/(ASCE)0733-947X(2005)131:7(506)

Chen, A., Chootinan, P., Recker, W., 2009. Norm approximation method for handling traffic count inconsistencies in path flow estimator. *Transportation Research Part B: Methodological* 43(8), 852-872. https://doi.org/10.1016/j.trb.2009.02.007

Chen, A., Chootinan, P., Ryu, S., Lee, M., Recker, W., 2012. An intersection turning movement estimation procedure based on path flow estimator. *Journal of Advanced transportation* 46(2), 161-176. https://doi.org/10.1002/atr.151

Chen, A., Ryu, S., Chootinan, P., 2010. L_∞-norm path flow estimator for handling traffic count inconsistencies: Formulation and solution algorithm. *ASCE Journal of Transportation Engineering* 136(6), 565-575. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000122

Chen, A., Zhou, Z., Ryu, S., 2011. Modeling physical and environmental side constraints in traffic equilibrium problem. *International Journal of Sustainable Transportation* 5(3), 172-197. https://doi.org/10.1080/15568318.2010.488277

Chen, J., Liu, Z. Y., Wang, S., Chen, X., 2018. Continuum approximation modeling of transit network design considering local route service and short-turn strategy. *Transportation Research Part E: Logistics and Transportation Review* 119, 165-188. https://doi.org/10.1016/j.tre.2018.10.001

Cheng, L., Iida, Y., Uno, N., 2002. Travel time reliability based on sensitivity of capacitated user equilibrium flow. *Traffic and Transportation Studies*, 682-689. https://doi.org/10.1061/40630(255)96

Chootinan, P., Chen, A., 2011. Confidence interval estimation for path flow estimator. *Transportation Research Part B: Methodological* 45(10), 1680-1698. https://doi.org/10.1016/j.trb.2011.07.001

Chootinan, P., Chen, A., Recker, W., 2005. Improved path flow estimator for estimating origin-destination trip tables. *Transportation Research Record* 1923, 9-17. https://doi.org/10.1177%2F0361198105192300102

Chriqui, C., Robillard, P., 1975. Common bus lines. *Transportation Science* 9(2), 115-121. https://doi.org/10.1287/trsc.9.2.115

Codina, E., 2013. A variational inequality reformulation of a congested transit assignment model by Cominetti, Correa, Cepeda, and Florian. *Transportation Science* 47(2), 231-246. https://doi.org/10.1287/trsc.1120.0427

Codina, E., Rosell, F., 2017. A heuristic method for a congested capacitated transit assignment model with strategies. *Transportation Research Part B: Methodological* 104, 293-320. https://doi.org/10.1016/j.trb.2017.07.008

Cominetti, R., Correa, J., 2001. Common-lines and passenger assignment in congested transit networks. *Transportation Science* 35(3), 250-267. https://doi.org/10.1287/trsc.35.3.250.10154

Cui, A., 2006. Bus passenger origin–destination matrix estimation using automated data collection systems. *M.S. Thesis*, Massachusetts Institute of Technology, Cambridge, MA. http://hdl.handle.net/1721.1/37970

Dan, T., Lodi, A., Marcotte, P., 2021. An exact algorithmic framework for a class of mixed-integer programs with equilibrium constraints. *SIAM Journal on Optimization* 31(1), 275-306. https://doi.org/10.1137/18M1208769

de Cea, J., Bunster, J. P., Zubieta, L., Florian, M., 1988. Optimal strategies and optimal routes in public transit assignment models: an empirical comparison. *Traffic Engineering & Control* 29(10), 520-426.

de Cea, J., Fernández, E., 1993. Transit assignment for congested public transport systems: an equilibrium model. *Transportation Science* 27(2), 133-147. https://doi.org/10.1287/trsc.27.2.133

de Cea, J., Fernández, J. E., Dekock, V., Soto, A., 2005. Solving network equilibrium problems on multimodal urban transportation networks with multiple user classes. *Transport Reviews* 25(3), 293-317. https://doi.org/10.1080/0144164042000335805

Dial, R. B., 1971. A probabilistic multipath traffic assignment algorithm which obviates path enumeration. *Transportation Research* 5(2), 83-111.

Dijkstra, E. W., 1959. A note on two problems in connexion with graphs. *Numerische Mathematik* 1, 269-271.

Emme/4.3.2 software, 2020. INRO Consultants, Montréal.

Fan, L., Mumford, C., 2010. A metaheuristic approach to the urban transit routing problem. *Journal of Heuristics* 16(3), 353-372. https://doi.org/10.1007/s10732-008-9089-8

Farahani, R., Miandoabchi, E., Szeto, W., Rashidi, H., 2013. A review of urban transportation network design problems. *European Journal of Operational Research* 229(2), 281-302. https://doi.org/10.1016/j.ejor.2013.01.001

Fisk, C., 1988. On combining maximum entropy trip matrix estimation with user optimal assignment. *Transportation Research Part B: Methodological* 22(1), 69-73. https://doi.org/10.1016/0191-2615(88)90035-5

Florian, M., 1977. A traffic equilibrium model of travel by car and public transit modes. *Transportation Science* 8, 166-179. https://doi.org/10.1287/trsc.11.2.166

Florian, M., Spiess, H., 1982. The convergence of diagonalization algorithms for asymmetric network equilibrium problems. *Transportation Research Part B: Methodological* 16(6), 477-483. https://doi.org/10.1016/0191-2615(82)90007-8

Gao, G., Sun, H., Wu, J., 2019. Activity-based trip chaining behavior analysis in the network under the parking fee scheme. *Transportation* 46(3), 647-669. https://doi.org/10.1007/s11116-017-9809-8

García-Ródenas, R., Marín, A., 2009. Simultaneous estimation of the origin-destination matrices and the parameters of a nested logit model in a combined network equilibrium model. *European Journal of Operational Research* 197(1), 320-331. https://doi.org/10.1016/j.ejor.2008.05.032

Guan, J. F., Yang, H., Wirasinghe, S. C., 2004. Simultaneous optimization of transit line configuration and passenger line assignment. *Transportation Research Part B: Methodological* 40(10), 885-902. https://doi.org/10.1016/j.trb.2005.12.003

Guihaire, V., Hao, J.K., 2008. Transit network design and scheduling: a global review. *Transportation Research Part A: Policy and Practice* 42(10), 1251-1273. https://doi.org/10.1016/j.tra.2008.03.011

He, F., Yin, Y., Zhou, J., 2015. Deploying public charging stations for electric vehicles on urban road networks. *Transportation Research Part C: Emerging Technologies* 60, 227-240. https://doi.org/10.1016/j.trc.2015.08.018

Huang, D., Liu, Z. Y., Liu, P., Chen, J., 2016. Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure. *Transportation Research Part E: Logistics and Transportation Review* 96, 1-19. https://doi.org/10.1016/j.tre.2016.10.004

Huang, H., Li, Z., 2007. A multiclass, multicriteria logit-based traffic equilibrium assignment model under ATIS. *European Journal of Operational Research* 176(3), 1464-1477. https://doi.org/10.1016/j.ejor.2005.09.035

Jansuwan, S., Chen, A., Ryu, S., 2012. An alternative planning tool for a small metropolitan planning organization in Utah. *Transportation Research Record* 2307, 68-79. https://doi.org/10.3141%2F2307-08

Jansuwan, S., Ryu, S., Chen, A., 2017. A two-stage approach for estimating statewide truck trip table. *Transportation Research Part A: Policy and Practice* 102, 274-292. https://doi.org/10.1016/j.tra.2016.09.013

Ji, Y., Mishalani, R. G., Mccord, M. R., 2015. Transit passenger origin-destination flow estimation: efficiently combining onboard survey and large automatic passenger count datasets. *Transportation Research Part C: Emerging Technologies* 58, 178-192. https://doi.org/10.1016/j.trc.2015.04.021

Kitthamkesorn, S., Chen, A., 2013. A path-size weibit stochastic user equilibrium model. *Transportation Research Part B: Methodological* 57, 378-397. https://doi.org/10.1016/j.trb.2013.06.001

Kitthamkesorn, S., Chen, A., 2014. Unconstrained weibit stochastic user equilibrium model with extensions. *Transportation Research Part B: Methodological* 59, 1-21. https://doi.org/10.1016/j.trb.2013.10.010

Kitthamkesorn, S., Chen, A., 2017. Alternate weibit-based model for assessing green transport systems with combined mode and route travel choices. *Transportation Research Part B: Methodological* 103, 291-310. https://doi.org/10.1016/j.trb.2017.04.011

Kitthamkesorn, S., Chen, A., Xu, X., 2015. Elastic demand with weibit stochastic user equilibrium flows and application in a motorised and non-motorised network. *Transportmetrica A: Transport Science* 11(2), 158-185. https://doi.org/10.1080/23249935.2014.944241

Kitthamkesorn, S., Chen, A., Xu, X., Ryu, S., 2016. Modeling mode and route similarities in network equilibrium problem with go-green modes. *Networks and Spatial Economics* 16(1), 33-60. https://doi.org/10.1007/s11067-013-9201-y

Kurauchi, F., Bell, M. G. H., Schmöcker, J. D., 2003. Capacity constrained transit assignment with common lines. *Journal of Mathematical Modelling and Algorithms* 2(4), 309-327. https://doi.org/10.1023/B:JMMA.0000020426.22501.c1

Lam, W. H. K., Gao, Z.Y., Chan, K. S., Yang, H., 1999. A stochastic user equilibrium assignment model for congested transit networks. *Transportation Research part B: Methodological* 33, 351-368. https://doi.org/10.1016/S0191-2615(98)00040-X

Lam, W. H. K., Wu, Z. X., Chan, K. S., 2003. Estimation of transit origin-destination matrices from passenger counts using a frequency-based approach. *Journal of Mathematical Modelling and Algorithms* 2(4), 329-348. https://doi.org/10.1023/B:JMMA.0000020423.93104.14

Lam, W. H. K., Xu, G., 1999. A traffic flow simulator for network reliability assessment. *Journal of Advanced Transportation* 33(2),159-182. https://doi.org/10.1002/atr.5670330206

Lam, W. H. K., Zhou, J., Sheng, Z. H., 2002. A capacity restraint transit assignment with elastic line frequency. *Transportation Research Part B: Methodological* 36(10), 919-938. https://doi.org/10.1016/S0191-2615(01)00042-X

Li, T., 2016. A demand estimator based on a nested logit model. *Transportation Science* 51(3), 918-930. https://doi.org/10.1287/trsc.2016.0671

Li, T., Baik, H., Trani, A. A., 2013. A method to estimate the historical us air travel demand. *Journal of Advanced Transportation* 47(3), 249-265. https://doi.org/10.1002/atr.1200

Li, Y. W., Cassidy, M. J., 2007. A generalized and efficient algorithm for estimating transit route ODs from passenger counts. *Transportation Research Part B: Methodological* 41(1), 114-125. https://doi.org/10.1016/j.trb.2006.04.001

Li, Z. C., Lam, W. H. K., Wong, S. C., Sumalee, A., 2012. Design of a rail transit line for profit maximization in a linear transportation corridor. *Transportation Research Part E: Logistics and Transportation Review* 48(1), 50-70. https://doi.org/10.1016/j.tre.2011.05.003

Liu, H., He, X., He, B. S., 2009. Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem. *Network and Spatial Economics* 9(4), 485-503. https://doi.org/10.1007/s11067-007-9023-x

Lu, X. S., Liu, T. L., Huang, H. J., 2015. Pricing and mode choice based on nested logit model with trip-chain costs. *Transport Policy* 44, 76-88. https://doi.org/10.1016/j.tranpol.2015.06.014

Lundgren, J. T., Peterson, A., 2008. A heuristic for the bi-level origin-destination-matrix estimation problem. *Transportation Research Part B: Methodological* 42(4), 339-354. https://doi.org/10.1016/j.trb.2007.09.005

Mahdavi Moghaddam, S. M. H., Rao, K. R., Tiwari, G., Biyani, P., 2019. Simultaneous bus transit route network and frequency setting search algorithm. *Journal of Transportation Engineering, Part A: Systems* 145(4), 04019011. https://doi.org/10.1061/JTEPBS.0000229

Maher, M. J., Zhang, X., Van Vliet, D., 2001. A bi-level programming approach for trip matrix estimation and traffic count problems with stochastic user equilibrium link counts. *Transportation Research Part B: Methodological* 35, 23-40. https://doi.org/10.1016/S0191-2615(00)00017-5

Maricopa Association of Governments and Valley Metro Transit System, 2015. 2014-2015 Valley Metro onboard survey-final report. https://www.valleymetro.org/sites/default/files/uploads/event-resources/2014-2015_onboard_survey_final.pdf.

Maruyama, T., Harata, N., 2005. Incorporating trip-chaining behavior into network equilibrium analysis. *Transportation Research Record* 1921, 11-18. https://doi.org/10.1177%2F0361198105192100102

Maruyama, T., Harata, N., 2006. Difference between area-based and cordon-based congestion pricing: investigation by trip-chain-based network equilibrium model with non-additive path costs. *Transportation Research Record* 1964, 1-8. https://doi.org/10.1177%2F0361198106196400101

Maruyama, T., Sumalee, A., 2007. Efficiency and equity comparison of cordon- and area-based road pricing schemes using a trip-chain equilibrium model. *Transportation Research Part A: Policy and Practice* 41(7), 655-671. https://doi.org/10.1016/j.tra.2006.06.002

Munizaga, M. A., Palma, C., 2012. Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies* 24, 9-18. https://doi.org/10.1016/j.trc.2012.01.007

Nguyen, S., Morello, E., Pallottino, S., 1988. Discrete time dynamic estimation model for passengers origin-destination matrices on transit networks. *Transportation Research Part B: Methodological* 22(4), 251-260. https://doi.org/10.1016/0191-2615(88)90002-1

Nguyen, S., Pallottino, S., 1988. Equilibrium traffic assignment for large scale transit networks. *European Journal of Operational Research* 37(2), 176-186. https://doi.org/10.1016/0377-2217(88)90327-X

Nie, Y., Zhang, H. M, Recker, W., 2005. Inferring origin destination trip matrices with a decoupled GLS path flow estimator. *Transportation Research Part B: Methodological* 39(6), 497-518. https://doi.org/10.1016/j.trb.2004.07.002

Nuzzolo, A., Crisalli, U., 2001. Estimation of transit origin-destination matrices from traffic counts using a schedule-based approach. *Proceedings of the AET 2001*, Homerton College, Cambridge.

Oppenheim, N., 1995. Urban Travel Demand Modeling. John Wiley and Sons Inc., New York.

Primerano, F., Taylor, M., Pitaksringkarn, L., Tisato, P., 2008. Defining and understanding trip chaining behaviour. *Transportation* 35(1), 55-72. https://doi.org/10.1007/s11116-007-9134-8

Ryu, S., Chen, A., Choi, K., 2017. Solving the combined modal split and traffic assignment problem with two types of transit impedance function. *European Journal of Operational Research* 257(3), 870–880. https://doi.org/10.1016/j.ejor.2016.08.019

Ryu, S., Chen, A., Su, J., Choi, K., 2018. Two-stage bicycle traffic assignment model. *Journal of Transportation Engineering, Part A: Systems* 144(2), 04017079. https://doi.org/10.1061/JTEPBS.0000108

Ryu, S., Chen, A., Xu, X., Choi, K., 2014a. A dual approach for solving the combined distribution and assignment problem with link capacity constraints. *Network and Spatial Economics* 14, 245-270. https://doi.org/10.1007/s11067-013-9218-2

Ryu, S., Chen, A., Zhang, H.M., Recker, W., 2014b. Path flow estimator for planning applications of small communities. *Transportation Research Part A: Policy and Practice* 69, 212-242. https://doi.org/10.1016/j.tra.2014.08.019

Sheffi, Y., 1985. Urban transportation networks: equilibrium analysis with mathematical programming methods. *Prentice-Hall*, Englewood Cliff, New Jersey.

Shi, F., Zhou, Z., Yao, J., Huang, H., 2012. Incorporating transfer reliability into equilibrium analysis of railway passenger flow. *European Journal of Operational Research* 220(2), 378-385. https://doi.org/10.1016/j.ejor.2012.02.012

Shimamoto, H., Higuchi, T., Uno, N., Shiomi, Y., 2016. A trip-chain-based combined mode and route-choice network equilibrium model. *Transportmetrica B: Transport Dynamics* 5(3), 307-324. https://doi.org/10.1080/21680566.2016.1207577

Smith, M. J., 1983. The existence and calculation of traffic equilibria. *Transportation Research Part B: Methodological* 17(4), 291-303. https://doi.org/10.1016/0191-2615(83)90047-4

Spiess, H., 1984. Contributions à la théorie et aux outils de planification de réseaux de transport urbain. *Ph.D. thesis*, Département d'informatique et de recherche opérationnelle, Centre de recherche sur les transports, Université de Montréal.

Spiess, H., Florian, M., 1989. Optimal strategies: a new assignment model for transit networks. *Transportation Research Part B: Methodological* 23(2), 83-102. https://doi.org/10.1016/0191-2615(89)90034-9

Stern, R., 1996. Passenger transfer system review. *Synthesis of Transit Practice* 19, Transportation Research Board.

Sun, S., Szeto, W. Y., 2018. Logit-based transit assignment: approach-based formulation and paradox revisit. *Transportation Research Part B: Methodological* 112, 191-215. https://doi.org/10.1016/j.trb.2018.03.018

Szeto, W. Y., Jiang, Y., 2014. Transit assignment: Approach-based formulation, extragradient method, and paradox. *Transportation Research Part B: Methodological* 62, 51-76. https://doi.org/10.1016/j.trb.2014.01.010

Szeto, W. Y., Wu, Y., 2011. A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong. *European Journal of Operational Research* 209(2), 141-155. https://doi.org/10.1016/j.ejor.2010.08.020

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge: Cambridge University Press.

Transport Department, 2014. Travel characteristics survey 2011 final report.

Trépanier, M., Tranchant, N., Chapleau, R., 2007. Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems* 11, 1-14. https://doi.org/10.1080/15472450601122256

Vovsha, P., 1997. Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area. *Transportation Research Record* 1607(1), 6-15. https://doi.org/10.3141/1607-02

Wang, J., Peeta, S., He, X. Z., Zhao, J. B., 2018. Combined multinomial logit modal split and paired combinatorial logit traffic assignment model. *Transportmetrica A: Transport Science* 14(9), 737-760. https://doi.org/10.1080/23249935.2018.1431701

Wang, T. G., Xie, C., Xie, J., Waller, S. T., 2016. Path-constrained traffic assignment: a trip chain analysis under range anxiety. *Transportation Research Part C: Emerging Technologies* 68, 447-461. https://doi.org/10.1016/j.trc.2016.05.003

Wong, S. C., Tong, C. O., 1998. Estimation of time-dependent origin-destination matrices for transit networks. *Transportation Research Part B: Methodological* 32(1), 35-48. https://doi.org/10.1016/S0191-2615(97)00011-8

Wu, J. H., Florian, M., Marcotte, P., 1994. Transit equilibrium assignment: a model and solution algorithms. *Transportation Science* 28(3), 193-203. https://doi.org/10.1287/trsc.28.3.193

Wu, Z. X., Lam, W. H. K., 2006. Transit passenger origin-destination estimation in congested transit networks with elastic line frequencies. *Annals of Operations Research* 144(1), 363-378. https://doi.org/10.1007/s10479-006-0002-2

Wu Z. X., Lam W. H. K., 2003. Combined modal split and stochastic assignment model for congested networks with motorized and nonmotorized transport modes. *Transportation Research Record* 1831, 57-64. https://doi.org/10.3141%2F1831-07

Xie, C., Wang, T. G., Pu, X., Karoonsoontawong, A., 2017. Path-constrained traffic assignment: modeling and computing network impacts of stochastic range anxiety. *Transportation Research Part B: Methodological* 103, 136-157. https://doi.org/10.1016/j.trb.2017.04.018

Xu, Z., Xie, J., Liu, X., Nie, Y. M., 2020. Hyperpath-based algorithms for the transit equilibrium assignment problem. *Transportation Research Part E: Logistics and Transportation Review* 143, 102102. https://doi.org/10.1016/j.tre.2020.102102

Yang, H., 1995. Heuristic algorithms for the bi-level origin-destination matrix estimation problem. *Transportation Research Part B: Methodological* 29(4), 231-242. https://doi.org/10.1016/0191-2615(95)00003-V

Yang, H., Meng, Q., Bell, M. G. H., 2001. Simultaneous estimation of the origin-destination matrices and travel cost coefficient for congested networks in a stochastic user equilibrium. *Transportation Science* 35(2), 107-123. https://doi.org/10.1287/trsc.35.2.107.10133

Yen, J. Y., 1971. Finding the k shortest loopless paths in a network. *Management Science* 17(11), 712-716. https://doi.org/10.1287/mnsc.17.11.712

Yu, Q., Fang, D., Du, W., 2014. Solving the logit-based stochastic user equilibrium problem with elastic demand based on the extended traffic network model. *European Journal of Operational Research* 239, 112-118. https://doi.org/10.1016/j.ejor.2014.04.009

Zhao, F., 2006. Large-scale transit network optimization by minimizing user cost and transfers. *Journal of Public Transportation* 9(2), 107-129. http://doi.org/10.5038/2375-0901.9.2.6

Zhao, F., Ubaka, I., 2004. Transit network optimization – minimizing transfers and optimizing route directness. *Journal of Public Transportation* 7(1), 67-82. http://doi.org/10.5038/2375-0901.7.1.4

Zhao, J., Rahbee, A., Wilson, N., 2007. Estimating a rail passenger trip origin–destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering* 22(5), 376-387. https://doi.org/10.1111/j.1467-8667.2007.00494.x

Zhou, B., Li, X., He, J., 2014. Exploring trust region method for the solution of logit-based stochastic user equilibrium problems. *European Journal of Operational Research* 239, 46-57. https://doi.org/10.1016/j.ejor.2014.05.002